DIFUSIÓN CULTURAL

III CONGRESO INTERNACIONAL DE FONÉTICA EXPERIMENTAL

XUNTA DE GALICIA

III CONGRESO INTERNACIONAL DE FONÉTICA EXPERIMENTAL

Actas do congreso organizado
pola Dirección Xeral de Creación e Difusión Cultural,
a Universidade de Santiago de Compostela e a Real Academia Galega.
Santiago, 24 -26 de outubro de 2005

XUNTA DE GALICIA
Consellería de Cultura e Deporte
Dirección Xeral de Creación e Difusión Cultural
2007

III Congreso Internacional de Fonética Experimental (3º, 2005. Santiago de Compostela)
Actas do 3º Congreso Internacional de Fonética Experimental. Santiago de Compostela, 24 - 26 de outubro de 2005 / organizado pola Dirección Xeral de Creación e Difusión Cultural da Xunta de Galicia, a Universidade de Santiago e a Real Academia Galega - Santiago de Compostela: Xunta de Galicia, 2006 - 1 vol.: 24 cm
D.L.: C 2027-2007, ISBN: 978-84-453-4451-4
Textos en galego, castelán, inglés, francés e portugués - Bibliografía.
I. González González, Manuel, ed. lit. II. Fernández Rei, Elisa, ed. lit.
III. González Rei, Begoña, ed. lit. IV. Xunta de Galicia
1. Fonética - Congresos e asembleas

Ánxela Bugallo Rodríguez
Conselleira de Cultura e Deporte

Luís Bará Torres
Director xeral de Creación e Difusión Cultural

Francisco Xosé Fernández Naval
Subdirector xeral de Creación e Difusión Cultural

Coordinación: Charo Ferreiro e Inma Pena
Deseño de cuberta: Antonio Castro Ferreiro
Revisión do texto: Carlos Arias

Santiago de Compostela, 2007
ISBN: 978-84-453-4451-4
Dep. legal: C 2027-2007

Edita: Xunta de Galicia
Imprime: Litonor

III CONGRESO INTERNACIONAL DE FONÉTICA EXPERIMENTAL

Manuel González González
Elisa Fernández Rei
Begoña González Rei
Editores

ÍNDICE

COMITÉS

Comité Organizador

Presidente: Manuel González González
Secretaria: Elisa Fernández Rei
Vogais: Carmen García Mateo
Begoña González Rei
Xosé Luís Regueira Fernández
Luís A. Xuncal Pereira

Comité Científico

Manuel González González	Universidade de Santiago
X. L. Regueira Fernández	Universidade de Santiago
Elisa Fernández Rei	Universidade de Santiago
Carmen García Mateo	Universidade de Vigo
Eduardo Rodríguez Banga	Universidade de Vigo
Eugenio Martínez Celdrán	Universitat de Barcelona
Joaquim Llisterri Boix	Universitat Aut. de Barcelona
Antonio Pamies Bertrán	Universidad de Granada
Antonio Rubio Ayuso	Universidad de Granada
Ana Elejabeitia	Universidad de Deusto
Josefa Dorta Luís	Universidad de La Laguna
Michel Contini	Université Stendhal-Grenoble III
Lurdes de Castro Moutinho	Universidade de Aveiro
Antonio Romano	Universitá di Torino

Organizan

XUNTA DE GALICIA
CONSELLERÍA DE CULTURA E DEPORTE

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA
FACULTADE DE FILOLOXÍA

PRESENTACIÓN

Ánxela Bugallo Rodríguez
Conselleira de Cultura e Deporte

Organizado pola Consellería de Cultura e Deporte, a través da Dirección Xeral de Creación e Difusión Cultural, en colaboración coa Real Academia Galega e mais coa Facultade de Filoloxía da Universidade de Santiago, tivo lugar, no edificio desta última, o III Congreso de Fonética Experimental, que reuniu especialistas de nivel internacional para debater e intercambiar experiencias, opinións e liñas de traballo verbo da investigación relacionada coa realización da linguaxe.

Entre os días 24 e 26 de outubro de 2005, o foro de debate afondou nunha das disciplinas de maior evolución na nosa sociedade, con infinidade de aplicacións en áreas tan variadas como a apredizaxe de idiomas, a investigación verbo das manifestacións nas distintas estruturas sociais ou a evolución dos sistemas de intercambio en tempo real de mensaxes de voz formulados en linguas distantes.

Daquela, estas actas recollen a valiosísima achega colectiva das e dos especialistas que acudiron para integrar experiencias e darlles un impulso mercé ao contraste e intercambio de materiais e enfoques. Deste xeito, Galicia trouxo ao seu territorio un foco de primeira orde da ciencia moderna.

Confío plenamente en que as seguintes páxinas figurarán como obra de referencia do público especializado, polo que lles expreso ás persoas que contribuíron no seu funcionamento eficaz o meu recoñecemento máis profundo.

PALABRAS DE APERTURA DO III CONGRESO DE FONÉTICA EXPERIMENTAL

MANUEL GONZÁLEZ GONZÁLEZ
Presidente do Comité organizador
Universidade de Santiago

Excmo. Sr. Presidente da Real Academia Galega, Ilmo. Sr. Director Xeral de Cultura, Ilmo. Sr. Decano da Facultade de Filoloxía da Universidade de Santiago, queridos colegas e amigos participantes neste III Congreso de Fonética Experimental:

Os días 22, 23 e 24 de febreiro de 1999 tiña lugar en Tarragona, na Universitat Rovira i Virgili, o primeiro Congreso de Fonética Experimental, coa finalidade principal de poñer en común os traballos dos fonetistas do Estado Español, de maneira especial, os que traballan nas distintas Universidades. Nacía este primeiro Congreso tamén coa vontade de continuidade, e daquela considerouse que sería desexable organizar este evento cunha cadencia de dous anos. Cadencia que se cumpriu na súa segunda edición, xa que o 5,6 e 7 de marzo de 2001 tivo lugar o II CFE na Universidade de Sevilla. Foi aquí cando se tomou o acordo de que o próximo se celebrase en Santiago de Compostela, pero non no 2003, como correspondería, xa que no mes de agosto dese ano tiña lugar en Barcelona a XV edición do evento máis importante no ámbito da fonética, o Congreso Internacional de Ciencias Fonéticas. Chegamos deste xeito a este III Congreso de Fonética Experimental, que hoxe inauguramos na Facultade de Filoloxía da Universidade de Santiago.

Tamén a Asemblea da SEAF, no seu último Congreso de Barcelona, no ano 2003, decidiu que este ano se celebrase en Santiago, e pediume que puxese en marcha a súa organización. Dada a proximidade entre a fonética experimental e a acústica forense, e, tendo en conta, que non poucos investigadores de acústica forense son figuras da fonética experimental, pareceunos que era unha boa idea que os dous congresos se colleran da man, e se celebrasen un a continuación do outro, facilitando deste xeito a participación de moitas persoas que teñen interese en ámbolos campos.

No Congreso que inauguramos hoxe, o de Fonética Experimental, exporanse ao longo de tres días os 55 relatorios admitidos polo comité científico, o que supón ao redor do 80% dos presentados, e que aparecen asinados por 110 investigadores, estruturados nas seguintes áreas temáticas

- Fonética articulatoria
- Fonética acústica
- Fonética perceptiva
- Fonética xeral
- Prosodia
- Patoloxías lingüísticas
- Síntese de voz
- Recoñecemento de voz
- Tecnoloxías aplicadas

As sesións de cada día abriranse coas tres conferencias plenarias por invitación, impartidas por personalidades relevantes da fonética actual (Joaquim Llisterri, Eugenio Martínez Celdrán e Michel Contini). Están programadas dúas mesas redondas, unha delas destinada a ver os proxectos máis importantes que se están levando a cabo neste momento nos distintos laboratorios de fonética, e outra destinada a debater a importancia das tecnoloxías da fala na sociedade da información, tema fundamental se temos en conta a importancia que a síntese de voz, o recoñecemento de voz, os sistemas de diálogo home-máquina teñen na automatización das comunicacións da sociedade actual, pero sobre todo da sociedade do futuro inmediato. Debemos facer todos un esforzo para dar a coñecer ao conxunto da sociedade a importancia do noso traballo, que moitas veces non é comprendido, e sinto certo rubor ao confesalo, ás veces nin sequera por compañeiros do mundo universitario, debemos ser capaces de facerlles chegar a trascendencia das nosas investigacións non só no campo teórico, senón tamén no campo aplicado, nas tecnoloxías da fala, na facilitación da adquisición de segundas e terceiras linguas en sociedades cada vez máis plurilingües, na logopedia, no ámbito xudicial, etc.

Esta é unha das razóns polas que quixemos que neste Congreso houbese tamén un un espazo dedicado á exposición de ferramentas, programas e outros produtos relacionados coa fonética e coas tecnoloxías da fala. Estes produtos e ferramentas poderán ser analizados polos congresistas na sala habilitada para tal efecto nos horarios que se indican no programa.

Desexo que este Congreso sexa un agradable lugar de encontro para todos aqueles que, desde diversos países do mundo, se achegan a Compostela, destino final de tantos camiños ao longo da historia, mesmo que sexa para falar ou discutir de fonética. Estou certo de que será unha plataforma eficiente e de resonancia para os participantes expoñeren os seus estudos, os seus proxectos e as súas descubertas. Espero sobre cada un dos temas exista un diálogo e un debate franco, aberto e froitífero, que sirva para enriquecernos a todos, pero sobre todo para o avance da ciencia e da técnica, avance que no mundo de hoxe é practicamente imposible sen o intercambio científico.

Teño que agradecer publicamente á Consellería de Cultura da Xunta de Galicia o compromiso que adquiriu para a organización deste Congreso e todos os medios que dunha maneira xenerosa puxo ao noso dispor para que esta iniciativa cumpra cos seus obxectivos, agradecemento que quero facer extensivo á Real Academia Galega e á Universidade de Santiago, as outras dúas entidades coorganizadoras.

Que teñan unha estadía agradable en Santiago, e que leven un grato recordo deste Congreso, desta cidade e de Galicia.

CONFERENCIAS PLENARIAS

EL PAPEL DE LA FONÉTICA EN LAS TECNOLOGÍAS DEL HABLA

JOAQUIM LLISTERRI
Departamento de Filología Española. Universitat Autonoma de Barcelona

"Can we leave it to the computer to learn about speech or shall we insist on developing our own insights in the many dimensions of the speech code?" (Fant 1983: 17)
"Computing power can not substitute crucial knowledge" (Fant 2004: 11)

1. LA RELACIÓN ENTRE LA FONÉTICA Y LAS TECNOLOGÍAS DEL HABLA

Cuando en 1983 Gunnar Fant se dirigía a los participantes en el décimo Congreso Internacional de Ciencias Fonéticas, planteaba una pregunta, seguramente retórica, a la que daba respuesta en 2004, afirmando que, frente al dilema entre las aproximaciones al estudio del habla basadas en el conocimiento y las que se fundamentan en el uso de técnicas informáticas, la potencia de cálculo no puede sustituir al conocimiento (Fant 1983, 2004). Ésta es, en esencia, una de las cuestiones centrales cuando se aborda la relación entre la fonética y las tecnologías del habla, al igual que sucede al considerar el papel de otras disciplinas lingüísticas en el procesamiento del lenguaje natural (Rodríguez 2004).

Si bien en sus inicios las tecnologías del habla pretendían incorporar varios niveles de conocimiento fonético, parece existir en la actualidad un cierto consenso para reconocer que los últimos años no han sido precisamente los más fértiles en lo que se refiere al uso de información fonética en la síntesis y en el reconocimiento; así lo señalan, entre otros, Strik (2005: 168) -"...in the last decades we have witnessed a decrease in the amount of phonetic knowldege used in ASR and TTS"- y Barry *et al* (2005: 1) -"... the linguistic approach soon lost terrain, in recognition applications at least, to (nonlinguistically oriented) engineers who were less concerned with formal linguistic insights, treating the signal as a pattern just like any other, and this with outstanding success"-, esgrimiendo las razones que discutiremos más adelante.

Otro de los hechos que más llaman la atención cuando se analiza la relación entre las dos disciplinas que nos ocupan es el contraste entre las constantes referencias a la necesidad de colaboración entre fonetistas y tecnólogos y la situación que se desprende de las publicaciones: si éstas reflejan acertadamente la realidad, no parece que, al menos en nuestro contexto más inmediato, sobren motivos para el optimismo en lo que a la interdisciplinariedad se refiere.

A modo de ejemplo, un estudio de las comunicaciones presentadas en los congresos anuales de la Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN) celebrados entre 2000 y 2005 revela que, de 71 contribuciones que podemos considerar relacionadas con las tecnologías del habla, únicamente tres (4%) aparecen firmadas conjuntamente por autores vinculados a departamentos tecnológicos y a departamentos de filología; si añadimos a ello que las tres proceden del mismo equipo (González y García 2000; González *et al.* 2001; Rodríguez *et al.* 2002), podemos darnos cuenta de que lo que en teoría se considera deseable es, en España, prácticamente inexistente[1]. Por ello no sorprende, tampoco, que los cuatro ponentes en el taller sobre "Tecnologías del habla: pasado, presente y futuro. Particularización sobre tecnología del español" (*sic*), que tuvo lugar en ocasión del congreso de la SEPLN el 2003, pertenecieran a Escuelas Técnicas Superiores de Ingenieros de Telecomunicación.

Es también sintomático que en los congresos de Fonética Experimental celebrados en 1999, 2001 y 2005, únicamente 9 (8,5%) del total de las 106 comunicaciones presentadas se centraran en las tecnologías del habla; si computamos solamente aquellas cuyos autores están vinculados a una universidad española, el porcentaje baja al 4,7% (5 comunicaciones). Otro dato que no deja de ser significativo es la presencia de 14 autores de comunicaciones que, entre 1999 y 2005, han participado tanto en los congresos de la SEPLN como en los de Fonética Experimental; esto representa aproximadamente un 10% de los investigadores que presentaron contribuciones a la SEPLN y un 13% de los que lo hicieron a los congresos de Fonética Experimental; sin embargo, entre ellos se cuentan diez investigadores asociados al mismo proyecto (Battaner *et al.* 2005a, b) y dos que firman conjuntamente sus trabajos (Escudero y Cardeñoso 2001, 2002) con lo que, en realidad, el número de equipos presentes en ambos congresos se reduce a tres[2]. La división entre la comunidad dedicada a la fonética y la centrada en las tecnologías del habla parece pues, en lo que se refiere a nuestro país, claramente marcada.

En las páginas que siguen, presentaremos, en primer lugar, algunos de los campos en los que las tecnologías del habla pueden beneficiarse del conocimiento fonético y exploraremos, a continuación, las razones que obstaculizan la integración entre ambas disciplinas. Esbozaremos, para concluir, algunas propuestas orientadas al futuro.

---

1 De este dato, sin embargo, no puede desprenderse que no existan otros equipos o proyectos en los que colaboran lingüistas y expertos en informática o en ingeniería de telecomunicación; más bien cabe pensar que, en algunos casos, la participación de los lingüistas no siempre se refleja en la autoría de las publicaciones (Hernando *et al.* 2002).

2 Los dos anteriormente mencionados y Pérez *et al.* (2002)

2. LA INCORPORACIÓN DE CONOCIMIENTOS FONÉTICOS A LAS TECNOLOGÍAS DEL HABLA

Las ventajas y los problemas derivados de incorporar información fonética a los sistemas desarrollados en el marco de las tecnologías del habla y la relación entre ambas disciplinas son cuestiones que aparecen de modo recurrente en diversos trabajos, tanto en reflexiones generales (Fant 1983, 1989, 2004, 2005; Moore 1995; Rossi 1996; Pols 1999, 2001; Greenberg 2001a, b, 2005; Öhman 2001; Ainsworth 2005; Barry *et al.* 2005; Strik 2005), como en aportaciones centradas en la síntesis (Carlson y Granstrom 1991; Fant 1991; Huckvale 2002; van Santen 2005), el reconocimiento (Greenberg 1998; Dusan y Rabiner 2005) o los sistemas de diálogo (Shriberg 2005).

Como resultado de la colaboración en diversos proyectos en el campo de las tecnologías del habla, hemos intentado describir, en varias publicaciones de nuestro grupo, el papel del fonetista en este contexto (Aguilar *et al.* 1997; Garrido *et al.* 2000; Llisterri 1990, 2003a, b; Llisterri *et al.* 2003a, b, 2004). Esencialmente, y coincidiendo también con parte de las tareas que en Acero (1995) se describen como propias de un especialista en fonética o en lingüística –siempre, naturalmente, en estrecha colaboración con el resto de los profesionales que intervienen en el desarrollo de un sistema-, podríamos mencionar los ámbitos que se detallan a continuación.

(1) Conversión de texto en habla

- Reglas para el procesamiento previo del texto que contemplen la expansión de los signos de puntuación sin valor lingüístico, las expresiones numéricas, las siglas y las abreviaturas.
- Reglas para el procesamiento morfológico y sintáctico en los casos en que el conversor contempla un análisis lingüístico automático del texto de entrada o supervisión manual del resultado de un proceso de etiquetado automático.
- Reglas de transcripción fonética automática, que establecen la correspondencia entre grafías y alófonos, la silabación y la acentuación, complementadas por diccionarios de pronunciación para el tratamiento de las excepciones.
- Modelos de duración segmental que consideren los diversos factores que influyen en la duración y basados en datos procedentes de corpus representativos.
- Modelos de intensidad segmental que, igualmente, consideren los factores que inciden en la intensidad y procedan de corpus representativos.
- Modelos de asignación de pausas que contemplen tanto las marcadas mediante signos de puntuación como las no marcadas y que establezcan, además, diferencias de duración entre los distintos tipos de pausas.
- Modelos de entonación que permitan generar una curva melódica natural, teniendo en cuenta factores fonéticos, sintácticos, semánticos y pragmáticos.

- Establecimiento del inventario de fonemas y alófonos de la lengua para la constitución del diccionario de unidades de síntesis.
- Diseño del corpus de unidades de síntesis teniendo en cuenta las restricciones fonotácticas de la lengua y la frecuencia de aparición de las unidades.
- Selección del locutor para la grabación del corpus de síntesis.
- Supervisión de la grabación del corpus de síntesis para asegurar una realización adecuada de los elementos segmentales y suprasegmentales, tanto en los sistemas de síntesis por concatenación como en los basados en selección de unidades.
- Segmentación (o supervisión de una segmentación semiautomática) del corpus de unidades de síntesis.
- Evaluación objetiva de los distintos módulos del conversor, en un proceso iterativo que permita la corrección de errores.
- Diseño de pruebas de evaluación subjetiva tanto de la inteligibilidad como de la naturalidad del resultado de la conversión de texto en habla.

(2) Reconocimiento del habla

- Definición del inventario de fonemas y alófonos de la lengua para determinar las unidades del sistema de reconocimiento.
- Diseño del corpus de entrenamiento teniendo en cuenta el inventario de unidades previamente definido y las restricciones sobre su aparición.
- Selección de la muestra de población para la grabación del corpus de entrenamiento, considerando factores de variación individual, geográfica, social y de registro.
- Segmentación (o supervisión de una segmentación automática) del corpus de entrenamiento del reconocedor.
- Realización o validación de los diccionarios de pronunciación, que incorporan las formas canónicas y las variantes encontradas en el corpus.
- Análisis fonético acústico de corpus significativamente amplios para obtener información sobre los factores que condicionan la variabilidad segmental y suprasegmental, contemplando tanto el nivel fonético como otros niveles más altos del análisis lingüístico.

(3) Sistemas de diálogo

- Transcripción, anotación y estudio de corpus de interacciones naturales entre personas para definir el dominio del sistema, diseñar posibles estrategias de gestión del diálogo y establecer los escenarios que se emplearán en el corpus de entrenamiento.
- Selección de la muestra de población para la obtención del corpus de entrenamiento, teniendo en cuenta factores de variación individual, geográfica, social y de registro.
- Transcripción, anotación y análisis de corpus de interacciones simuladas (obtenidas mediante el protocolo del Mago de Oz) para el entrenamiento del sistema.

- Diseño de estrategias de acceso, de salida y de confirmación adecuadas desde el punto de vista pragmático.
- Estudio de los fenómenos propios del habla espontánea para modelarlos adecuadamente en el sistema de diálogo.
- Análisis de la relación entre el nivel fonético y el nivel pragmático, especialmente en lo que se refiere a las manifestaciones prosódicas de los actos que habla y a los correlatos acústicos de las emociones.
- Verificación del grado de corrección y adecuación lingüística del diálogo.

Cabe recordar, naturalmente, que parte del conocimiento que se adquiere mediante las tareas que acabamos de enumerar puede obtenerse a través de técnicas de aprendizaje automático aplicadas a grandes corpus de habla. Como bien apunta Greenberg (2005: 111) "Speech technology can proudly point to its apparent success with speech recognition and concatenative synthesis in defense of its machine-learning-centric approach"; sin embargo, precisa más adelante que "...imperfect science is capable of providing an effective foundation for technology - as long as the demands of the market are not exceedingly stringent or profund". Por tanto, existen no sólo argumentos teóricos para insistir en la necesidad de incorporar conocimiento, sino también consideraciones prácticas para intentar superar las limitaciones con las que actualmente nos encontramos.

3. LOS OBSTÁCULOS PARA LA INTEGRACIÓN DEL CONOCIMIENTO FONÉTICO EN LAS TECNOLOGÍAS DEL HABLA

Pese a que se trata, sin duda, de una cuestión compleja, podrían distinguirse dos clases de obstáculos que surgen en el momento de incorporar conocimiento fonético a las tecnologías del habla: por una parte, los derivados de la propia naturaleza de la información fonética disponible y, por otra, los que obedecen a las distintas tradiciones académicas o al contexto en el que se lleva a cabo la investigación.

Strik (2005: 177) formula el problema de un modo muy directo: "...phonetics does not provide ready-made quantitative models that can be plugged directly into a system", lo que nos lleva al primer tipo de dificultades al que hacíamos referencia. En efecto, en muchas ocasiones la información fonética sobre un determinado fenómeno no es suficientemente detallada, no está cuantificada o no se expresa con el formalismo adecuado para un entorno computacional. Si, por ejemplo, nos ceñimos a la descripción del español observamos que, pese a la proliferación de estudios acústicos, estamos lejos de disponer de los datos que se requieren en muchas de las aplicaciones de las tecnologías del habla (Gil y Llisterri 2004). Siguiendo la distinción que establecen Barry *et al.* (2005: 10) entre hacerse una idea de un fenómeno, obtener datos cuantitativos sobre el mismo y presentarlos en un formato que permita incorporarlos a una aplicación, podríamos decir que nos encontramos, tal vez, todavía en el primer estadio.

Otra de las limitaciones de los estudios fonéticos que habitualmente se señalan es que los datos provienen, en su mayor parte, de trabajos de laboratorio, obtenidos me-

diante un diseño experimental en el que se establece un control de las variables y, a menudo, en condiciones muy distintas a las del uso real de las tecnologías del habla. Por tal motivo se ha propuesto que el fonetista podría emplear con provecho los grandes corpus recogidos para el entrenamiento de sistemas de reconocimiento u otras bases de datos que respondan a situaciones más realistas. Greenberg (2005), por citar un ejemplo, ofrece una muestra de este tipo de estudio, que responde a los planteamientos de Barry *et al.* (2005: 11): "The greater access phonetically trained researchers have to the databases and tools used in mainline techology applications, the more likely is that quantitative answers to phonetic questions can be presented in a way which can be useful for speech technology applications". Sin embargo, es conocido el problema de la disponibilidad de los corpus, tanto por motivos económicos como por los derivados de políticas científicas que favorecen el desarrollo de recursos que no se reutilizan (Llisterri 2004b). Para el español existen las bases de datos distribuidas por ELDA (*Evaluation and Language resources Distribution Agency*) y por el LDC (*Linguistic Data Consortium*) además de las creadas en el marco de diversos proyectos que han gozado de financiación pública (Llisterri *et al.* 2005) pero aun así, no siempre es fácil localizar y conseguir acceder a estos recursos.

Un tercer aspecto inherente al propio carácter de la fonética como disciplina científica es la proliferación de modelos que, en el terreno que nos ocupa, se hace especialmente patente en el campo de la prosodia. Desde la perspectiva del desarrollo de las tecnologías del habla, "There is too much emphasis on theoretical concepts and on the discussion of which one is better suited for the description of a special language or of languages in general." (Batliner y Möbius 2005: 25), lo que a menudo introduce, en opinión de los autores citados, niveles de abstracción que tal vez no son necesarios en el contexto de las aplicaciones, de modo que se pierde parcialmente la distinción entre el conocimiento básico y el que, de alguna manera, está mediatizado por un modelo[3]. Quizás por este motivo, en las comunicaciones presentadas en los congresos de la SEPLN a los que hacíamos referencia anteriormente encontramos trabajos sobre modelos prosódicos llevados a cabo por investigadores vinculados a departamentos de informática (Escudero y Cardeñoso 2001) o de ingeniería de telecomunicación (Fernández y Rodríguez 2000; Navas *et al.* 2002; Agüero y Bonafonte 2003) sin que en ellos intervengan especialistas en fonética y en los que, como es de esperar, se emplean técnicas de aprendizaje automático basadas en corpus prescindiendo de aportaciones procedentes del campo de la fonética.

La segunda categoría de obstáculos es de naturaleza académica –o, en ocasiones, el resultado de una determina política científica- y responde a la tradicional separación entre estudios humanísticos y tecnológicos. Una consecuencia importante es la falta de formación interdisciplinar que dificulta la formación de equipos mixtos y, a la vez, no facilita la incorporación de lingüistas o filólogos a un ámbito que podría ofrecer salidas profesionales. Compárese, por ejemplo, el perfil que define Acero (1995: 175) con los

3 A modo de ejemplo, Batliner y Möbius (2005: 26) señalan que "Phonological systems like the ToBI approach only introduce a quantisation error: the whole variety of F0 values available in acoustics is reduced to a mere binary opposition L *vs.* H, and to some few additional, diacritic distinctions".

planes de estudio vigentes -y, probablemente, futuros- en las facultades de letras españolas: "... a successful phonetician working on a spoken language system will need some knowledge of computers, algorithms, statistics and signal processing [...] Also desired is proficiency with common computing environments such as Windows, UNIX and Macintosh, text editors, and speech analysis packages". Insiste el actual responsable de la investigación en el área de tecnologías del habla de Microsoft que el fonetista no debe ser un experto en todos estos campos, pero sí debe tener un conocimiento básico suficiente que le permita incorporarse a un equipo. Al no darse estas condiciones, "One of the reasons why there are not more linguists working in building Spoken Language Systems is that in many cases, lack of training in these other disciplines prevent them to be as effective in the team as an engineer or a programmer" (Acero 1995: 175).

Cierto es que algunas universidades ofrecen cursos de postgrado en los que se pretende proporcionar una formación integral como la que acabamos de describir. La oferta actual en España no parece ser especialmente amplia; por ejemplo, el máster en Lingüística y Aplicaciones Tecnológicas que programa la Universitat Pompeu Fabra, ofrece únicamente 3 créditos de fonética y fonología y 3 créditos sobre tratamiento del habla, mientras que en el *European Masters in Language and Speech* que se imparte en la Universitat Politècnica de Catalunya es obligatoria una asignatura sobre procesamiento del habla (6 créditos) y pueden cursarse, como asignaturas de libre elección, una introducción a la fonética y a la fonología (6 créditos) y una materia sobre percepción del lenguaje (3 créditos), ambas ofrecidas por profesores de la Universitat de Barcelona. Por el momento, pues, la formación de postgrado en fonética y en tecnologías del habla no tiene un espacio propio en nuestro país, pese a la existencia de grupos consolidados y productivos en ambos campos (Llisterri 2004[a]; Rubio y Hernáez 2005).

De un modo acorde a lo que sucede en la formación, la investigación interdisciplinar en España no parece estar especialmente favorecida, al menos en la práctica cotidiana. Al margen de los problemas burocráticos que suele plantear la colaboración entre departamentos, los propios mecanismos mediante los que se evalúa la investigación financiada con fondos públicos hacen posible que proyectos que en principio requerirían, por su temática, la colaboración de expertos procedentes de campos diversos, se lleven a cabo entre equipos de la misma especialidad (Llisterri 2004a)

Todos los factores considerados llevan a la separación entre las dos "culturas", la lingüística y la tecnológica, que intentábamos mostrar con algunos datos en el apartado 1. El resultado lo describe claramente van Santen (2005: 149) "...the phonetics community has not focused on questions most relevant for speech technology while the speech technology community has not developed algorithms and data structures that are optimally receptive for the incorporation of phonetic knowledge".

4. ALGUNAS PERSPECTIVAS DE FUTURO

Como mencionábamos anteriormente, algunas razones para abordar de nuevo la incorporación de conocimiento fonético son puramente prácticas, y responden a las limi-

taciones que se encuentran en los sistemas de reconocimiento de habla espontánea y en la falta de flexibilidad propia de la síntesis por selección de unidades.

En cuanto al reconocimiento, es preciso recordar, como hace Ainsworth (2005), que se obtienen buenos resultados en función de un corpus de entrenamiento de gran tamaño y de diccionarios de pronunciación que incorporan la variación documentada en el corpus, pero que la adaptación a nuevas situaciones no siempre se realiza con éxito. En la síntesis por selección de unidades, pese a su elevada naturalidad en dominios restringidos, la dependencia entre la fuente y el filtro reduce las posibilidades expresivas y prácticamente obliga a disponer de un nuevo corpus cada vez que se requiere una nueva voz o una nueva aplicación (Fant 2004).

Existen también áreas de investigación emergentes y cada vez más populares como la síntesis y el reconocimiento de las emociones, que probablemente no se están abordando de un modo completamente adecuado por falta de conocimiento lingüístico sobre la interacción comunicativa humana. Dada la dificultad de recopilar un corpus realista, se suele recurrir a actores y a un repertorio de emociones básicas que no parecen ser las que aparecen habitualmente en el habla, tal como recoge Campbell (2004: X): "... there was very little expression of the big-six emotions. Instead, there were a great variety of different speaking styles that changed as a consequence of listener and subject differences". Partiendo de estos datos relativamente artificiales, se aplican algoritmos de aprendizaje automático hasta encontrar el que es capaz de obtener mejores resultados en el reconocimiento. Es interesante destacar que mientras que las contribuciones sobre emociones en los últimos congresos de la Sociedad Española para el Procesamiento del Lenguaje Natural aparecen firmada por autores procedentes de departamentos de ingeniería de telecomunicación o de informática (Adell *et al.* 2005; Francisco *et al.* 2005; Luengo *et al.* 2005), en los tres congresos de Fonética Experimental no hemos podido encontrar ninguna comunicación que contenga la palabra *emoción* en el título.

Además de estos nuevos desarrollos, se plantean también en la actualidad una serie de tareas relacionadas con la mejora de las tecnologías existentes. Así, Batliner y Möbius (2005: 38-39) proponen una línea de investigación en prosodia que pueda resultar útil tanto para la síntesis como para el reconocimiento, basada en la realización de inventarios de funciones lingüísticas, paralingüísticas, léxicas y sintáctico/semánticas de la prosodia, en el diseño de un sistema de anotación motivado por consideraciones prácticas y orientado a la forma más que a la función, en el establecimiento de procedimientos para modelar rasgos prosódicos a partir de bases de datos que no representen necesariamente a un hablante específico y, finalmente, en el reconocimiento de que los parámetros prosódicos no pueden modelarse independientemente, ya que en el habla se producen de forma conjunta.

Por su parte, van Santen (2005: 162-163) establece un conjunto de tareas relevantes para la conversión de texto en habla: estudio de la percepción de las discontinuidades espectrales propias de la concatenación; percepción de las discontinuidades en los contornos melódicos; análisis de los aspectos subsegmentales en la organización temporal de la producción del habla; modelado de la reducción vocálica; estudio de la variación inter e intralocutor; determinación de los umbrales diferenciales en la percepción de

curvas melódicas; estudio de la percepción de las emociones generadas mediante síntesis; posibilidad de diseñar un modelo alternativo a ToBI para la descripción fonológica de la entonación; y, coincidiendo plenamente con Batliner y Möbius, análisis y modelado multidimensional de la interacción entre rasgos prosódicos.

Podríamos tal vez tener argumentos para pensar que nos encontramos en una etapa de transición en lo que se refiere a la relación entre la fonética y las tecnologías del habla. Por una parte, existe una gran cantidad de conocimiento fonético útil -paradójicamente "in part hidden in text-to-speech programs" (Fant 2002: 10)-, aunque no se encuentre en el formato adecuado; también siguen siendo válidas las aproximaciones basadas en reglas[4] y, si se dieran las condiciones favorables, el fonetista podría tener a su disposición corpus con los que realizar estudios específicamente adaptados a las necesidades de las tecnologías del habla, sin por ello renunciar a los modelos empleados en fonética; esta es, en parte, la perspectiva que sugiere Ainsworth (2005: 17): "the way to integrate phonetic knowledge into speech technology is not by deriving the detailed acoustic structure of phones from sets of phonetic rules, but by basing both speech recognition and speech synthesis on more realistic models of speech production. The details are probably best derived from speech databases as at present", coincidiendo con Barry *et al.* (2005: 11) en lo que se refiere al uso adecuado de los recursos lingüísticos.

Es factible, por tanto, realizar los necesarios avances en el estudio del habla como código siguiendo los objetivos tradicionales de la fonética, como propone Fant (2002: 10), y pensar con Greenberg (2005: 129) que "Over the coming decades this tension [entre la fonética y las tecnologías del habla] is likely to dissolve into a collaborative relationship melding linguistic knowledge with machine-learning and statistical methods as a means of developing mature science and technology".

Desde un punto de vista más práctico, investigadores como van Santen (2005: 164) proponen medidas concretas: incorporar materias relacionadas con la fonética y la lingüística en la formación de los tecnólogos y cursos de matemáticas, informática y procesamiento de señales en la de los fonetistas; organizar postgrados especializados en tecnologías del habla o licenciaturas conjuntas entre departamentos de lingüística, de ingeniería de telecomunicación y de informática para propiciar la formación de expertos en lo que se podría denominar "fonética computacional" (Moore 1995); y, por último, organizar congresos en los que, contrariamente a lo que ahora sucede, se programen sesiones en las que participen simultáneamente especialistas con formación humanística y tecnólogos.

No cabe duda de que la fonética puede y debe jugar un papel relevante en el desarrollo de las tecnologías del habla; si bien es cierto que existen obstáculos nada desdeñables, algunos de ellos pueden superarse apropiándose de nuevos enfoques y problemas, estableciendo mecanismos de interacción y coordinación que favorezcan la discusión en foros conjuntos y, especialmente, recordando a menudo las palabras de Fant con las que encabezábamos este trabajo.

---

4 Véase como muestra, la afirmación de Fant (2004: 9): "From detailed acoustic phonetic studies of text reading during the last 15 years, we have now been able to develop quite efficient prosodic rules for text-to-speech synthesis".

BIBLIOGRAFÍA[5]

ACERO, A. (1995) "The role of phoneticians in speech technology", in BLOOTHOOFT, G.- HAZAN, V.- HUBER, D.- LLISTERRI, J. (Eds.) *European Studies in Phonetics and Speech Communication*. Utrecht: OTS Publications. pp. 170-175.

http://research.microsoft.com/srg/papers/1995-alexac-esca.pdf

ADELL, J.- BONAFONTE, A.- ESCUDERO, D. (2005) "Analysis of prosodic features: towards modelling of emotional and pragmatic attributes of speech", *Procesamiento del Lenguaje Natural* 35: 277-283.

http://gps-tsc.upc.es/veu/research/pubs/download/Ade_Ana_05.pdf

AGÜERO, P.D.- BONAFONTE, A. (2003) "Phrase break prediction: a comparative study", *Procesamiento del Lenguaje Natural* 31: 107-114.

http://www.sepln.org/revistaSEPLN/revista/31/31-Pag107.pdf

AGUILAR, L.- GARRIDO, J.M.- LLISTERRI, J. (1997) "Incorporación de conocimientos fonéticos a las tecnologías del habla", in SERRA, E.- GALLARDO, B.- VEYRAT, M.- JORQUES, D.- ALCINA, A. (Eds.) *Panorama de la investigació lingüística a l'Estat Espanyol. Actes del I Congrés de Lingüística General. Volum III. Comunicacions: Fonètica i Fonologia. Semàntica i Pragmàtica.* València: Universitat de València. pp. 5-13.

http://liceu.uab.es/~joaquim/publicacions/valencia_94.html

AINSWORTH, W.A. (2005) "Can phonetic knowledge be used to improve the performance of speech recognisers and synthesisers?", in BARRY, W.J.- van DOMMELEN, W.A. (Eds.) *The Integration of Phonetic Knowledge in Speech Technology*. Dordrecht: Springer. pp. 13-20.

BARRY, W.J.- van DOMMELEN, W.- KOREMAN, J. (2005) "Phonetic knowledge in speech technology - and phonetic knowledge from speech technology?", in BARRY, W.J.- van DOMMELEN, W.A. (Eds.) *The Integration of Phonetic Knowledge in Speech Technology*. Dordrecht: Springer. pp. 1-12.

http://www.coli.uni-saarland.de/~koreman/Publications/2005/Eurospeech2001.pdf

BATLINER, A.- MÖBIUS, B. (2005) "Prosodic models, automatic speech understanding, and speech synthesis: Towards the common ground?", in BARRY, W.J.- van DOMMELEN, W.A. (Eds.) *The Integration of Phonetic Knowledge in Speech Technology*. Dordrecht: Springer. pp. 21-44.

BATTANER, E.- CARBÓ, C.- GIL, J.- LLISTERRI, J.- MACHUCA, M.J.- MADRIGAL, N.- MARRERO, V.- de la MOTA, C.- RIERA, M.- RÍOS, A. (2005a)

5 Las direcciones de Internet citadas en la bibliografía se han consultado en octubre de 2005.

"VILE: Estudio acústico de la variación inter e intralocutor en español", *Procesamiento del Lenguaje Natural* 35: 435-436.

http://liceu.uab.es/~joaquim/phonetics/VILE/VILE_SEPLN05.pdf

BATTANER, E.- CARBÓ, C.- GIL, J.- LLISTERRI, J.- MACHUCA, M.J.- MADRIGAL, N.- MARRERO, V.- de la MOTA, C.- RIERA, M.- RÍOS, A. (2005) "VILE: Estudio acústico de la variación inter e intralocutor en español", in *Actas del III Congreso de Fonética Experimental*. Universidade de Santiago de Compostela (en prensa).

http://liceu.uab.es/~joaquim/phonetics/VILE/VILE_IIICFE05.pdf

CAMPBELL, N. (2004) "Getting to the Hearth of the Matter: Speech is more than just the Expression of Text or Language", in *LREC 2004. Proceeedings of the 4th International Conference on Language Resources and Evaluation*. Paris: ELRA, European Language Resources Association. Vol. 5, pp. VII-X.

http://feast.atr.jp/nick/pubs/lrec-keynote.pdf

CARLSON, R.- GRANSTRÖM, B. (1991) "Speech synthesis development and phonetic research - a personal introduction", *Journal of Phonetics* 19, 1: 3-8.

http://www.speech.kth.se/~rolf/papers/wwjphonint.pdf

DUSAN, S.- RABINER, L.R. (2005) "On integrating insights from human speech perception into automatic speech recognition", in *EUROSPEECH 2005 - INTERSPEECH 2005. Proceedings of the 9th European Conference on Speech Communication and Technology*. pp. 1233-1236.

http://www.caip.rutgers.edu/~sdusan/2177anav.pdf

ESCUDERO, D.- CARDEÑOSO, V. (2001) "Modelo cuantitativo de entonación del español", *Procesamiento del Lenguaje Natural* 27: 233-240.

http://www.sepln.org/revistaSEPLN/revista/27/27-articulo27.pdf

ESCUDERO, D.- CARDEÑOSO, V. (2002) "Una experiencia en reconocimiento automático de tipos de unidades melódicas a partir de su perfil de entonación", in *Actas del II Congreso de Fonética Experimental*. Sevilla: Laboratorio de Fonética, Facultad de Filología, Universidad de Sevilla. pp. 161-166.

http://WWW.infor.uva.es/~descuder/investig/pdfs/cfeII.pdf

FANT, G. (1983) "Phonetics and Speech Technology", *Speech Transmission Laboratory - Quarterly Progress and Status Report* 2-3: 20-35; in *Proceedings of the 10th International Congress of Phonetic Sciences*. Dordrecht: Foris, 1984. pp. 13-24.

http://www.speech.kth.se/qpsr/pdf/1983/1983_24_2-3_020-035.pdf

FANT, G. (1989) "Speech Research in Perspective", *Speech, Music and Hearing - Quarterly Progress and Status Report* 30, 4:1-7; in *EUROSPEECH 1989. European Conference on Speech Communication and Technology*. Edinburgh: CEP Consultants Ltd. Vol 1, pp. 3-4.

http://www.speech.kth.se/qpsr/pdf/1989/1989_30_4_001-007.pdf

FANT, G. (1991) "What can basic research contribute to speech synthesis?", *Journal of Phonetics* 19, 1: 75-90.

FANT, G. (2004) "More than half a century in phonetics and speech research", in FANT, G. *Speech Acoustics and Phonetics*. Dordrecht: Kluwer. pp. 1-14.

http://www.speech.kth.se/~gunnar/halfcentury.pdf

FANT, G. (2005) "Historical notes", *Speech, Music and Hearing - Quarterly Progress and Status Report* 47: 9-19.

http://www.speech.kth.se/qpsr/tmh/2005/05-47-009-019.pdf

FERNÁNDEZ SALGADO, X.- RODRÍGUEZ BANGA, E. (2000) "Proposición de un marco adecuado para el estudio de contornos de F0 para síntesis de voz", *Procesamiento del Lenguaje Natural* 24: 175-182.

http://www.sepln.org/revistaSEPLN/revista/26/fernandez-salgado.pdf

FRANCISCO, V.- GERVÁS, P.- HERVÁS, R. (2005) "Análisis y síntesis de la expresión emocional en cuentos leídos en voz alta", *Procesamiento del Lenguaje Natural* 35: 293-300.

GARRIDO, J.M.- ORTÍN, I.- QUAZZA, S.- SALZA, P.L.- MANCINI, F. (2000) "Desarrollo de un módulo de asignación de parámetros prosódicos para la versión en español del sistema de conversión texto-habla ACTOR®", *Procesamiento del Lenguaje Natural* 24: 183-190.

http://www.sepln.org/revistaSEPLN/revista/26/garrido-alminana.pdf

GIL, J.- LLISTERRI, J. (2004) "Fonética y fonología del español en España (1978-2003)", *Lingüística Española Actual* 26, 2: 4-44.

http://liceu.uab.es/~joaquim/publicacions/Gil_Llisterri_LEA_2003.pdf

GONZÁLEZ REI, B.- CARDENAL LÓPEZ, A.- DOCÍO FERNÁNDEZ, L.- GARCÍA MATEO, C. (2001) "Problemática de la recogida y anotación de una base de datos oral para el gallego", *Procesamiento del Lenguaje Natural* 27: 37-44.

http://www.sepln.org/revistaSEPLN/revista/27/27-articulo4.pdf

GONZÁLEZ REI, B.- GARCÍA MATEO, C. (2000) "Diseño de una base de datos tipo SpeechDat para el idioma gallego", *Procesamiento del Lenguaje Natural* 26: 197-204.

http://www.sepln.org/revistaSEPLN/revista/26/gonzalez-rei.pdf

GREENBERG, S. (1998) "Recognition in a new key - Towards a science of spoken language", in *ICASSP 1998. Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing*. pp. 1401-1405.

http://www.icsi.berkeley.edu/~steveng/PDF/Recognition_in_a_New_Key.pdf

GREENBERG, S. (2001a) "From here tu utility - Melding phonetic insight with speech technology", in *Eurospeech 2001. Proceedings of the 7th European Conference on Speech Communication and Technology*. Vol 4. pp. 2485-2488.

http://www.icsi.berkeley.edu/ftp/global/pub/speech/papers/euro01-utility.pdf

GREENBERG, S. (2001b) "Whither Speech Technology? - A Twenty-First Century Perspective", in *Eurospeech 2001. Proceedings of the 7th European Conference on Speech Communication and Technology*. Vol 1, pp. 3-6.

http://www.icsi.berkeley.edu/ftp/global/pub/speech/papers/euro01-whither.pdf

GREENBERG, S. (2005) "From here to utility", in BARRY, W.J.- van DOMMELEN, W.A. (Eds.) *The Integration of Phonetic Knowledge in Speech Technology*. Dordrecht: Springer pp. 107-132.

http://www.icsi.berkeley.edu/%7Esteveng/PDF/Phonetic_Insight.pdf

HERNANDO, J.- PADRELL, J.- RODRÍGUEZ, H. (2002) "Sistema de información meteorológica automática por teléfono ATTEMPS", *Procesamiento del Lenguaje Natural* 29: 311-312.

http://www.sepln.org/revistaSEPLN/revista/29/29-Pag311.pdf

HUCKVALE, M. (2002) "Speech Synthesis, Speech Simulation and Speech Science", in *ICSLP 2002 - INTERSPEECH 2002. Proceedings of the 7th International Conference on Spoken Language Processing*. pp. 1261-1264.

http://www.phon.ucl.ac.uk/home/mark/papers/icslp02synth.pdf

LLISTERRI, J. (1990) "Algunes reflexions sobre el paper de la lingüística en la tecnologia de la veu", *Límits. Revista d'Assaig i d'Informació sobre les Ciències del Llenguatge* 8: 19-32.

http://liceu.uab.es/~joaquim/publicacions/llisterri_88.html

LLISTERRI, J. (2003a) "Las tecnologías del habla: Entre la ingeniería y la lingüística", in *Actas del Congreso Internacional "La Ciencia ante el Público. Cultura humanística y desarrollo científico y tecnológico"*. Salamanca: Instituto Universitario de Estudios de la Ciencia y la Tecnología. Edición en CD-ROM. pp. 44-67.

http://liceu.uab.es/~joaquim/publicacions/TecnolHab_Salamanca_02.pdf

LLISTERRI, J. (2003b) "Lingüística y tecnologías del lenguaje", *Lynx. Panorámica de Estudios Lingüísticos* 2: 9-71.

http://liceu.uab.es/~joaquim/publicacions/TecnoLing_Lynx02.pdf

LLISTERRI, J. (2004a) "Las tecnologías del habla para el español", in SEQUERA, R. (Ed.) *Ciencia, tecnología y lengua española: la terminología científica en español*. Madrid: Fundación Española para la Ciencia y la Tecnología. pp. 123-141.

http://liceu.uab.es/~joaquim/publicacions/TecnolHablaEsp_FECyT03.pdf

LLISTERRI, J. (2004b) "Las tecnologías lingüísticas en España", in *El español en el mundo. Anuario del Instituto Cervantes 2004*. Madrid: Instituto Cervantes – Círculo de Lectores – Plaza & Janés. pp. 229-251.

http://cvc.cervantes.es/obref/anuario/anuario_04/llisterri/default.htm

LLISTERRI, J.- CARBÓ, C.- MACHUCA, M. J.- de la MOTA, C.- RIERA, M.- RÍOS, A. (2003a) "El papel de la lingüística en el desarrollo de las tecnologías del habla",

in CASAS, M. (Dir.) - VARO, C. (Ed.) *VII Jornadas de Lingüística*. Cádiz: Servicio de Publicaciones de la Universidad de Cádiz. pp. 137-191.

http://liceu.uab.es/publicacions/Linguistica_TH_Cadiz02.pdf

LLISTERRI, J.- CARBÓ, C.- MACHUCA, M. J.- de la MOTA, C.- RIERA, M.- RÍOS, A. (2004) "La conversión de texto en habla: aspectos lingüísticos", in MARTÍ, M. A. - LLISTERRI, J. (Eds.) *Tecnologías del texto y del habla*. Barcelona. Edicions de la Universitat de Barcelona – Fundación Duques de Soria. pp. 145-186.

http://liceu.uab.es/publicacions/Linguistica_CTH_FDS02.pdf

LLISTERRI, J.- MACHUCA, M.J.- de la MOTA, C.- RIERA, M.- RÍOS, A. (2005) "Corpus orales para el desarrollo de las tecnologías del habla en español", *Oralia. Análisis del discurso oral* 8 (en prensa).

http://liceu.uab.es/~joaquim/publicacions/Oralia_04.pdf

LLISTERRI, J.- MACHUCA, M.J.- de la MOTA, C.- RIERA, M.- RÍOS, M. (2003b) "Entonación y tecnologías del habla", in PRIETO, P. (Ed.) *Teorías de la entonación*. Barcelona: Ariel. pp. 209-243.

http://liceu.uab.es/~joaquim/publicacions/Ariel_Aplicaciones.pdf

LUENGO, I.- NAVAS, E.- HERNÁEZ, I.- SÁNCHEZ, J. (2005) "Reconocimiento automático de emociones utilizando parámetros prosódicos", *Procesamiento del Lenguaje Natural* 35: 13-20.

MOORE, R. (1995) "Computational Phonetics", in *ICPhS 95. Proceedings of the XIIIth International Congress of Phonetic Sciences*. Vol 4, pp. 68-71.

NAVAS, E.- HERNÁEZ, I.- SÁNCHEZ, J.M. (2002) "Modelo de duración para la conversión de texto a voz en euskera", *Procesamiento del Lenguaje Natural* 29: 147-152.

http://www.sepln.org/revistaSEPLN/revista/29/29-Pag147.pdf

ÖHMAN, S. (2001) "Why current speech technology is false phonetics", *Working Papers (Lund University, Department of Linguistics)* 49: 180-183.

http://www.ling.lu.se/disseminations/pdf/49/bidrag46.pdf

PÉREZ, G.- LÓPEZ, T.- QUESADA, J.F. (2002) "Modelado de los candidatos seleccionados por un reconocedor de voz mediante técnicas de análisis gramatical", in *Actas del II Congreso de Fonética Experimental*. Sevilla: Laboratorio de Fonética, Facultad de Filología, Universidad de Sevilla. pp. 279-285.

POLS, L C.W. (1999) "Flexible, robust, and efficient human speech processing versus present-day speech technology", in *ICPhS 99. Proceedings of the 14th International Congress of Phonetic Sciences*. Vol 1, pp. 9-16.

http://www.fon.hum.uva.nl/pols/ICPHS'99.VS2.doc

POLS, L. (2001) "Acquiring and implementing phonetic knowledge", in *Eurospeech 2001. Proceedings of the 7th European Conference on Speech Communication and Techno-*

logy. Vol 1. pp. K3-K6; in *IFA Proceedings (Institute of Phonetic Sciences, University of Amsterdam)* 24: 39-46.

http://www.fon.hum.uva.nl/Proceedings/Proceedings24/Proc24Pols_corrton.html

RODRÍGUEZ BANGA, E.- CAMPILLO DÍAZ, F.- FERNÁNDEZ REI, E.- MÉNDEZ PAZÓ, F. (2002) "Sistema de conversión texto-voz en lengua gallega basado en la selección combinada de unidades acústicas y prosódicas", *Procesamiento del Lenguaje Natural* 29: 153-158.

http://www.sepln.org/revistaSEPLN/revista/29/29-Pag153.pdf

RODRÍGUEZ, H. (2004) "Lingüística y estadística, ¿incompatibles?", in MARTÍ, M. A. – LLISTERRI, J. (Eds.) *Tecnologías del texto y del habla*. Barcelona. Edicions de la Universitat de Barcelona – Fundación Duques de Soria. pp. 89-117.

ROSSI, M. (1996) "Connaissances et traitement automatique de la parole", in MÉLONI, H. (Coord.) *Fondements et Perspectives en Traitement Automatique de la Parole*. Paris: Éditions AUPELF-UREF. pp. 19-30.

http://www.bibliotheque.refer.org/html/parole/rossi/rossi.htm

RUBIO, A.- HERNÁEZ. I. (2005) *Libro blanco de Tecnologías del Habla*. Granada: Universidad de Granada - Red Temática en Tecnologías del Habla.

http://www.rthabla.org/LibroBlancoTecnologiasDelHabla.pdf

SANTEN, J.P.H. van (2005) "Phonetic knowledge in text-to-speech synthesis", in BARRY, W.J.- van DOMMELEN, W.A. (Eds.) *The Integration of Phonetic Knowledge in Speech Technology*. Dordrecht: Springer. pp. 149-166.

SHRIBERG, E. (2005) "Spontaneous Speech: How People Really Talk and Why Engineers Should Care", in *EUROSPEECH 2005 - INTERSPEECH 2005. Proceedings of the 9th European Conference on Speech Communication and Technology*. pp. 1781-1784.

http://www.speech.sri.com/papers/eurospeech2005-shriberg-keynote2005-10-22

STRIK, H. (2005) "Is phonetic knowledge of any use for speech technology?", in BARRY, W.J.- van DOMMELEN, W.A. (Eds.) *The Integration of Phonetic Knowledge in Speech Technology*. Dordrecht: Springer. pp. 167-180.

http://lands.let.kun.nl/literature/strik.2005.1.pdf

LA INTERDEPENDENCIA ENTRE ACENTO LÉXICO Y ACENTO TONAL EN LAS FRASES DEL ESPAÑOL

EUGENIO MARTÍNEZ CELDRÁN
martinezceldran@ub.edu
Universidad de Barcelona

1. PRESENTACIÓN[1]

> "En la simple enunciación de la palabra aislada, fuera de relación sintáctica, el tono sigue a la intensidad con regular correspondencia. La sílaba fuerte de la palabra es al mismo tiempo la sílaba tónica, es decir, la que se pronuncia con mayor altura de tono [...] dentro de la frase, la palabra no tiene entonación propia" (T. Navarro Tomás, 1944: 21-22).

Hemos comenzado con esa cita en la que ya se establecen algunos de los puntos que vamos a tratar. En primer lugar, el acento y el tono tienen una relación muy estrecha: la sílaba fuerte es la que mayor altura de tono posee. En segundo lugar, hay que indicar que no es lo mismo considerar las palabras de forma aislada que concatenadas en una frase; en esta segunda situación los tonos de las sílabas de la palabra se someten a la entonación de la frase.

Hay que señalar algunos cambios modernos en la consideración del acento. Todavía Navarro Tomás hace referencia a la intensidad como principal parámetro del acento. Desde este punto de vista, son muchos los autores que han descartado la intensidad como la principal responsable del fenómeno acentual. En la actualidad, en palabras aisladas se ha defendido tanto la F0 como la duración, como los parámetros principales del acento léxico (Quilis 1971, Canellada y Madsen 1978, Solé 1984, Llisterri *et alii* 2003, etc.). El mismo Navarro Tomás ya dice en su cita que la sílaba fuerte es a la vez la sílaba tónica, es decir, la de mayor altura tonal. Por tanto, admitía que el tono

1 Este trabajo ha contado con un ayuda de la SGPI del MCYT, ref. BFF2003-08487

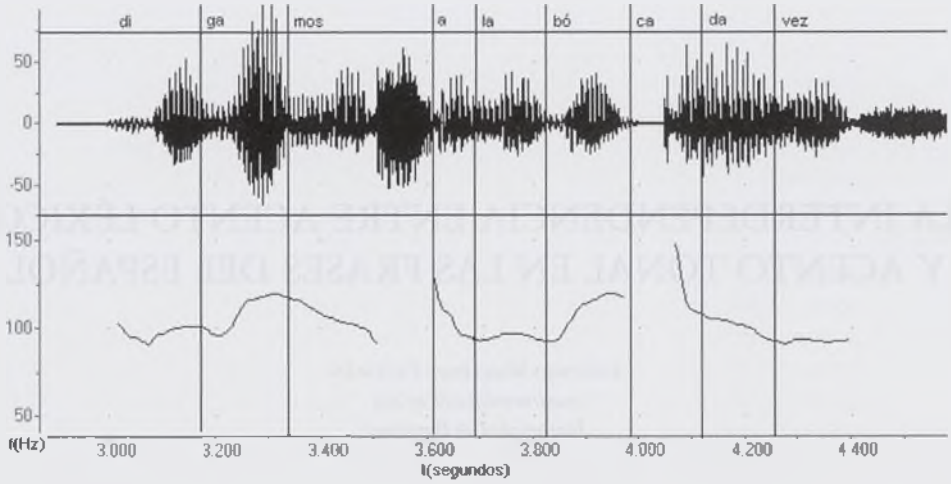


Figura 1. Curva de F0 y oscilograma de la frase con la palabra aguda

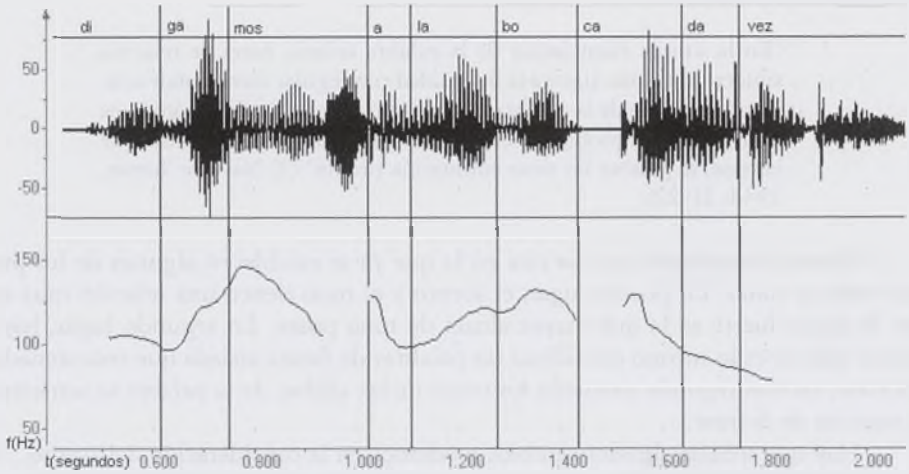


Figura 2. Curva de F0 y oscilograma con la palabra llana.

era también un parámetro importante en la detección del acento. Los autores citados lo confirman y añaden que la duración es el segundo parámetro más relevante. Descartamos, pues, la intensidad como parámetro determinante. Sirva de ejemplo las frases de las figuras 1 y 2.

En estas figuras hemos utilizado frases con sonidos sonoros para obtener una curva casi continua, pero delimitando la palabra, que queremos estudiar, con sonidos sordos.

Como se ve, la sílaba fuerte en la palabra aguda *alabó* posee un valle en su inicio y a partir de él comienza una ascensión que llega hasta la sílaba postónica. La diferencia entre el valle y el pico de la postónica es de 52 Hz (98-150). También hay diferencias por la duración; la sílaba *–bó-* mide 157 ms y la precedente y siguiente miden unas 129 ms. Esta diferencia (28 ms) supera el umbral diferencial de la percepción de la duración, ya que en se suele considerar 20 ms dicho umbral (Roulet 1999); aunque mucho más exigentes son Pamies y Fernández (2004) que el umbral lo colocan con diferencias de un tercio, lo cual implicaría diferencias mayores de 50 ms en este caso, según estos autores la diferencia estaría bajo umbral. La intensidad ya se ve que es aproximadamente la misma por las alturas representadas en el oscilograma. En el caso de la palabra llana *alabo*, nuevamente hay diferencias evidentes entre el valle (100 Hz) y el pico situado en la sílaba postónica (130 Hz) nuevamente. La duración es sólo un poco mayor en la sílaba tónica (169 ms) que en las átonas que la rodean (150 ms). La diferencia de 19 ms no supera ninguno de los dos criterios sobre el umbral de la percepción. En este caso la intensidad sí es un poco mayor en la sílaba tónica. La diferencia principal es siempre la del F0 ya que en ambas palabras se supera con mucho el umbral de la percepción situado en 1.5 st (Rietveld y Gussenhoven 1985) tomando como elemento principal el ascenso desde el valle situado en las inmediaciones del comienzo de la sílaba tónica y el pico situado en la postónica.

Cuando las palabras se pronuncian aisladas se puede constatar que los tres parámetros: F0, duración e intensidad, se destacan con diferencias notables, como se puede apreciar en la figura 3 (D=débil, F=fuerte; Lg=larga, B=breve; L=low (baja), H=high (alta)), cosa que no sucede dentro de la frase, donde uno de los tres se destaca por encima de los demás. Además, hay una diferencia importante, el pico aparece dentro de la sílaba tónica.

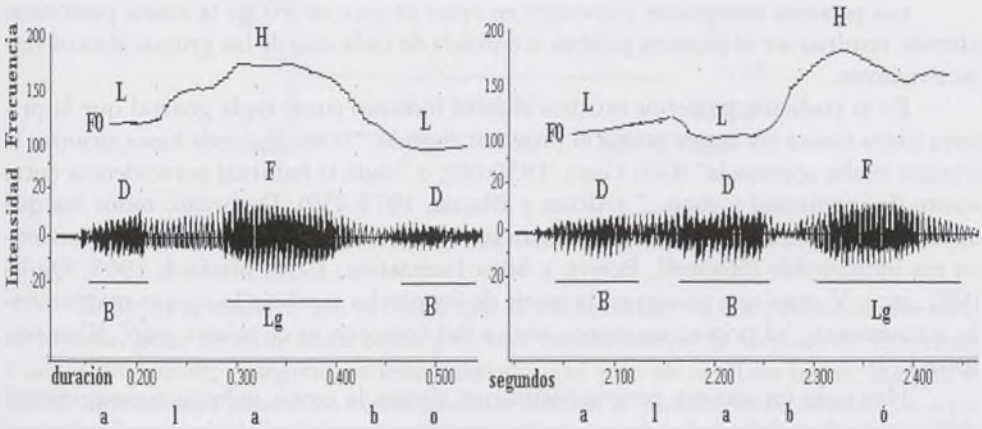


Figura 3. Parámetros del acento en palabras aisladas.

Así, pues, podemos concluir que Navarro Tomás está en lo cierto cuando afirma que la sílaba fuerte posee mayor tono en las palabras aisladas y también que posee mayor intensidad, pero que se somete a la entonación general de la frase. No habla para nada de la duración como parámetro que soporta la prominencia silábica en la misma frase, sino de la intensidad, pero actualmente se le da la relevancia tanto al tono como a la duración, según las circunstancias, por lo que el acento no desaparece. El hablante seguirá teniendo conciencia de la presencia del acento en las palabras dentro de una frase.

Navarro Tomás sigue afirmando que "de un modo general en la serie sucesiva de los núcleos silábicos de una frase, el tono en español se eleva más o menos sensiblemente en las sílabas fuertes y desciende en las inacentuadas (1944: 23)". Y añade que "no se trata de una correspondencia que se produzca con una regularidad invariable..." (Ibíd.). Cuando este autor expone el esquema general de la entonación enunciativa presta bastante relevancia a la inflexión inicial: "la inflexión ascendente de la voz desde la sílaba inicial débil a la acentuada..." (p. 46). Más adelante indicará que "la nota alcanzada en la primera sílaba fuerte apenas experimenta modificaciones importantes en las sílabas siguientes..." (p. 48); pero una página más adelante (49) dirá: "Es frecuente que la sílaba débil que sigue inmediatamente a la primera acentuada resulte algo más alta que ésta, con diferencia de uno o dos semitonos."

Se puede decir que, por regla general, Navarro Tomás considera que hay una correspondencia entre primera sílaba acentuada y pico de F0, aunque constata que muchas veces la postónica es más alta que la misma tónica. El mismo proporciona el siguiente ejemplo tomado de un texto literario de Valle Inclán:

Los niños fueron a sentarse || en las vradas del altar;
Sus vestidos eran albos || como el lino de los paños litúrgicos.

Las palabras subrayadas coinciden en tener el pico de F0 en la sílaba postónica; además, resultan ser la primera palabra acentuada de cada uno de los grupos fónicos respectivamente.

En la tradición posterior muchos autores tomaron como regla general que la primera sílaba tónica era la que poseía el pico más elevado: "la voz asciende hasta alcanzar la primera sílaba acentuada" (Gili Gaya, 1950:60); o "dada la habitual coincidencia entre acento de intensidad y tono..." (Alcina y Blecua, 1975:459). De hecho, todos los que siguieron la teoría de los niveles de la escuela americana suponían que esa correspondencia era indiscutible (Stocwell, Bowen y Silva Fuenzalida, 1956; Matluck 1965; Quilis 1981; etc.). Y otros que no siguen la teoría de los niveles también lo siguen manteniendo actualmente: "el primer segmento tónico del contorno es el primer pico" (Cantero, 2002:172).

Han sido los autores generativistas que siguen la teoría métrica autosegmental (AM) los que han defendido que en español la estructura tonal o "pitch accent" principal de la zona prenuclear es L*+H (D'Introno *et alii* 1995; Sosa, 1995; Prieto, 1995; Face, 2003, etc...). Es decir, la sílaba fuerte es baja [L*] y la que sigue (postónica) es alta [H].

Por ejemplo D'Introno y Alt. (1995: 439-442) proporcionan la siguiente fórmula para la enunciativa no marcada del español:

$$\{L\% -\{L^*+H\}^n\}-L^*L\%$$

L% significa en ambos casos las junturas inicial y terminal que delimitan la frase. L* anterior a la juntura terminal constituye lo que se denomina el núcleo de la frase en la tradición británica, que en la fonología española Navarro Tomás denominó el tonema, y la fórmula entre corchetes indica que cada uno de los acentos tonales anteriores al tonema (uno o más), en la parte que se suele denominar prenuclear o pretonemática, tendrá la estructura siguiente: inflexión ascendente en la sílaba fuerte que culmina con un pico en la sílaba siguiente que es la postónica. Como se ve, con esta fórmula estos autores están seguros de que en español la estructura del prenúcleo siempre tiene la estructura L*+H. Ellos la ejemplifican con una frase con una curva de F0 dibujada por ellos, pero yo prefiero exponerla con la pronunciación real y con la intención de reproducir su esquema:

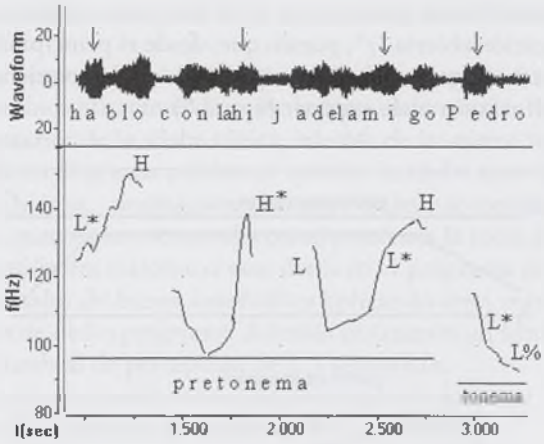


Figura 4. Correspondencia de los acentos tonales con los acentos léxicos(marcados con flechas)

La figura 4 muestra que es cierto que la estructura L*+H es predominante en el pretonema, pero no es la única como pretende hacernos creer la fórmula de D'Introno *et alii*. Como se ve, el segundo acento tonal tiene el pico en la sílaba fuerte, aunque es posible que se haya puesto un cierto grado de énfasis al pronunciarla. Muchos de estos autores generativistas se remontan a Navarro Tomás para indicar que esa es la estructura del pretonema; pero este autor sólo dijo que era frecuente y en sus propios ejemplos aparece básicamente en el primer acento tonal de la frase, no en los demás.

Todo lo dicho hasta ahora se refería básicamente a la entonación de frases enunciativas. Nuevamente, fue Navarro Tomás el que dio unas pautas del comportamiento de las interrogativas absolutas comparándolas con las enunciativas neutras en lo que al pretonema se refiere, pues el tonema es totalmente contrario. Podríamos resumirlas en cuatro puntos (1944: 98):

1° Las sílabas débiles iniciales comienzan un movimiento ascendente y, desde el principio, parten de una nota más alta.
2° En la primera sílaba acentuada la voz se eleva unos semitonos sobre la altura media de la enunciación normal. En las sílabas siguientes se produce gradualmente un descenso más o menos amplio.
3° El nivel de la pregunta representa una base tónica más alta relativamente que la de la enunciación ordinaria.
4° La entonación enunciativa y la interrogativa se distinguen entre sí por la diferente disposición de sus líneas melódicas: horizontal en la primera y oblicua en la segunda.

Por eso Navarro Tomás justifica que en la ortografía española se exija empezar con el signo de interrogación abierta "¿", puesto que, desde el principio, hay diferencias entre la enunciativa y la interrogativa. Todo lo expuesto lo podemos ejemplificar mediante las curvas típicas que él mismo suele exponer (figura 5).

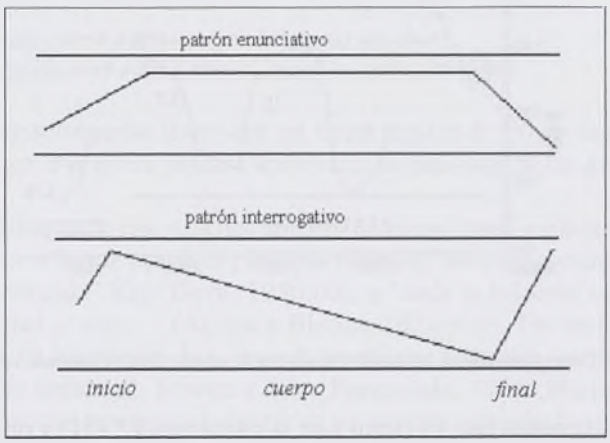


Figura 5. Curvas prototípicas de una enunciativa neutra y una interrogativa absoluta, según Navarro Tomás.

En este caso Navarro no detecta que haya desajuste entre el acento y el pico de F0 pues, además de lo dicho con anterioridad, añade más adelante que "la sílaba débil que sigue a la primera acentuada presenta el mismo tono que ésta, y las demás sílabas débiles

se acomodan a la línea de descenso marcada por la acentuada. Pero Sosa (1999:210) ponía de nuevo de manifiesto que en las interrogativas "el pretonema consiste en acentos tonales L*+H y el tonema de anticadencia L*+H%". También Cantero (2002:171) indica que "(del primer segmento del contorno /+interrogativo/) parte una inflexión ascendente, cuyo segundo segmento (átono) es el primer pico del contorno"; por tanto, los autores posteriores no dan a Navarro Tomás la razón en la correspondencia de primera sílaba tónica y primer pico de F0. Sí se le suele dar la razón en todo lo demás que la caracteriza.

Después de todo lo expuesto pensamos (Martínez Celdrán; Fernández Planas y Fullana Rivera, 2003) realizar un experimento con el que pudiéramos explorar estos problemas y sobre todo la correspondencia de los acentos tonales y los acentos léxicos o, su equivalente, picos de F0 y sílabas fuertes. Nuestra hipótesis partía del hecho de que los ejemplos utilizados por Navarro Tomás o por los generativistas tenían sobre todo palabras llanas; por tanto, pensamos investigar si podía haber diferencia por el tipo de palabra. Es decir, la hipótesis era que si el tipo de palabra cambiaba a palabras agudas no existiría de manera tan frecuente un desplazamiento del pico de F0 hacia la postónica.

2. MÉTODO

2.1. Puntos de análisis

Determinamos obtener el valor del F0 en Hz del centro de las vocales solamente, para despreciar los cambios micromelódicos de la frase. Además sólo se analizaron las posiciones anterior y posterior de la sílaba tónica, además de la misma tónica; de modo que si esa sílaba no existía en la misma palabra se tomaba la sílaba anterior o posterior de la palabra contigua: en...*la mesa*... se analizaban las tres vocales y se tomaba como pretónica la vocal del artículo; en...*mansión se*... se tomaba como postónica la vocal del *se*.

Nuestra pretensión era introducir esos datos en el programa de estadística SPSS y obtener los acentos tonales de forma automática aplicando unas reglas precisas programadas con las sintaxis de dicho programa. Además utilizamos un filtro para la definición de las estructuras: el umbral de percepción de 1.5 semitonos.

2.2. Acentos tonales utilizados en el modelo AM (Ladd 1996)

- L*+H: ascenso en la tónica con pico en la postónica
- L+H*: ascenso en la pretónica con pico en la tónica
- H*+L: pico en la tónica con descenso hacia la postónica
- H+L*: pico en la pretónica con descenso en la tónica
- H*: pico en tónica sin diferencias con las sílabas débiles vecinas o con diferencias bajo umbral.
- L*: la tónica muestra un descenso que será aún mayor en la postónica ya que sólo puede existir en el tonema descendente.

Los acentos bitonales poseen siempre diferencias por encima de 1.5 semitonos. Cuando las tres sílabas poseen diferencias superiores a 1.5 semitonos entonces se coloca

L a la más baja de todas, siendo H las demás; por ejemplo, (H+L*)+L, con valores en Hz 145 – 130* – 110, se analiza como H*+L, pues la diferencia de la pretónica con la tónica es menor (1.89 st) que entre la tónica y la postónica (2.89 st); en cambio, (L+H*)+H con valores en Hz 110 – 120* – 145 sería L*+H, pues ahora la diferencia entre pretónica y tónica es de 1.5 st, mientras que la tónica se diferencia de la postónica en 3.27 st.

2.3.Estímulos e informantes

Se grabó un corpus que consistía en cuatro bloques de ocho frases cada uno. Cada frase estaba compuesta por palabras llanas (bloque 1) y agudas (bloques 2): por ejemplo, *las casonas estaban en la colina, las manzanas estaban en la mesita, etc.*, como representantes de las llanas; *la mansión se dibujó sobre el azul, el peral se destacó sobre el jardín, etc.*, donde todas las palabras acentuadas eran agudas. Además de estos dos bloques, se incluyeron otros dos más añadiendo un adjetivo a cada frase, para que funcionara como expansión del SN sujeto (bloques 3 y 4): *las casonas **marinas** estaban en la colina, la mansión **añil** se dibujó sobre el azul...*; por tanto, había un total de 32 frases. 16 de las cuales tenían tres estructuras tonales y las otras 16 tenían cuatro. Este hecho determina 112 acentos tonales por hablante que por cuatro hablantes (dos masculinos y dos femeninos) resultan 448 acentos tonales; por dos repeticiones, resultan ser 896 en las enunciativas y otras tantas en las interrogativas, ya que las mismas frases se pronunciaron en la modalidad interrogativa absoluta por otros cuatro hablantes diferentes. En cada una de las modalidades hubo dos voces masculinas y dos femeninas.

En este estudio hemos excluido las palabras esdrújulas, pero eso no afecta demasiado a nuestra hipótesis. Hay que tener en cuenta que las palabras llanas representan en el vocabulario español un 79,50% y las agudas un 17,75%. Las esdrújulas sólo son un 2,75%, según datos de A. Quilis (1981); por tanto, éstas últimas tienen un porcentaje bajísimo por lo que podemos prescindir de ellas y comprobar nuestra hipótesis con sólo las dos primeras.

Los cuatro informantes hablaban español peninsular estándar. Con ello queremos decir que no tenían ningún acento regional marcado: andaluz, catalán, aragonés, etc.

El programa utilizado fue el Speech Analyzer v. 1.5 del SIL.

3. RESULTADOS

Presentamos en dos tablas los resultados referentes a la zona prenuclear o pretonemática: la primera está referida a las frases enunciativas y, la segunda, a las interrogativas. Ofrecemos la frecuencia de aparición de los acentos tonales según el tipo de palabra y teniendo en cuenta en qué posición de la frase aparecen: primera, segunda o tercera. Se ha destacado la frecuencia mayor en negrita:

Acento tonal	Palabras agudas (%) 1ª - 2ª - 3ª	Palabras llanas (%) 1ª - 2ª - 3ª
1. H*	6.3 32.8 20.3	3.9 21.9 48.4
2. L+H*	71.1 35.9 45.3	17.2 6.3 3.1
3. L*+H	21.1 16.4 34.4	77.3 41.3 15.6
4. H+L*	3.1	30.5 32.8
5. H*+L	1.6 11.7	1.6

Tabla I Frecuencia de los acentos tonales en el pretonema de las frases enunciativas

En primer lugar, se ve claramente que en las agudas y en el primer acento la estructura destacada es L+H*; por tanto, el pico coincide sensiblemente con la sílaba fuerte, pues en el 71,1% de los casos obtenemos esta estructura. En las demás posiciones sigue siendo esta estructura la más numerosa pero se reparten más las frecuencias en otras estructuras. En las llanas, sucede lo contrario, el pico aparece en la postónica por lo que la estructura del primer acento es L*+H en un 77.3% de los casos, como ya habían defendido los generativistas. Sigue siendo en la segunda posición la estructura más frecuente pero ya se reparten más las frecuencias hacia otras estructuras, siendo H* la predominante en la tercera posición. No obstante, la primera conclusión es que no se puede afirmar que la única estructura del pretonema sea L*+H, como pretendían D'Introno *et alii* (1995) y Sosa (1995). Nuestros datos permiten matizar mejor y ofrecen la variedad de acentos existente teniendo en cuenta tanto el tipo de palabra como el orden en que aparece el acento tonal en la frase.

Ofrecemos a continuación dos gráficos (figuras 6 y 7) con frases donde se muestran los acentos más frecuentes que aparecen en la tabla I.

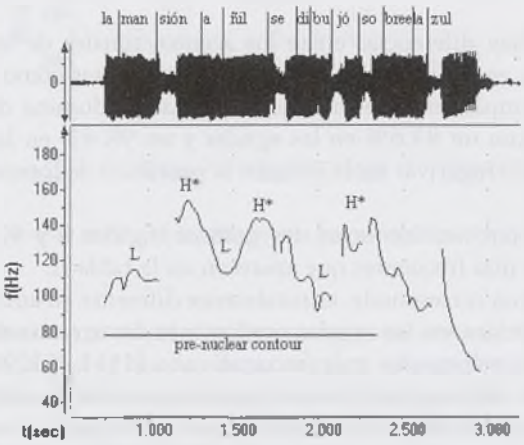


Figura 6. La curva de F0 y oscilograma de una frase enunciativa con palabras agudas

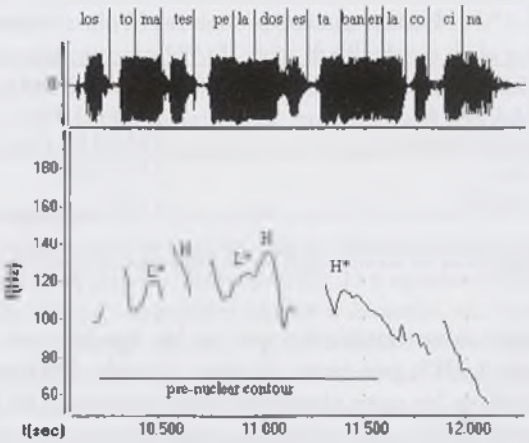


Figura 7. La curva de F0 y oscilograma de una frase enunciativa con palabras llanas.

Acento tonal	Palabras agudas (%) 1ª - 2ª - 3ª	Palabras llanas (%) 1ª - 2ª - 3ª
1. H*	0.8 40.6 76.6	0.8 14.1 50
2. L+H*	14.1 3.1 12.5	0.8
3. L*+H	83.6 14.1 3.1	98.4 24.2
4. H+L*	0.8 30.5 1.6	28.9 25
5. H*+L	0.8 11.7 6.3	32.8 25

Tabla II Frecuencia de los acentos tonales en el pretonema de las frases interrogativas

Nuevamente hay diferencias entre los acentos tonales de las palabras agudas y de las palabras llanas en las interrogativas, aunque no tantas como se apreciaban en las enunciativas; por ejemplo, en el primer acento tonal predomina de forma mayoritaria la estructura L*+H con un 83.6% en las agudas y un 98.4% en las llanas. Es decir, el primer pico de las interrogativas suele estar en la postónica de forma casi exclusiva, dado su alto porcentaje.

Ofrecemos a continuación otros dos gráficos (figuras 8 y 9) con frases donde se muestran los acentos más frecuentes que aparecen en la tabla II.

El tonema, como corresponde, es totalmente diferente en ambas modalidades descritas: en las enunciativas, en las agudas predominan los acentos tonales L* (64.8%) y H+L* (24.2%) y en las llanas los más frecuentes son H*+L (52.3%) y L* (38.3%). Es lógico el predominio del tono bajo puesto que la característica principal de esta modalidad es el descenso de las últimas sílabas. En las interrogativas, en el tipo de palabra aguda los acentos tonales predominantes son L+H* (55.5%) y H* (43%) y en el tipo de palabra llana L*+H (93.8%). El ascenso final conlleva que los acentos acaben en un tono

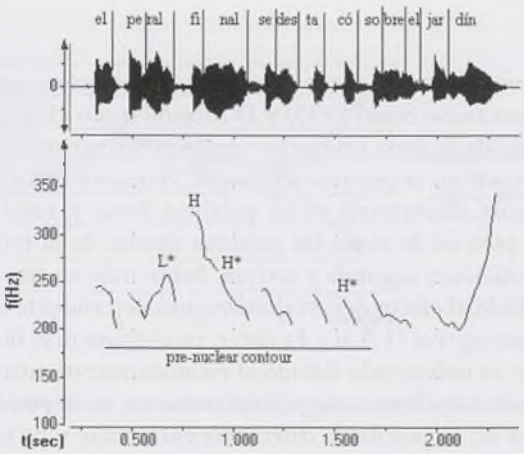


Figura8. La curva de F0 y oscilograma de una frase interrogativa con palabras agudas.

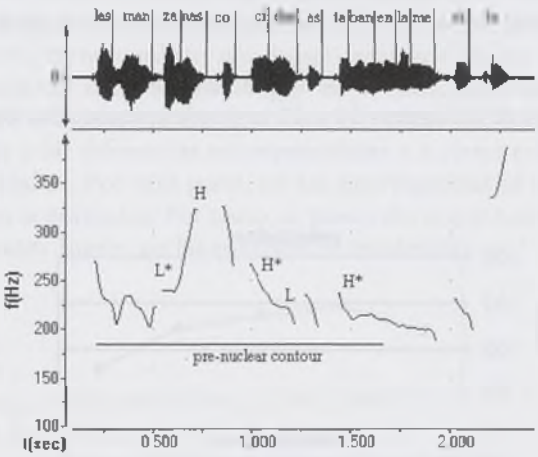


Figura 9. La curva de F0 y oscilograma de una frase interrogativa con palabras llanas.

alto lógicamente. Por otra parte, el distinto comportamiento de agudas y llanas al final de la frase se debe al hecho de que las llanas siempre poseen una sílaba débil antes de la pausa, mientras que las agudas acaban en sílaba fuerte; por tanto, los acentos tonales en esa posición han de ser diferentes; es imposible la estructura L*+H en las interrogativas con tonema acabado en palabra aguda, como el H*+L en las enunciativas en ese mismo contexto porque falta la sílaba débil que permite dicha estructura.

4. DISCUSIÓN

Como se desprende de los datos expuestos, parece claro que no son exactas las afirmaciones de autores como Sosa (1995) y D'Introno *et alii* (1995) que pretendían que cualquier acento tonal de la zona prenuclear tenía la estructura L*+H; es decir, que el pico de F0 estaba situado en la postónica siempre. Nuestros datos indican que efectivamente es una estructura mayoritaria en las palabras llanas y, sobre todo, en la primera posición de la frase; pero no lo es en las palabras agudas de la modalidad enunciativa. Ni tampoco en las posiciones segunda y tercera. Sobre todo en esta última el acento que predomina es H* debido al efecto del escalonamiento descendente (inglés "downstep") y el filtro del umbral perceptivo (1.5 st). Es decir, en el transcurso de la frase la diferencia entre valles y picos se va reduciendo debido al escalonamiento mencionado por lo que el filtro de 1.5 st pone de manifiesto que perceptivamente en la posición inmediatamente anterior al pretonema no se percibe la diferencia entre valle y pico; por ejemplo, en las interrogativas y para la voz masculina, hemos obtenido los promedios siguientes en las postónicas:

149 – 125 – 109 Hz

En las enunciativas, la pendiente no es tan grande (5% vs. 9% aproximadamente); por este motivo la frecuencia de H* es menor. Los datos de los picos en las enunciativas son:

146 – 137 – 124 Hz

En los gráficos de las figuras 10 y 11 se puede apreciar este fenómeno que comentamos:

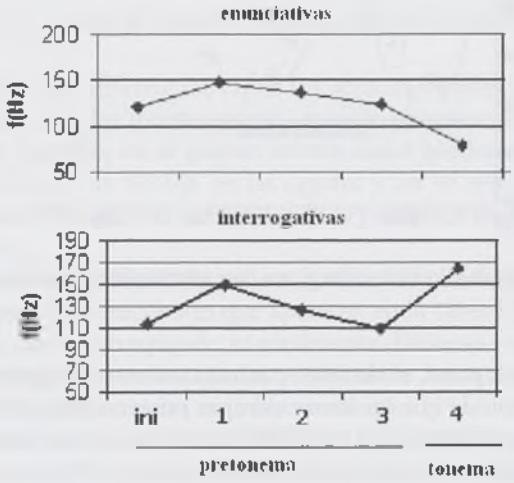


Figura 10. Modelos de curvas siguiendo la frecuencia en Hz de la sílaba postónica

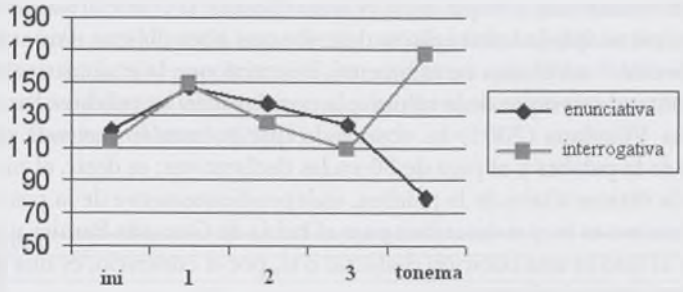


Figura 11. Curvas superpuestas que muestran las diferencias de perfil según la modalidad oracional

Si comparamos la figura 5 con las figuras 10 y 11, podremos observar hasta qué punto Navarro Tomás llevaba razón. En lo que se refiere a lo que él denominaba el cuerpo de la frase, él indicaba que en la enunciativa era prácticamente horizontal, mientras que en la interrogativa era oblicuo. Ciertamente el cuerpo de la interrogativa está más inclinado, hemos calculado un 9% de pendiente, pero el cuerpo de la enunciativa no es exactamente horizontal tiene también una ligera inclinación cuya pendiente es de un 5%, bien es cierto que este autor reconoció que "no se trata, por supuesto, de una igualdad de tono que deba imaginarse como una línea estrictamente horizontal" (1944: 49), pero esto lo achacaba a las diferencias microprosódicas y a cierto énfasis que se pudiera colocar en algunas sílabas. Por otra parte, en las interrogativas el primer pico se sitúa abrumadoramente en la postónica. Por tanto, se puede decir que nuestros datos afinan las afirmaciones de Navarro Tomás, no las contradicen totalmente.

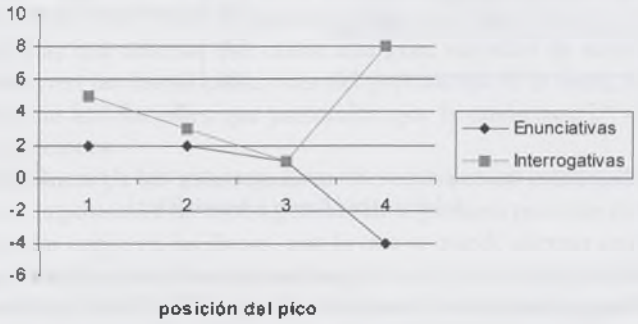


Figura 12. Diferencias entre tónica y postónica en las medias de los semitonos de las palabras llanas según la posición en la frase

Si se atiende a la diferencia en semitonos de tónica y postónica (figura 12), obsérvese que nos aproximamos más a lo que decía Navarro Tomás: la enunciativa tiene el cuerpo casi horizontal, mientras que la interrogativa describe una línea oblicua. Por otra parte, la interrogativa asciende 7 semitonos en el tonema, mientras que la enunciativa desciende cinco semitonos, contando siempre de la tónica a la postónica en las palabras llanas.

Estebas Vilaplana (2003) ha observado que en catalán hay una correspondencia entre el final de la palabra y el pico de F0 en las declarativas; es decir, el pico de F0 tiende a situarse en la última sílaba de la palabra, independientemente de la posición del acento léxico. Esto mismo es lo que describen para el habla de Granada Pamies y Amorós (2005) y se plantean si esto es una cuestión dialectal o si, por el contrario, es una peculiaridad de todo el español (véase la figura 13).

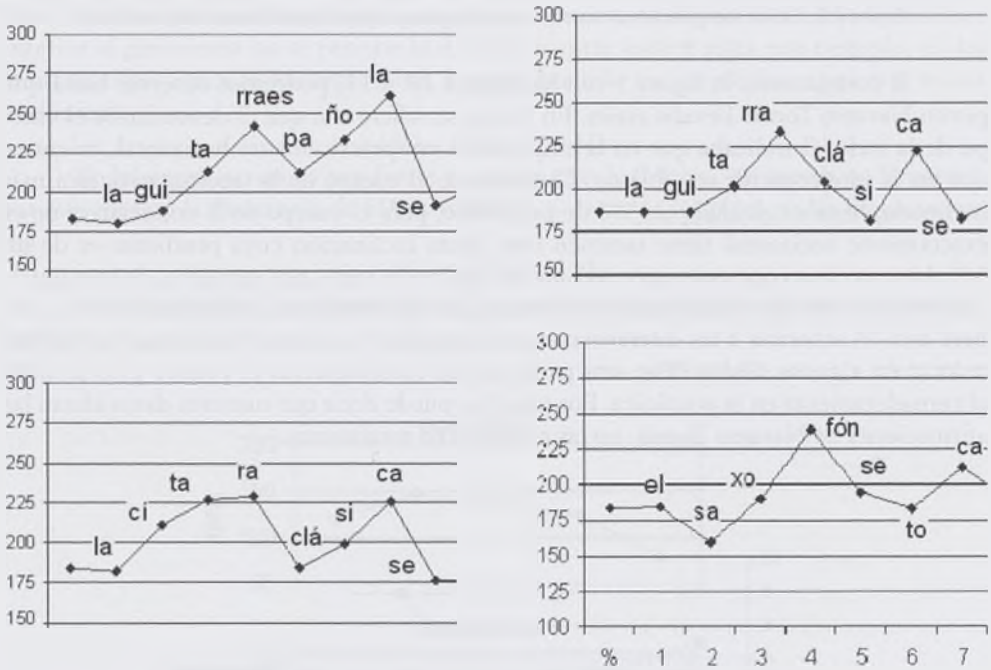


Figura 13. En los gráficos se muestran diversas estructuras acentuales y cómo el pico se desplaza hasta el final de la palabra (Apud Pamies y Amorós 2005).

En nuestros datos sobre esta lengua parece que hay la misma tendencia en las declarativas: en las agudas existe esta coincidencia en el 78.9% de la primera posición; en el 80.5% de la segunda y en el 75,6% en la tercera. En las llanas, coincide con la estructura L*+H: 77,3% en la primera posición; 41.3% en la segunda y 15.6% en la tercera. En las llanas, pues, no es tan llamativa esta coincidencia.

En este trabajo no hemos prestado atención a las esdrújulas, pero obsérvese la figura 14, tomada de Sosa (1999:195), donde en la palabra *número*, que aparece en la segunda posición, el tono sigue ascendiendo hasta la última sílaba de la palabra. Según el análisis de este autor L* va en la antepenúltima sílaba, mientras que H la coloca en la última, siendo la sílaba intermedia un sílaba de transición. Esto coincide con las afirmaciones de Estebas (2003) y Pamies y Amorós (2005).

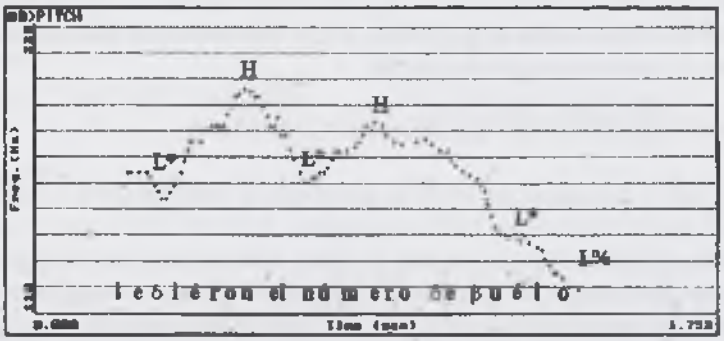


Figura 14. Palabra esdrújula con H en la última sílaba.

5. CONCLUSIONES PARA EL HABLA DE LABORATORIO

En muchos aspectos, este estudio viene a precisar y a matizar las afirmaciones de Navarro Tomás:

1º Coincidimos con Navarro Tomás en que es frecuente el pico de F0 en la postónica en las enunciativas, sobre todo en la primera posición de la frase y en las palabras llanas, que estadísticamente son las más abundantes en español. Sin embargo, hay que afirmar que existe una gran variedad de acentos tonales en las agudas y en las demás posiciones del pretonema de la frase, lo cual contradice todas las afirmaciones que pretenden que la estructura L*+H es la única en esas posiciones.

2º No coincidimos en las interrogativas. Resulta que en esta modalidad el pico aparece en la postónica de modo general en la primera posición de la frase tanto en las agudas como en las llanas, por lo que se puede afirmar que el fenómeno es más general que en las enunciativas.

3º La enunciativa posee un cuerpo casi horizontal, al menos se puede afirmar que es menos inclinado u oblicuo que el cuerpo de la interrogativa.

4º El tonema de la enunciativa que es descendente, Navarro Tomás lo llamaba "cadencia" y decía de él que tenía "unos ocho semitonos por debajo de la lí-

nea del cuerpo" (1944:51). Según nuestros datos, la bajada es de unos nueve semitonos, calculando como línea media del cuerpo la media de la postónica de la segunda posición en la voz masculina: 137 Hz y, como valor final, 78 Hz. En las interrogativas Navarro indicaba que la anticadencia subía "cuatro o cinco semitonos por encima del cuerpo" (p. 51). Nuestros datos proporcionan efectivamente unos 5 semitonos: subida desde 125 Hz hasta 167 Hz. Por tanto, la conclusión es que nuestros datos confirman las afirmaciones de Navarro Tomás.

5° Finalmente, podemos indicar que existe una tendencia a alinear el pico de F0 con la última sílaba de la palabra, auque son necesarios más estudios para acabar de confirmar esta apreciación.

6. POSICIÓN DEL PICO EN HABLA ESPONTÁNEA

Face (2003) encuentra un 25% de picos que se realizan en la tónica en la zona prenuclear de las declarativas en habla espontánea. Además indica que es relativamente habitual hallar palabras, incluso cadenas de ellas, que no poseen ascenso en sus sílabas tónicas, quizás porque sólo suben aquellas que poseen gran importancia comunicativa. También señala la posibilidad de que exista un cierto énfasis siempre que el pico aparece en la misma tónica, aunque indica que son necesarios más estudios sobre la cuestión del énfasis.

En las interrogativas, Cantero et alii (2001) describen cuatro patrones, de los cuales dos tienen primer pico desplazado, pero no ofrecen porcentajes.

Toledo (2000, 2002, 2003, 2004) encuentra que, en todos los corpus analizados en Buenos Aires, España y Cuba, el pico está mayoritariamente en la tónica.

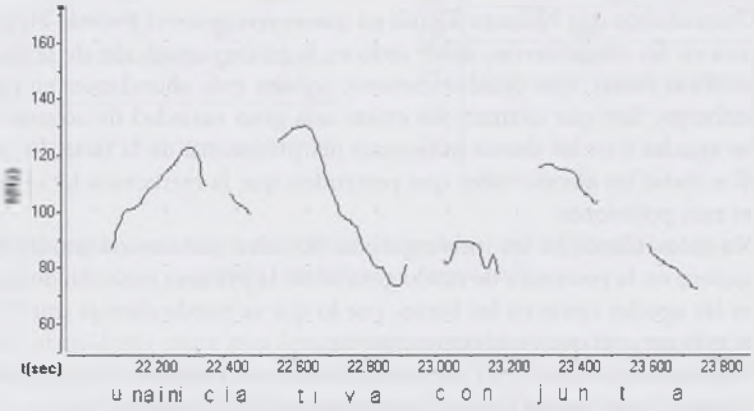


Figura 15. Picos sobre las tónicas en habla espontánea. El énfasis hace que la sílaba átona –ni- posea un pico.

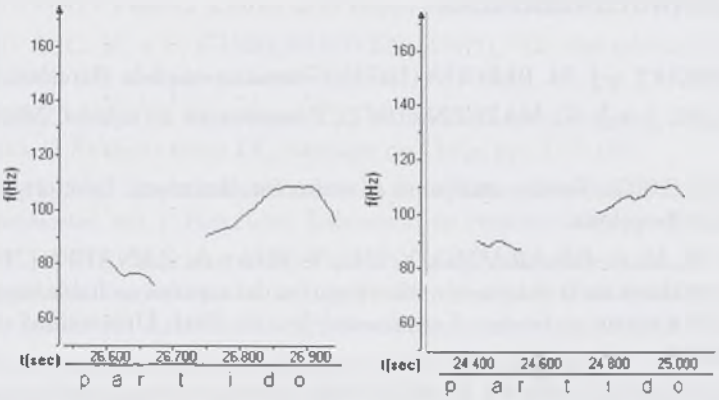


Figura 16. Picos desplazados a la postónica en habla espontánea

Por nuestra parte podemos ofrecer datos del análisis de habla espontánea a partir de una entrevista televisiva a un político. Hemos analizado el habla del entrevistado y hemos encontrado párrafos donde el pico coincide con la tónica en un 50% de los casos y otros párrafos donde eso sucede en un 100%. Somos del parecer de Face: es evidente que el énfasis comporta la coincidencia del pico con la sílaba acentuada. Cuanto mayor es el énfasis, tanto más sube la coincidencia y viceversa.

Las figuras 15 y 16 muestran estas tendencias.

En un estudio sobre el habla semiespontánea a partir del "Map Task" utilizado en las grabaciones de AMPER, la estructura vuelve a ser L*+H en la mayoría de casos cuando las frases son neutras (Fig. 17) y el pico se encuentra en la tónica siempre que se aprecia algo de énfasis.

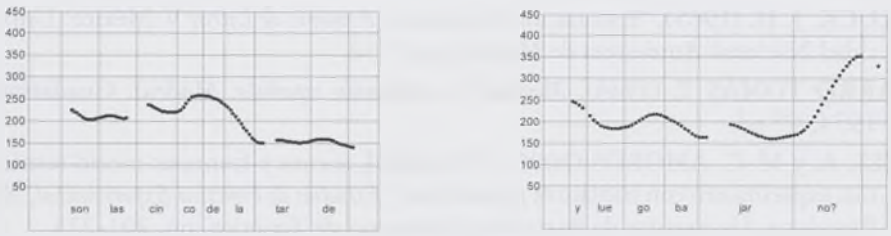


Figura 17 Frases de habla espontánea con picos desplazados a partir del "Map Task"

REFERENCIAS BIBLIOGRÁFICAS

ALCINA FRANCH J. y J. M. BLECUA (1975), *Gramática española*. Barcelona: Ariel.

CANELLADA, M. J. y J. K. MADSEN (1987), *Pronunciación del español*. Madrid: Castalia.

CANTERO, F. J. (2002), *Teoría y análisis de la entonación*. Barcelona: Edicions de la Universitat de Barcelona.

CANTERO, F. J.; M. A. DE ARAÚJO; Y. LIU; Y. WU y A. ZANATTA (2002), "Patrones melódicos de la entonación interrogativa del español en habla espontánea", *Actas del II Congreso de Fonética Experimental*, Sevilla 2001, Universidad de Sevilla, pp. 118-123.

D'INTRONO, F.; E. DEL TESO y R. WESTON (1995), *Fonética y fonología actual del español*. Madrid: Cátedra.

ESTEBAS VILAPLANA, E. (2003), "The modelling of prenuclear accents in central Catalan declaratives", *Catalan Journal of Linguistics 2, Special Issue on Romance Intonation*, vol.2. Barcelona, UAB, pp. 97-114.

FACE T. L., (2003), "Intonation in Spanish declaratives. Differences between lab speech and spontaneous speech", *Catalan Journal of Linguistics 2, Special Issue on Romance Intonation*, vol.2, UAB, pp. 115-112.

GILI GAYA, S. (1950), *Elementos de fonética general*. Madrid: Gredos. Edición de 1975.

LADD, D. R. (1996), *Intonational Phonology*, Cambridge, Cambridge University Press.

LLISTERRI, J.; M. MACHUCA; C. DE LA MOTA; M. RIERA y A. RÍOS (2003), "The perception of lexical stress in Spanish", *Proceedings of the 15th International Congress of Phonetics Sciences*, Barcelona, UAB, pp. 2023-2026.

MARTÍNEZ CELDRÁN E.; A. MA. FERNÁNDEZ PLANAS y N. FULLANA RIVERA (2003), "Pre-nuclear Tonal Inventories of Spanish Intonation", *Proceedings of the 15th International Congress of Phonetics Sciences*, Barcelona, UAB, pp. 595-598.

MATLUCK, J. H. (1965), "Entonación hispánica", *Anuario de Letras* V. México: Universidad Nacional Autónoma de México, pp. 5-32.

NAVARRO TOMÁS T. (1944), *Manual de entonación española*. Madrid: Guadarrama, 1974, 4th ed.

PAMIES, A. y M. C. AMORÓS (2005), "Pico tonal, acento y fronteras morfo-semánticas: experimento con hablantes granadinos", *Estudios de Fonética Experimental*, XIV, Barcelona, Universitat de Barcelona, Laboratori de Fonètica, pp. 201-223.

PRIETO P.; J. VAN SANTEN y J. HIRSCHBERG (1995), "Tonal alignment patterns in Spanish" *Journal of Phonetics*, vol.23, 4, pp.429-451.

QUILIS, A. (1971), "Caracterización fonética del acento español", *Travaux de Linguistique et Littérature*, IX, pp. 53-72.

QUILIS, A. (1981), *Fonética acústica de la lengua española*. Madrid: Gredos.

RIETVELD A. C. M. y C. GUSSENHOVEN (1985), "On the relation between pitch excursion size and prominence", *Journal of Phonetics*, 13, pp. 299-308.

SILVA-FUENZALIDA, I. (1956-57), "La entonación en el español y su morfología", *Boletín de Filología* tomo IX, Santiago de Chile, pp. 177-187.

SOLÉ M. J. (1984), "Experimentos sobre la percepción del acento", *Estudios de Fonética Experimental*, vol. I. Barcelona. Laboratori de Fonètica UB, pp. 131-241.

SOSA J. M. (1995), "Nuclear and pre-nuclear tonal inventories and the phonology of Spanish declarative intonation", in K. Elenius and P. Branderud (eds.) *Proceedings of the ICPhS 95*. Arne Stombergs: Stockholm, vol. 4, pp. 646-649.

SOSA, J. M. (1999), *La entonación del español*. Madrid: Cátedra.

STOCKWELL, ROBERT P.; J. DONALD BOWEN y I. SILVA-FUENZALIDA (1956), "Spanish juncture and intonation", *Language* vol. 32, nº 4, pp. 641-665.

COMPARACIÓN DE LAS ESTRUCTURAS ENTONATIVAS DE LAS HABLAS ROMÁNICAS O LA APUESTA DE AMPER

MICHEL CONTINI
Universidad Stendhal, Grenoble

En el título de mi comunicación, tenía ganas de definir AMPER como un Atlas 'atípico', considerando que en los trabajos de geolingüística, terminados o en curso de realización –al menos en lo que concierne al dominio románico– jamás se había consagrado ningún Atlas a la prosodia y menos aún a la entonación dialectal. Como mucho, en los Atlas de primera generación, los autores se limitaban a señalar sobre los mapas, en las palabras transcritas fonéticamente al lado de cada punto de encuesta, la vocal acentuada, con diferentes diacríticos, colocados encima, debajo o delante de ella. Sé, sin embargo, por mi experiencia personal, que muchos dialectólogos eran conscientes de la laguna que suponía la falta de anotaciones de un hecho lingüístico mayor como la entonación, percatándose del importante papel que jugaba en la caracterización de las variedades dialectales, caracterización que, por otro lado, era percibida inconscientemente por los propios informantes, sin que pudieran definir su naturaleza precisa. Así, el deseo de una investigación que explorara este dominio relativamente poco conocido, difícil de comprender y de cuantificar, fue formulado por muchos de ellos sin encontrar un hecho favorable: la colaboración con los fonetistas parecía la única solución. Por eso, si ninguna empresa geolingüística tuvo en cuenta jamás el parámetro 'prosodia', esto se debía a la falta de preparación de los dialectólogos para abordar el análisis de este aspecto de la lengua, que necesita un conocimiento de la fonética acústica y una formación imprescindible en el enfoque instrumental indispensable, este último, para la cuantificación de una realidad multiparamétrica. Yo añadiría también un cierto recelo de los dialectólogos, bastante acusado en el pasado, respecto de los enfoques cuantitativos sobre los diferentes aspectos de la lengua, aunque se han conocido afortunadas excepciones. Pienso, en particular, en los análisis dialectométricos emprendidos por J. Seguy en el marco del *ALG*, por H.

Guiter en el marco del *ALPO*, que han experimentado posteriormente un desarrollo considerable, gracias a los trabajos de H. Goebl (1981). En lo que respecta a la prosodia, también pueden señalarse excepciones: podría citar a autores como M. Rossi (1977) para una variedad dialectal del toscano, F. Cartón (1972) para las hablas de oïl del norte de Francia, E. Gårding (1975, 1976) y G. Bruce y Gårding (1978) para las estructuras acentuales de los dialectos suecos o, aún más, mis trabajos para algunos dialectos sardos (Contini 1971, 1984, 1989, 1995). Otros trabajos, en los que yo mismo he participado, han adoptado también una visión contrastiva considerando las variedades regionales de una lengua nacional, como las del dominio ítalo-románico (M. Contini, O. Profili, 1989). Sin embargo, más allá de su carácter puntual, reflejan en conjunto las diferentes visiones, tanto en relación con la constitución de los corpora, como con las estrategias de análisis, de interpretación de los datos y la definición del marco lingüístico considerado. A partir de los años 70 los trabajos sobre entonación se han desarrollado considerablemente gracias especialmente a los avances de la síntesis. Los viejos fonetistas, como yo, recuerdan que les inspiraban mucho, en el ámbito lingüístico, los trabajos ya antiguos de K. Pike (1945) sobre el inglés americano, que proponían un análisis de la entonación por niveles: (me refiero, por ejemplo, a los trabajos de P. Delattre (1966), de P. León et M. Léon (1964), en los del equipo del Instituto de Fonética de Aix en Provence, con G. Faure, M. Rossi, M. Chafcouloff, A. Di Cristo y también en nuestros trabajos en el Instituto de Fonética de Grenoble (Contini 1983). Más tarde, el enfoque sintáctico de la entonación, postulado particularmente por los trabajos de Ph. Martin (1975, 1979, 1981) y de M. Rossi (1989), nos había parecido más adecuado para la representación de esta realidad difícil de asumir: la entonación podía ser descrita con un número limitado de rasgos, en oposición binaria, que caracterizan los segmentos situados en las posiciones clave de la frase (fronteras principales o secundarias). Al número tan restringido de rasgos (3/4) considerados por Ph. Martin y relacionados únicamente con las variaciones de F0 y la duración, yo proponía un abanico de rasgos mucho más extenso (hasta 16), permitiendo una mejor caracterización de la configuración de las curvas de F0, una representación de los máximos y los mínimos de los principales parámetros prosódicos (H, I, D) en el interior de cada unidad acentual y con respecto a toda la frase, su posicionamiento en relación con el F0 medio, así como la declinación general de la pendiente (Contini 1991; Contini, O. Profili 1989). Una frase podía estar representada así por un haz de rasgos que evidencian un análisis lingüístico muy próximo a la sustancia fónica. Yo había probado este método en las investigaciones sobre el sardo y, con O. Profili, sobre el italiano regional. Y cuando comencé a pensar en un proyecto de análisis comparativo entre las variedades dialectales románicas, a principios de los 90 (Contini 1991) es así como yo consideraba el enfoque comparativo. Debo decir que, además de mi formación inicial como fonetista, en esta época me estaba convirtiendo ya en dialectólogo. Ya había trabajado especialmente en la extensa obra del *ALE*, primer trabajo de geolingüística a escala continental, a mediados de los 70 en una época en que llevaba la dirección del Instituto de Fonética, y desde 1987, lanzaba el proyecto del *AliR* con mi colega G. Tuaillon, al que reemplacé al año siguiente en la dirección del Centro de Dialectología de la misma Universidad.

Digamos que desde siempre he sido un fonetista-dialectólogo en la medida en que mis investigaciones han estado siempre consagradas al análisis fonético del espacio dialectal, lo que me hacía estar particularmente atento al fenómeno de la variación fonética en su dimensión diatópica, aunque sin olvidar sin embargo las otras dimensiones, diastrática y diacrónica. Se trataba sobretodo de la variación de las unidades segmentales: pero ¿cómo representar la variación prosódica? En cuanto al nuevo proyecto del Atlas Lingüístico Románico que reflejaba también mi interés por el análisis del espacio dialectal de un dominio lingüístico genéticamente homogéneo, éste tampoco abordaba un análisis de la variabilidad prosódica, siendo su objetivo esencialmente el análisis del léxico, aunque también el de la estructura fonética, fonológica y morfológica. En resumen, la prosodia volvía a ser una vez más el pariente pobre: y el ALiR se había puesto en marcha sin tener en cuenta esta dimensión.

He aquí, pues, la toma de conciencia de este vacío que está en el origen del proyecto AMPER, nacido del encuentro de las exigencias y de los deseos del dialectólogo y del fonetista. El dialectólogo, por su experiencia adquirida en el análisis de la variabilidad en el espacio lingüístico, formulaba un cierto número de exigencias que debían ser respetadas, a su parecer, por todos los colegas comprometidos con el proyecto, a saber:

- un criterio mínimo común para la elección de los informantes;
- la definición de la red de encuestas basada en un conocimiento previo del espacio dialectal;
- la realización de un cuestionario para la constitución de un corpus objetivo de frases, que debe servir de base al análisis instrumental, cuestionario que debe ser el mismo para las encuestas;
- la elaboración de una estrategia de encuesta idéntica para todas las variedades consideradas.

Para la elección del locutor, me parecen esenciales dos criterios: en primer lugar, la utilización cotidiana de su habla en todos los usos; en segundo lugar, el reconocimiento de esta habla por parte de la comunidad lingüística como representativa de la variedad estudiada. El primer punto me parece de extrema importancia porque nos enfrentamos hoy al problema de la desaparición de los dialectos, sumergidos por la creciente influencia de las lenguas nacionales. Yo conozco muy bien esta situación para los dominios ítalo-románico y sardo y para el galo-románico. En Francia, por ejemplo, en la región en que yo vivo, el franco-provenzal en tanto que lengua viva, empleada cotidianamente, prácticamente ha desaparecido. Solamente el Valle de Aosta conoce aún un franco-provenzal vivo: lo que me había llevado a pedir a S. Roullet (1999) hacer una Tesis Doctoral dedicada a la entonación de esta región de Italia, Doctorado que representa una de las primeras encuestas en la orientación del proyecto AMPER. Por eso, un análisis de la prosodia necesita, obligatoriamente, un empleo corriente de la lengua. En el espacio franco-provenzal de Francia, salvo casos excepcionales, nos encontramos ante una lengua recordada. Los informantes conocen sin duda las palabras aisladas, las expresiones, pero

tienen dificultades para entablar una conversación. Un análisis de la entonación es, en este caso, irrealizable. Lo que, en cambio, puede interesarnos es la entonación del francés regional hablado por los informantes que se podría definir como 'un sustrato franco-provenzal' y que, indirectamente, puede proporcionarnos información sobre la entonación de la lengua actualmente olvidada. Las investigaciones de S. Roullet, A. Romano (1999) y J. P. Lai (2002) sobre otras variedades dialectales (sardas y del Salento), grabando a los mismos locutores para obtener enunciados comparables en el dialecto y en la lengua nacional, han demostrado efectivamente la persistencia, en esta última, de los esquemas prosódicos dialectales.

Para terminar con este aspecto, pienso que situaciones semejantes pueden encontrarse en todos los dominios lingüísticos: la opinión de los dialectólogos especializados en las diferentes áreas dialectales de cada país se hace en este caso indispensable.

Otro problema surge al tratar de elegir, en un espacio dialectal dado, las variedades más representativas, pues se sabe muy bien que un Atlas como el nuestro no podrá incluir una red tan cerrada como la de un atlas léxico. En este caso también, la colaboración de los dialectólogos parece indispensable porque sólo ellos tienen una visión global del espacio dialectal y pueden contribuir a la selección de las variedades que deberían considerarse en nuestra red. Se debería comenzar por aquellas que son más fácilmente identificables por los propios dialectófonos. Para ello, nos podríamos basar en el test de percepción que realizó A. Romano para seleccionar, entre las variedades del Salento, al sur de Italia, aquellas que debían ser objeto prioritario de sus investigaciones. Este test consistía en la grabación de secuencias emitidas en diferentes localidades del espacio dialectal para, posteriormente, hacer escuchar estas grabaciones a un público no especializado pidiéndole identificar el origen geográfico de los dialectófonos. La elección se basó en las variedades más marcadas, aquellas que fueron más fácilmente identificada por los oídos 'ingenuos', fundamentadas en una simple impresión perceptiva, aún sabiendo, sin embargo, que la identificación del origen geográfico de un habla o de una variedad regional de la lengua nacional no se basa forzosamente en un solo parámetro, sino que puede deberse a fenómenos segmentales, suprasegmentales, morfológicos e, incluso, al empleo de elementos léxicos particulares. Se trata de un ejercicio al que nosotros mismos estamos acostumbrados, en nuestra deformación profesional de fonetistas, cuando nos esforzamos, al oír hablar a una persona desconocida, en averiguar su procedencia: «¡Señor, es usted de Málaga!». «¿Pero cómo lo ha adivinado?»... Es un secreto...

Para la metodología de la encuesta, deberíamos basarnos también en el trabajo de los dialectólogos, dada la orientación geolingüística de nuestro proyecto. Después de más de un siglo de investigaciones sobre la variabilidad lingüística, habían decidido la utilización de un único cuestionario que debía ser usado para todas las encuestas, para cada punto de encuesta, requisito indispensable para la realización de un Atlas. Habría que inspirarse también en los dialectólogos que, con el fin de hacer un análisis comparativo del léxico o de la morfo-sintaxis, habían creado una misma estrategia de encuesta para todos los puntos del mapa y, en particular, el empleo de la pregunta indirecta, con la misma fórmula, que permite obtener del informante dialectal una respuesta dada, como

por ejemplo la designación de un objeto sin nombrarlo (por ejemplo: '¿cómo llama usted al objeto metálico que sirve para segar el trigo?'). Se sabía que el hecho de nombrar el objeto (en este caso, la azada) en la pregunta, podía influir en la respuesta, tanto como al pedir una traducción a partir de un texto escrito.

Parece, desde el comienzo, que este método de la dialectología debía ser adoptado por los fonetistas que tomaban el relevo en el proyecto AMPER: he aquí la razón por la que la colaboración con los colegas dialectólogos implicados en el proyecto del Atlas Lingüístico Románico me parecía y me parecerá siempre más que deseable.

Al menos en teoría, porque es evidente que una encuesta que permite constituir un corpus de frases destinado al análisis de la entonación no es del todo la misma que la que pretende recopilar palabras aisladas: es más fácil, sin duda, obtener la designación de un objeto que una frase entera con una entonación particular. Pero el espíritu de la encuesta dialectal debería ser respetado: un modo de proceder 'mayéutico' que pretende que el informante 'dé a luz' la frase deseada. La cosa no es fácil. En los años 70, con mi colega L. J. Boë, habíamos emprendido investigaciones sobre la entonación de la frase afirmativa y la pregunta total en francés, en el Instituto de Fonética de Grenoble (Contini, Boë 1973; Boë, Contini 1975). Habíamos utilizado 10 locutores (5 hombres y 5 mujeres) que presentaban una entonación no marcada por características regionales y pertenecientes a la misma categoría socio-profesional (se trataba de universitarios), para pronunciar 90 frases de diferente complejidad (45 afirmativas y 45 interrogativas), con 10 repeticiones cada una, siendo presentadas en orden aleatorio para evitar efectos de serie. Una primera experiencia, a partir de la lectura de las frases, nos había llevado a abandonar este método que se traducía en una falta de espontaneidad de los locutores y en una tendencia a la disminución de las variaciones de F0. Esta constatación me había llevado, desde el lanzamiento del proyecto AMPER, a no recurrir a la lectura. Tanto es así que en este nuevo proyecto nos propusimos no contar con informantes pertenecientes a la categoría de intelectuales, totalmente familiarizados con el universo de lo escrito, sino con dialectófonos, cuyas hablas, lenguas de tradición oral, no dispusieran por lo general de una tradición escrita y quienes además, por su nivel cultural a menudo bastante bajo, no se encuentren en una situación comparable frente a la lectura de un texto. Por tanto, parece que para la comparación entre lenguas los enunciados deben ser producidos, todos, con un mismo método, sin mezclar los que hayan sido leídos y los otros: ya habíamos abordado este tema durante nuestra última reunión de Grenoble. Parece pues que, por honestidad científica, cuando se adopte el método de la lectura para una encuesta determinada, debería señalarse con un signo cualquiera (por ejemplo una L mayúscula) en la base de datos que estamos creando. Me permito, sin embargo, llamar vuestra atención sobre un punto débil como es la presentación, para el análisis comparativo, de una base de datos donde éstos hayan sido obtenidos con diferentes estrategias. Era ya el principal reproche que se podía hacer a los estudios, numerosos, sobre la entonación, en las diferentes lenguas: notables, sin duda, pero que no permitían un trabajo de comparación, en tanto que los métodos para la constitución del corpus, para el análisis y la interpretación de los datos eran diferentes. Añadiré que en ningún caso se marcaban como objetivo la

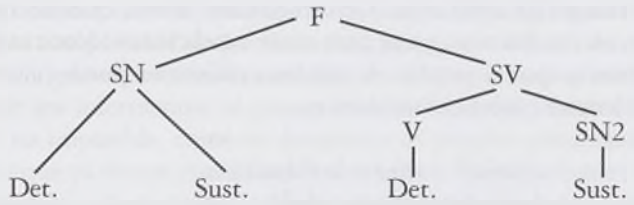
comparación entre lenguas, en el marco de un estudio geolingüístico, como es el caso de nuestro proyecto. No dejamos de reprochárnoslo. No recurrir a la lectura sin duda hace la encuesta más complicada y más larga: pero el resultado vale la pena. AMPER es el primer trabajo geoprosódico que abarca toda una familia lingüística genéticamente homogénea pero sin duda fuertemente diferenciada, especialmente en su estructura fónica segmental y suprasegmental, por la influencia de diversos sustratos y superestratos: su objetivo de comparar estructuras entonativas de diferentes variedades debe respetar reglas estrictas comunes a todos los participantes. Sobre el problema de la constitución del corpus, permitidme volver a mi experiencia de una treintena de años. Habiendo renunciado a las frases leídas, habíamos adoptado una nueva estratégia. El locutor y el encuestador se encerraban juntos en la cabina insonorizada: este último leía, con una entonación neutra, una frase dada indicando, al final, la modalidad (Ejemplo: *A Moulin on aimait les lilas*. 'En la ciudad de Moulin gustaban las lilas'. Interrogativa.). Después de la intervención del encuestador, transcurrían unos diez segundos, suficientes para eliminar una eventual influencia de la frase pronunciada sobre la frase esperada. Los resultados eran ya mucho más interesantes: permitían obtener un mayor grado de espontaneidad con una mejor dinámica de F0, lo que evidenciaba también una mayor variabilidad interlocutora, siendo necesaria, en consecuencia, la puesta en práctica de una estrategia de normalización (Bellet, Contini, Boë 1981-1982). Para las encuestas en el marco del proyecto AMPER hemos propuesto una estrategia diferente, pensada para la realización de investigaciones preliminares, basándonos siempre en la pregunta indirecta de los dialectólogos. Imaginemos que se quiera obtener del locutor una frase como 'El pájaro mira a la chiquita': se le pedirá: «Dime que el pájaro mira a la chiquita» o bien «Pregúntame si el pájaro mira a la chiquita». Seguramente, al principio, se puede observar cierta confusión o puede que se den respuestas que no se correspondan con nuestras expectativas. Los locutores también pueden estar desconcertados por el tipo de frases que les pedimos pronunciar, que pueden parecerles algo extrañas: era sin duda el caso de Amina Rhardisse (1995), cuando pedía a sus informantes aragoneses de Bielsa que pronunciaran frases como *El capotáz belsetán mira al caragol*; cuando L. Castro Moutinho (Castro Moutinho et al. 2004) pedía a los suyos pronunciar *O passaro toca no Toneca bisavô* 'el pájaro toca al bisabuelo Toneca' o también cuando J.P. Lai exigía a los sardófonos de Nuoro frases como *Sa pípera toca sa patata* 'la culebra toca la patata'. Pero, una vez explicada la razón de la utilización de frases a veces extravagantes, y que vamos más allá del contenido semántico de los enunciados, una vez explicado nuestro deseo de oír enunciados producidos con la máxima espontaneidad, los informantes son capaces de pronunciarlos sin dificultades. Lo que me parece deseable, en cambio, en la medida en que están previstas varias repeticiones de la misma frase (tres como mínimo), es hacer escuchar al informante sus propias grabaciones pidiéndole que señale las frases que le parecen menos naturales o menos espontáneas y, si fuera necesario, realizar nuevas grabaciones.

En conclusión: podemos preguntarnos por qué utilizar frases tan poco naturales, que pueden confundir a los informantes. ¿No sería más simple utilizar frases más banales, frases que se oyen todos los días en conversaciones corrientes? Sin duda sería la solución

ideal. Pero, ¿cómo comparar un número prácticamente ilimitado de enunciados, producidos en las situaciones de comunicación más diversas y en variedades lingüísticas tan distantes como la de la Barbagia, en las montañas de Cerdeña, la de la isla de San Miguel en las Azores o la del nordeste de Brasil?

El objetivo del proyecto AMPER es justamente esclarecer toda esta diversidad, partiendo de enunciados simples, con estructuras sintácticas comunes al conjunto de las variedades, para llegar al análisis comparativo de sus esquemas prosódicos: acatando el principio de que sólo se pueden comparar enunciados comparables, tanto en lo que respecta a su tamaño (número de sílabas) como al tamaño y la estructura acentual de las unidades que los componen.

Se comparará, por ejemplo, una frase simple del tipo



presente en todas las variedades analizadas. Se sabe, por otro lado, que los elementos acentuables que componen esta frase, situados en las fronteras principales (final del SN y final del SV), en todas las variedades románicas, pueden pertenecer a varias categorías acentuales, por lo que hay que tenerlo en cuenta: conociendo la situación en el conjunto del espacio románico, hemos considerado tres unidades lexicales (o palabras prosódicas) acentuadas en la sílaba final (oxítonas), en la penúltima (paroxítonas) y en la antepenúltima (proparoxítonas). Las unidades lexicales consideradas constan, generalmente, de tres o cuatro sílabas, más el determinante. Se sabe, sin embargo, que en las variedades románicas puede haber 'palabras prosódicas' acentuadas en la cuarta (por ejemplo: en sardo sos óminese 'los hombres') e incluso en la quinta sílaba antes del final (por ejemplo: en italiano telefonamelo 'telefonéamelo') y que, a la inversa, las variedades galo-románicas del dominio de oïl, y por tanto el francés, sólo presentan unidades oxítonas. El corpus de estas últimas será, en consecuencia, más reducido.

Así, la frase simple anterior será pronunciada con sustantivos, en las posiciones clave, pertenecientes a cada una de las categorías acentuales y con todas las combinaciones posibles entre los dos sintagmas: lo que representa, ya, con las dos modalidades y las repeticiones de cada frase, un conjunto de 90 enunciados. Pero es difícil economizar: se sabe que existe una estrecha relación entre la organización de las 'palabras prosódicas' y la entonación general de la frase y que, por consiguiente, sería imprudente comparar dos enunciados que llevaran, por ejemplo, al final del SN o del SV, un oxítono uno y un proparoxítono el otro.

Se ha de comparar, por ejemplo, dos frases con la misma estructura sintáctica

'el jabalí mira a la chiquita'
'el pájaro mira a la chiquita'

Las frases simples se han hecho después más complejas con la adición de expansiones en cada uno de los dos sintagmas: una con un adjetivo y otra con un sintagma preposicional que contiene un topónimo.

Ejemplos:

el pájaro bonito *mira a la chiquita*
el pájaro bonito de Lérida *mira a la chiquita*

Por las mismas razones ya explicadas, cada elemento de la expansión debe pertenecer a las tres categorías acentuales. Otro requisito: hemos querido que figurasen las mismas palabras en los dos sintagmas para tener condiciones idénticas de comparación evitando la influencia que el empleo de palabras diferentes puede tener sobre el informante, por su diferente contenido semántico.

Ejemplos:

el jabalí mira a la chiquita
la chiquita mira al jabalí

Por tanto, cada corpus consta solamente de tres sustantivos, tres adjetivos y tres topónimos. Añadimos que, por lo que respecta a la estructura fónica de estos últimos, hemos favorecido, en la medida de lo posible, un entorno consonántico sordo, así como la presencia de vocales de gran abertura, que permiten una mejor segmentación. Se entiende que, con todos estos requisitos, a menudo haya sido difícil confeccionar las frases: lo que justifica, entre otros, el empleo de topónimos en los sintagmas preposicionales. Dificultad acrecentada también por la necesidad de construir frases 'posibles' que puedan ser empleadas, si no en una conversación corriente, al menos en ciertas situaciones comunicativas. Así, más allá de las dificultades señaladas, se puede aceptar que las frases siguientes, a pesar de su diferencia por pertenecer a variedades románicas distintas, puedan ser perfectamente comparadas:

prt. *O pàssaro toca no Toneca.*
glc. *O pícaro xogaba co cadelo.*
cat. *La crítica no porta la caputxa.*
srd. *Sa pípera toca sa patata.*
rou. *O pas_re vede o fantoma.*
arg. *El pàjaro mira al misache.*

Para facilitar la obtención de los enunciados, también se han realizado dibujos o escenificaciones con las representaciones de cada uno de los elementos constitutivos de las frases. Pienso, entre otros, en el dibujo realizado por nuestros colegas de Aveiro con Un

pàssaro que toca o Tuneca bisavó (Castro Moutinho et al. 2004). Y funcionó bien.

En una segunda fase, debería poder establecerse un corpus de control utilizando un enfoque diferente, a saber, la técnica del Map Task. El informante se pone enfrente del encuestador o dos informantes uno enfrente del otro: tienen delante un mapa sobre el que están representados, entre otros, los referentes de las palabras utilizadas en el corpus principal. Sobre el mapa de uno de ellos figura también un itinerario que se desarrolla a través de los dibujos. Por medio de preguntas y respuestas recíprocas, el segundo informante debe trazar sobre su mapa el mismo recorrido y pronunciar así frases afirmativas o interrogativas totales, fuera del cuestionario. Esto permite verificar si los esquemas de las frases del primer corpus se repiten en el segundo, producido con un grado de espontaneidad más importante.

Un tercer corpus de grabaciones libres debería permitir, en fin, verificar la presencia de estructuras entonativas presentes en la lengua hablada, fuera de toda exigencia. Vemos así la gran complejidad de las encuestas que, teniendo en cuenta el tamaño del corpus que se ha de grabar, necesitan una colaboración importante y sobre todo una gran disponibilidad de los informantes, al menos dos por localidad, cosa no siempre fácil de encontrar. Pero no imposible, como lo demuestra el número creciente de encuestas ya realizadas, de las que ya dieron cuenta sus autores el año pasado durante el 2º Seminario AMPER de Grenoble y darán cuenta también en sus intervenciones durante esta jornada.

Para quienes pudieran considerar que este enfoque es demasiado simple, yo destacaría que, teniendo en cuenta todas las combinaciones posibles entre los dos sintagmas principales, y en las dos modalidades, se llega a un número considerable de enunciados (varios cientos) que, a mi parecer, merecerían todos ser analizados. Sólo después del análisis global de un cierto número de encuestas, se podrá, eventualmente, eliminar del cuestionario ciertas frases con expansión, habiendo constatado que no aportan ninguna modificación sobre la estructura entonativa de la frase más simple. Es muy evidente que, al final, la base de datos de AMPER debería estar compuesta por un corpus idéntico de frases para todas las variedades representadas: se vería mal que una variedad estuviera representada por 80 frases y otra por 40, la consulta de la base de datos tendría lagunas imprevistas y mal consideradas para ciertos puntos de la red.

Si para los aspectos precedentes (constitución del corpus, estrategia de encuesta) la experiencia de los dialectólogos ha podido ser valiosa, el trabajo de análisis de la sustancia fónica del (o de los) corpus concierne exclusivamente a la competencia de los fonetistas. También para esta fase se imponen tres requisitos, a saber

- la puesta en marcha de un procedimiento de análisis acústico estrictamente idéntico;
- la adopción de un mismo programa de digitalización de datos;
- la adopción de un sistema de codificación único.

Las elecciones se hicieron según las unidades segmentales a considerar y el tipo de medidas a realizar. Por lo que respecta a las primeras, sólo se han tenido en cuenta las vocales al considerar que son portadoras, sin duda, de la mayor parte de la información

prosódica. Somos conscientes de que esta elección es restrictiva en la medida en que se elimina información que también puede aportar el entorno consonántico: pero también sabemos que las consonantes, según su naturaleza, pueden introducir variaciones, especialmente en la altura y la duración de las vocales contiguas. Para reducir las influencias contextuales, hemos pretendido, en la medida de lo posible, el empleo de consonantes sordas. El ideal habría sido encontrar, en todas las variedades, unidades lexicales pertenecientes a las tres categorías acentuales y compuestas sólo por consonantes sordas y por vocales de timbre próximo, por ejemplo, vocales de gran abertura: la experiencia nos demostró que esto no era posible, al menos en todos los casos. Decimos que esta elección facilitaba sobre todo la segmentación de los enunciados y, por tanto, el trabajo de análisis y de interpretación de una realidad ya particularmente compleja. Añadimos que, para reducir además la influencia de las variaciones relacionadas con el entorno y con las características intrínsecas de las vocales sobre la estructura entonativa de la frase, hemos propuesto utilizar las mismas unidades lexicales en las fronteras principales.

En cuanto a la digitalización y el tratamiento de los datos, pensamos desde el principio que debía realizarse a partir de un mismo programa (Matlab) para el conjunto de las encuestas.

Siguiendo la estrategia adoptada desde hace tiempo por los equipos del Instituto de Fonética/IPC y el Centro de Dialectología de Grenoble, pero también por otros equipos (Instituto de Fonética de Aix-en-Provence, IPO de Eindoven), para el análisis del corpus se han considerado tres parámetros: la variación temporal de la frecuencia laríngea (F0), de la duración (D) y de la intensidad (I). Sobre cada segmento vocálico se realizaron tres medidas para la F0 (inicio, final y un punto intermedio), dos para la duración (inicio y final) y una para la intensidad (cumbre de la curva).

Por otro lado, se creó un sistema de codificación para el conjunto del proyecto AMPER, permitiendo el tratamiento y la comparación de los datos relativos a cada encuesta.

La repetición de cada frase (sería deseable tener al menos tres repeticiones) permite obtener las medias de F0 por medio de un tratamiento estadístico y describir, a partir de ellas, las curvas estilizadas con las que se puede realizar una modelización que pueda servir de base a un trabajo de comparación entre variedades.

El objetivo final del proyecto AMPER es la constitución de una base multimedia de datos prosódicos que sirva de soporte a lo que en otro tiempo se llamaba un Atlas lingüístico parlante. Uno de los primeros Atlas de esta nueva generación –y en todo caso el primero en Francia– se realizó hace exactamente 10 años en el Centro de Dialectología de Grenoble (Marquet 1995). Se trataba de un Atlas léxico, dedicado a un espacio dialectal que comprende una área alpina donde conviven hablas occitanas y franco-provenzales, cubierto por una red de una veintena de puntos de encuesta, y con un corpus de unas 800 palabras agrupadas en 16 categorías semánticas. El Atlas, que se puede consultar en la red en el sitio web de la Universidad Stendhal, permite elegir, dentro de una determinada categoría que aparece en pantalla (por ejemplo, la de los fenómenos atmosféricos), una palabra (por ejemplo, arco iris) y, además, haciendo clic sobre cada punto de encuesta

en un mapa visualizado también en pantalla, se puede oír la designación del referente en cuestión con la pronunciación de los propios informantes, grabados in situ durante las encuestas. Esto representa un progreso respecto de los Atlas precedentes, en soporte de papel, donde cada palabra se representaba en los mapas con trascripción fonética, la cual, como sabemos, no es más que una aproximación a la pronunciación real. Los Atlas parlantes, al introducir el parámetro Audio, se hacen accesibles también para un público no iniciado. Añadiré que también es posible agregarles el parámetro Vídeo, en la medida en que se puede colocar la transcripción fonética de las palabras escuchadas y a la vez visualizar la propia encuesta, que ha sido filmada, permitiendo observar al informante cuando pronuncia las diferentes palabras. Esto es lo que hizo A. M. Kristol (2000) para un Atlas parlante del Valais Románico y para el Atlas Lingüístico Multimedia de la región Rhône-Alpes y de las regiones limítrofes (*ALMURA*) que mi colega J. E. Médélice (2002) está realizando en el Centro de Dialectología de Grenoble.

No niego que esta experiencia ha jugado, sin duda, un papel importante en la decisión de lanzar el proyecto AMPER, pues la base de datos que pretendemos conseguir para este último presenta muchos rasgos comunes con las de los proyectos de Atlas léxicos de los que acabo de hablar.

¿Cuál será el producto final de nuestro trabajo?

Se tratará de una Base de Datos prosódicos compuesta por un conjunto de corpora de frases, con las diferentes estructuras consideradas, clasificadas en un mismo orden, de manera que cada corpus represente una variedad románica o, si se quiere, un punto de encuesta. El objetivo será comparar las estructuras entonativas que caracterizan cada tipo de frase y lo mismo para el conjunto de las variedades consideradas. Un modelo prototípico de presentación de esta base de datos ya ha sido realizado en Grenoble.

Por su fácil utilización, la consulta de la BD será accesible para los no especialistas, pero responderá también a las exigencias de los especialistas, dialectólogos y fonetistas.

Comenzaremos por la colocación en la pantalla de lo que podría llamarse 'el menú', es decir, el conjunto de las estructuras de frases disponibles, o sus referentes en cierto modo, presentados esquemáticamente.

Después de haber seleccionado una de estas frases en el 'menú', una vez elegida la modalidad, aparecerá un mapa del espacio románico de Europa o de América Latina, con las principales áreas dialectales. Se podrá efectuar un zoom sobre cada país para delimitar mejor sus áreas específicas, incluso clicar en alguna de ellas si allí se han realizado varias encuestas.

Clicando a continuación sobre uno de los puntos de encuesta y así sucesivamente sobre los otros, se podrá oír, en cada ocasión, la realización en la variedad local de la frase correspondiente a la estructura seleccionada. Al mismo tiempo, se podrá obtener sobre la pantalla el trazado de las curvas de F0 y los histogramas que representan las variaciones de duración y de intensidad de cada segmento vocálico, y asociar así a la audición de una frase la imagen de su estructura entonativa. También será posible visualizar –sobre la misma imagen (por ejemplo, con colores diferentes):

- las curvas de una misma frase, pronunciada en las dos modalidades, afirmativa e interrogativa;
- los histogramas con las duraciones y/o las intensidades de los diferentes segmentos constitutivos de la frase, siempre en las dos modalidades;
- los trazados de F0 y/o los histogramas de duración y de intensidad de distintas variedades permitiendo una inmediata comparación visual entre variedades (E. Martínez Celdrán et al. 2004);
- los trazados estilizados de las curvas de F0 de cada frase o de una misma estructura de frase pronunciada en diferentes variedades, como los que han sido realizados en numerosos trabajos.

El programa Matlab permitirá también llevar a cabo tests de percepción comparativos de la entonación neutralizando el contenido semántico de una frase y conservando sólo la información sonora relativa al contorno melódico.

La estructura definitiva del programa para la constitución de la BD-AMPER está siendo realizada en Grenoble en colaboración con el ICP (A. Rillard), con los equipos AMPER-Portugal de la Universidad de Aveiro (A. Texeira) y AMPER-Italia de Turín (A. Romano). Esta acogerá al conjunto de datos relativos a todas las encuestas realizadas en el espacio románico europeo y de América Latina.

La BD Amper será instalada en el sitio web del Centro de Dialectología de Grenoble.

Y habrá que acelerar la marcha, pues el proyecto crece y llama la atención de los especialistas, gracias, debe decirse, a la motivación, a la actividad investigadora y al dinamismo de los colegas ya comprometidos. Se puede decir que en la mayoría de los encuentros científicos (coloquios y congresos) de dialectología o de fonética organizados durante los últimos cinco años en los países románicos de Europa, el proyecto AMPER siempre ha estado representado. Este éxito, algo inesperado y que a la vez nos alegra, ha sido confirmado por la celebración de este 3er Congreso de Fonética Experimental de Santiago de Compostela (España), que ha dedicado a nuestro proyecto una jornada entera de las tres previstas. Esta atención nos emociona profundamente y, en nombre de todos los colegas, querría agradecer de todo corazón a los organizadores, por su acogida y su confianza.

Las 23 comunicaciones previstas para hoy se añadirán a los 76 artículos, comunicaciones o conferencias ya consagradas, en este día, al proyecto AMPER, en todos los países participantes en la 'empresa'. El número de publicaciones alcanzará esta tarde la cifra de 99, cifra destinada a crecer rápidamente con la ampliación de la red a los dominios latinoamericanos representados hoy por los colegas de Brasil, Chile y Cuba.

Al ir adquiriendo importancia, se hacía necesario dar al proyecto una estructura de organización científica que definimos con ocasión del 2º Seminario de Grenoble y que querría presentar brevemente a los nuevos colegas que se nos acaban de unir y también, y sobre todo, a los otros colegas fonetistas y lingüistas que participan en este Congreso. Para comenzar diré que cinco países románicos de Europa están representados actualmente en el proyecto: se trata, de oeste a este, de Portugal, España, Francia, Italia

y Rumania. Están representados por Comités nacionales o dominios lingüísticos. Por su situación lingüística particular, España cuenta con tres Comités que representan los dominios gallego, español y catalán. Los dominios portugués, galo-románico (de Francia, Bélgica y de Italia: con los dominios de oïl, occitano y franco-provenzal), ítalo-románico/sardo y rumano, están representados por un solo Comité. Los siete Comités están bajo la responsabilidad de eminentes colegas, fonetistas 'puros' o fonetistas-dialectólogos, que constituyen el Comité Científico Internacional del proyecto, creado con ocasión del 2º Seminario Internacional de AMPER que tuvo lugar en Grenoble el año pasado. Se trata de los colegas L. Castro Moutinho (dom. portugués), E. Fernández Rei (dom. gallego), E. Martínez Celdrán (dom. español), A. M. Fernández Plana (dom. catalán), M. Contini y J. P. Lai (dom. galo-románico), A. Romano (dom. ítalo-románico/sardo), A. Turculet (rumano). A los países europeos, acaban de sumarse países de América Latina (Brasil, Chile, Cuba): contrariamente al ALiR, pues el proyecto AMPER tiene en cuenta las variedades románicas de la Romania Nueva. Por el momento, las investigaciones de las variedades del español o del portugués del Nuevo Continente están coordinadas por los responsables de los mismos dominios en Europa: con la extensión del proyecto, se podría aspirar a la colaboración de responsables para los dos dominios lingüísticos del otro lado del Atlántico. Pienso que la cuestión interesa más particularmente a nuestros colegas L. Castro Moutinho y E. Martínez Celdrán, que encontrarán sin duda la mejor solución.

La coordinación general está a cargo de M. Contini (Grenoble) y A. Romano (Turín) y la secretaría científica de J. P. Lai (Grenoble), Ingeniero de estudios, del Centro de Dialectología de Grenoble, promotor del Atlas, que pone a su disposición una Secretaría administrativa, a cargo de C. Chevallier y, además del soporte informático indispensable para el desarrollo del proyecto, al que el Ministerio de Educación ha destinado (¡al menos para el plan cuatrienal en curso!) una partida presupuestaria.

Pienso que ya he dicho lo esencial.

Con una metodología común bien definida, una estrategia de encuesta que ya comienza a ser bien asimilada y objetivos bien fijados, y también con una dirección científica y administrativa establecida, el proyecto AMPER debería poder avanzar sin demasiada dificultad. La única dificultad será, sin duda, obtener cierta disciplina por parte de todos los participantes que permita proporcionar, con regularidad, los datos de cada una de las encuestas respetando, estrictamente, las reglas comunes (pienso especialmente en la codificación), pero también los detalles que iremos fijando a medida que avance el proyecto. Creed en mi experiencia de grandes empresas geolingüísticas como el *ALE* y el *ALiR*: no es fácil hacer trabajar juntos a decenas de colegas, pertenecientes a países diferentes, y sobre todo al mismo ritmo... pero podemos conseguirlo, si conservamos toda nuestra motivación y el deseo de concluir con éxito nuestra común empresa científica. Esta empresa se nos presenta, desde este momento, como un proyecto realizable. Sabemos que muchos trabajos recientes relativos, particularmente, a los dominios ítalo-románico, catalán o portugués, muestran desde ahora resultados muy interesantes para la comparación de estructuras entonativas de variedades diferentes. Todos esperamos con impaciencia poder disponer, rápidamente, de un gran número de encuestas para comparar distintas

variedades del espacio románico. Nuestra base de datos estará seguramente operativa en un corto plazo: una vez instaurada, será preciso 'alimentarla' regularmente. Se trata de una Base de Datos abierta. Esperamos que pueda recibir rápidamente los datos de las principales variedades románicas de Europa y de América Latina, cuyos representantes de los dominios brasileño, chileno y cubano, están presentes hoy en este Congreso, a quienes les doy la bienvenida.

Gracias por vuestra atención.

RÉFÉRENCES BIBLIOGRAPHIQUES

ALE, *Atlas Linguarum Europae* (1983-1990), I, 1-4, Assen-Maastricht, Van Gorcum; (1996) I, 5; (2004) I, 6 , Roma, Istituto Poligrafico e Zecca dello Stato.

ALG; Séguy, J., avec la collaboration de Allières J. et de Ravier X. (1954-1973), *Atlas linguistique et ethnographique de la Gascogne*, Paris, CNRS.

ALiR, *Atlas Linguistique Roman*, Rome, Istituto Poligrafico e Zecca dello Stato, vol. I (1996), 3 tomes; vol. IIa (2001), 2 Tomes.

ALPO; Guiter, H. (1966), *Atlas Linguistique des Pyrénées Orientales*, Paris, CNRS.

BELLET, G.; CONTINI, M.; BOE, L. J. (1981-1982), Normalisation temporelle et fréquentielle de Fo intra- et inter-individuelle, in *Bulletin de l'Institut de Phonétique de Grenoble*, 10/11, 265-283.

BOE, L.-J. et CONTINI, M. (1975), Etude de l'intonation de la phrase interrogative en français (Question totale). Premiers résultats, in *Bulletin de l'Institut de Phonétique de Grenoble*, 4, 85-102.

BRUCE, G.; GÅRDING, E. (1978), A prosodic typology for Swedish dialects, in *Nordic prosody, Travaux de l'Institut de Linguistique de Lund*, 219-228.

CABRERA-FRACHON, C. (1994), *Accent et intonation en castillan: phrase affirmative et interrogative*, Thèse de Doctorat en Sciences du Langage, Université Stendhal, Grenoble, 4 vol.

CARTON, F. (1972), *Recherches sur l'accentuation des parlers populaires de la region de Lille*, (Thèse d'Etat, Université de Strasbourg, 1970), Lille, Service des reproduction des Thèses, Univ. de Lille III.

CASTRO MOUTINHO, L.; COIMBRA, R. L.; PEREIRA BENDIHA, U. (2004), *Projecto de um Atlas Prosódico Multimédia do Espaço românico. Equipa portuguesa*, 2e Séminaire International du Projet AMPER (Grenoble 2004), Actes, Grenoble, Centre de Dialectologie/ELLUG, Hors série 3 de *Géolinguistique*, 7-18.

CASTRO MOUTINHO, L.; COIMBRA, R. L.; TEIXEIRA, A.; PEREIRA, M. (2005), *Variação entonacional em três areas dialectais de Portugal continental*, 2e Séminaire International du Projet *AMPER* (Grenoble 2004), Actes, Grenoble, Centre de Dialectologie/ELLUG, Hors série 3 de *Géolinguistique*, 19-37.

CELDRÁN MARTÍNEZ, E.; FERNÁNDEZ PLANA, A. M.; CARRERA SABATÉ, J.; ESPUNY MONSERRAT, J., (2005), *Approche de la carte prosodique dialectale de la langue catalane en Catalogne*, 2e Séminaire International du Projet *AMPER*

(Grenoble 2004), Actes, Grenoble, Centre de Dialectologie/ELLUG, Hors série 3 de *Géolinguistique*, 103-175.

CONTINI, M. (1971), Contribution à l'étude instrumentale de l'intonation en sarde, XIIIe *Congrès International de Linguistique et de Philologie Romanes*, Québec 1971, Actes, Québec, Presses de l'Université Laval, 1976, pp. 229-245.

CONTINI, M. (1984), L'intonation des phrases affirmatives et interrogatives avec inversion en sarde. Analyse et synthèse, *Bulletin de l'Institut de Phonétique de Grenoble*, XIII, pp.131-152.

CONTINI, M. (1983), *Etude de géographie phonétique et de phonétique instrumentale du sarde*, Thèse de Doctorat d'Etat, Strasbourg 1983, 4 volumes; publiée partiellement, avec le même titre, à Alessandria, Edizioni dell'Orso, 2 vol.

CONTINI, M. (1989), L'interjection en sarde. Une approche linguistique, in : *Espaces romans*, Grenoble, ELLUG, 320-329

CONTINI, M. (1991), Vers une géoprosodie, *Nazioarteko Dialektologia Biltzarra Agiriak*" (Bilbao, 1991), Bilbao, Publ. Real Academia de la Lengua Vasca, 1992, 83-109.

CONTINI, M. (1995), "Visti l'as? "Un trait syntaxique et intonatif sarde dans le catalan d'Alghero, in : *Estudis de lingüística i filologia oferts a Antoni M. Badia i Margarit*, Publicacions de l'Abadia de Monserrat, 1995, vol. 1, 221-247

CONTINI, M.; BOE, L. J. (1973), Contribution à l'étude quantitative de l'évolution de la fréquence laryngienne dans la phrase énonciative en français, in *Bulletin de l'Institut de Phonétique de Grenoble*, II, 77-92.

CONTINI, M.; BOE, L. J. (1979), Etude quantitative de l'intonation en français, in *Recherches sur la prosodie du français*, Grenoble, Publications de l'Université des Langues et Lettres, 117-129.

CONTINI, M.; PROFILI, O. (1989), L'intonation de l'Italien régional. Un modèle de description par traits, in *Mélanges de phonétique générale et expérimentale offerts à Péla Simon*, Strasbourg, Publications de l'Institut de Phonétique de Strasbourg, 855-870.

CONTINI, M.; FRANCHON CABRERA, Cl.; RHARDISSE, A. (1995), Analyse comparée de l'intonation en castillan et en aragonais, in *Atti del Congresso Internazionale di Linguistica e Filologia Romanza* (Palermo, 1995), 143-157.

CONTINI, M.; LAI, J.P.; ROMANO, A.; ROULLET, S. (2003), Vers un Atlas prosodique des variétés romanes, *Sempre los camps auràn segadas resurgantas*, in Mélanges offerts à Xavier Ravier, Toulouse, CNRS/ Université de Toulouse-Le Mirail, 73-84.

DELATTRE, P. (1966), Les dix intonations de base du français, in *French Review*, 40/1, 1-14.

ENDO, R.; Bertinetto, P.M. (1997), Aspetti dell'intonazione in alcune varietà dell'italiano, in *Atti delle VII G. di St. del GFS* (Napoli 1996), Roma, Esagrafica, 27-49.

GOEBL, H. (1981), Eléments d'analyse dialectométrique (avec application à l'AIS), in *Revue de Linguistique Romane*, T. 45, 349-420.

GÖRDING, E. (1975), Toward a Prosodic Typology for Swedish Dialects, *The Nordic Languages and Modern Linguistics*, 2, Stockholm, Ed. K. H. Dahlstedt Almqvist et Viksell, 466-474.

GÖRDING, E. (1977), *Vergleichende Studien zur Prosodie schwedischer Dialekte*, Phonetics Laboratory Department of General Linguistic, Cambridge, Cambridge University Press.

KRISTOL, A. M. (2000), L'Atlas linguistique audiovisuel du Valais romand (ALAVAL). Etat des travaux», in *Nouvelles du Centre d'Etudes francoprovençales Rene Willien* 41, 50-65 Saint-Nicolas (Aoste) (avec Federica Diémoz et Raphaël Maître).

LAI, J. P. (2002), *L'intonation du parler sarde de Nuoro*, Thèse de Doctorat en Sciences du Langage, Université Stendhal, 2 vol.

LAI, J. P. (2005), Etude de l'intonation d'un parler occitan : Gap, 2e Séminaire International du Projet AMPER (Grenoble 2004), in Actes, Grenoble, Centre de Dialectologie/ELLUG, Hors série 3 de *Géolinguistique*, 65-86.

LAI, J.P.; ROMANO, A.; ROULLET, S. (1997), Analisi dei sistemi prosodici di alcune varietà parlate in Italia: problemi metodologici e teorici, in *Bollettino dell'Atlante Linguistico Italiano*, 21, 23-70.

LÉON, P.; LÉON, M. (1964), *Introduction à la phonétique coirrective*, Paris, Hachette.

MARQUET, I. (1995), *Atlas linguistique parlant d'une région des Alpes occidentales*, Thèse de Doctorat, Grenoble, Université Stendhal, Centre de Dialectologie (l'Atlas est consultable sur Internet, sur le site de l'Université Stendhal).

MARTIN, Ph. (1975), Eléments pour une théorie de l'intonation, in *Rapport de l'Institut de phonétique de Bruxelles*, 9/1, 97-126.

MARTIN, Ph. (1979), Sur les principes d'une théorie syntaxique de l'intonation, in LÉON P. et ROSSI M. (Dir.), *Problèmes de prosodie*, vol. I, *Approches théoriques*, Studia phonética, 17, Ottawa, Didier (Canada), 91-101.

MARTIN, Ph. (1981), Pour une théorie de l'intonation. L'intonation est-elle une structure congruente à la syntaxe? in ROSSI M., DI CRISTO A., HIRST D., MARTIN Ph., NISCHIMURA Y. (Ed.), *L'intonation. De l'acoustique à la sémantique*, Paris, Klincksieck, 133-152.

MÉDÉLICE, J. E. (2002), Atlas Linguistique Multimédia de la Région Rhône-Alpes (ALMURA). Présentation du projet et avancement des travaux, in *Bollettino dell'Atlante Linguistico Italiano,* 26, 187-195.

RHARDISSE, A. (1995), *L'acent et l'intonation du parler de Bielsa (Aragon. Espagne)*, Thèse de Doctorat en Sciences du Langage, Université Stendhal/Centre de Dialectologie, Grenoble, 2 vol.

ROMANO, A. (2005), Utilisation des données *AMPER* pour une description de la variation linguistique. Tests de perception et contrôles statistiques 2e Séminaire International du Projet AMPER (Grenoble 2004), Actes, Grenoble, Centre de Dialectologie/ELLUG, Hors série 3 de *Géolinguistique*, 39-64.

ROMANO, A. (1999), *Analyse des structures prosodiques des dialectes et de l'italien régional parlés dans le Salento (Italie): approche linguistique et instrumentale.* Thèse de Doctorat

en Sciences du Langage, Université Stendhal Grenoble/Centre de Dialectologie, 2 vol.

ROSSI, M. (1974), *Contribution à la méthodologie de l'analyse linguistique avec application à la description phonétique et phonologique du parler de Rossano*, Italie, Paris, Champion, 3 tomes.

ROSSI, M. (1989), Ordre, organisation et intonation, in *Mélanges de Phonétique générale et expérimentale offerts à Pela Simon*, Strasbourg, Publications de l'Institut de Phonétique de Strasbourg, 715-733.

ROSSI, M. (1999), *L'intonation. Le système du français: description et modélisation*. Gap, Ophrys.

ROULLET, S. (1999), *Accent et intonation dans deux parlers francoprovençaux de la Vallée d'Aoste (Sarre et Cogne*, Thèse de Doctorat en Sciences du Langage, Université Stendhal, Grenoble, 2 vol.

PIKE, K. (1945), *The Intonation of American English*, Ann Arbor, University of Michigan.

COMUNICACIÓNS

EL TONEMA EN FRASES DECLARATIVAS –CON Y SIN EXPANSIÓN– EN EL HABLA GRANADINA

MARI CRUZ AMORÓS CÉSPEDES
Colegio San Alberto Magno. Alicante mcamoros@terra.es

1. INTRODUCCIÓN

En trabajos anteriores (Amorós 2003, Pamies & Amorós 2003, Amorós 2004) estudiamos los casos de *overshooting* en palabras agudas, llanas y esdrújulas en un corpus creado para este fin con hablantes de tres regiones diferentes: Granada, Madrid y Salamanca. Los resultados obtenidos mostraban que los casos de postrealización del acento eran insignificantes en las palabras agudas (sin estar al final de frase), en las palabras llanas el pico se realizaba en la sílaba postónica y en las palabras esdrújulas el pico tonal también se desplazaba hacia la derecha pero no en la primera postónica sino en la segunda postónica, es decir, al final de la palabra. Sin embargo, este comportamiento se veía muy determinado por la posición que la palabra estudiada ocupaba en la frase, más específicamente si se trataba del pretonema o el tonema.

2. OBJETIVOS Y METODOLOGÍA

El objetivo principal de este trabajo es estudiar en el tonema la relación existente entre dos factores suprasegmentales -el acento y la entonación- y el plano morfológico y sintáctico del discurso. Para ello, hemos aislado una parte específica del discuso: el tonema en frase declarativa. Analizamos el comportamiento de la curva melódica y del acento (prominencia tonal) en tres tipos de tonemas: agudos, llanos y esdrújulos en frases sin expansión en el predicado vs. con expansión en el predicado.

Así, tenemos tres palabras diferentes (*obsesión, paciencia y pánico*) analizadas en dos posiciones diferentes: "La cítara se toca con obsesión" y "La cítara se toca *con obsesión* y con amor". En este último grupo de frases, el tonema está formado por la expansión, que a su vez puede ser aguda, llana o esdrújula (*y con amor, finito/a, práctico/a*). Ello permite

comparar el pretonema y el tonema con la misma palabra, así como el núcleo gramatical con la expansión.

El número de frases analizadas ha sido de 63 (36 sin expansión en el predicado y 27 con expansión en el predicado) en tres repeticiones diferentes. Los valores analizados han sido la frecuencia fundamental en tres puntos de cada vocal (inicial, central y final) y los tonos de juntura inicial y final en todas las frases. Después se ha realizado una media entre las tres repeticiones de cada informante. Para la elaboración de los gráficos se ha tenido en cuenta (a parte de los tonos de juntura inicial y final) los valores centrales de F_0 de todas las vocales.

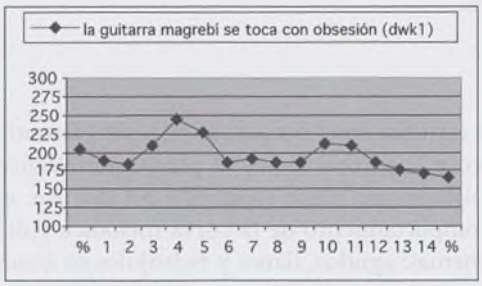
Las informantes han sido los mismos que en Pamies & Amorós 2005: dos jóvenes de Granada (Granada capital y El Padul), y las frases también proceden del corpus fijo de AMPER. Las grabaciones se han realizado con un DAT (Casio-DA1 con micrófono Shure SM58), capturadas y amplificadas con *Goldwave 4.25*; posteriormente segmentadas y analizadas con *Matlab 4.0* para el análisis automático y comprobadas con *Speech Analyzer 2.4* para el análisis manual de la F_0.

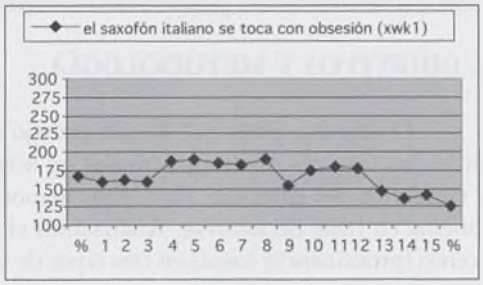
3. RESULTADOS

A- FRASES SIN EXPANSIÓN EN EL PREDICADO[1]

TONEMA AGUDO "... con obsesión."

- (GR): En todos los casos el tono ha descendido en escalera de forma progresiva (ejemplo dwk) menos dos casos de tónica más alta que la pretónica.
- (P): En 4/12 casos el tono ha descendido en escalera de forma progresiva; en 6/9 casos se ha producido sincronización entre el pico tonal y la sílaba tónica, ejemplo xwk; en 2/9 casos el valor de la tónica ha sido el mismo que el de la pretónica.

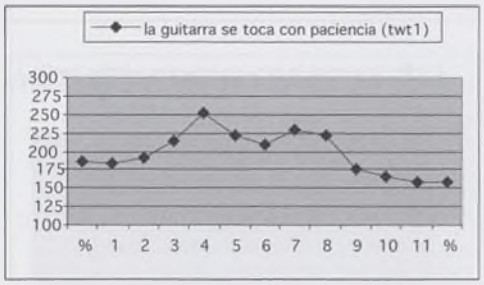


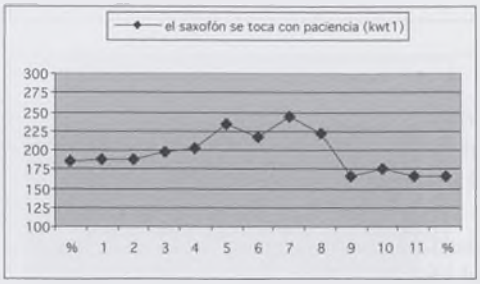


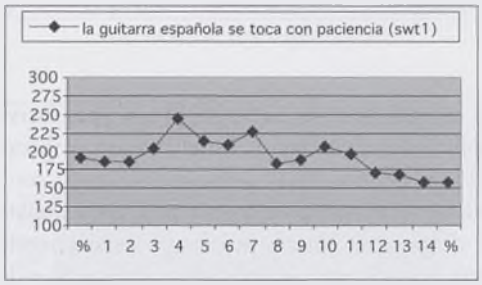
1 GR: Granada capital; P: El Padul

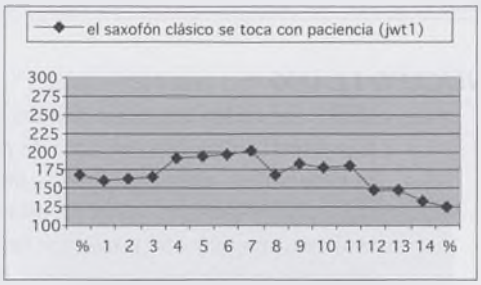
TONEMA LLANO "... con paciencia."

- (GR): En la mitad de los casos el tono ha descendido en escalera de forma progresiva (ejemplo twt); en la otra mitad se ha producido una sincronización entre sílaba tónica y pico tonal (ejemplo kwt).
- (P): En 5/12 casos el tono ha descendido en escalera de forma progresiva (ejemplo swt); en 7/9 casos el valor de la tónica ha sido el mismo que el de la pretónica y ha bajado en la postónica (ejemplo jwt).









TONEMA ESDRÚJULO "... con pánico."

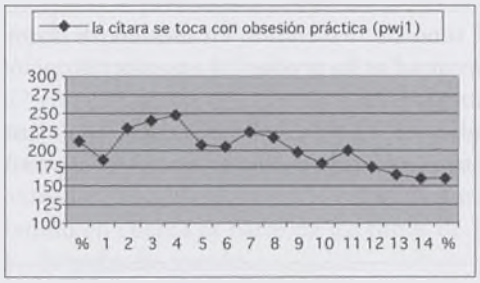
- (GR): En todos los casos el tono ha descendido en escalera de forma progresiva.
- (P): En todos los casos el tono ha descendido en escalera de forma progresiva.

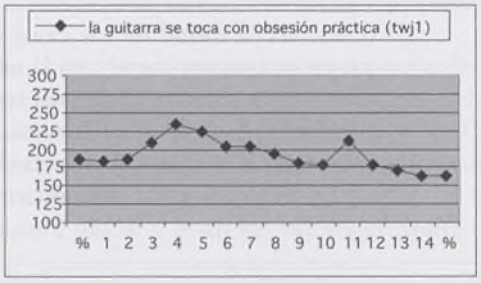
B- FRASES CON EXPANSIÓN EN EL PREDICADO

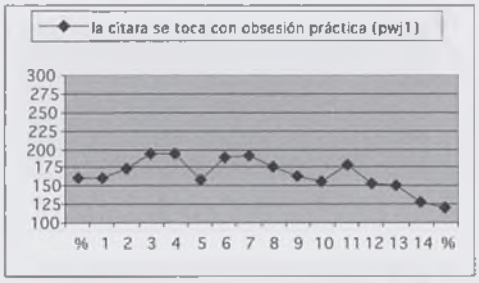
B1- ANÁLISIS DEL NÚCLEO GRAMATICAL:

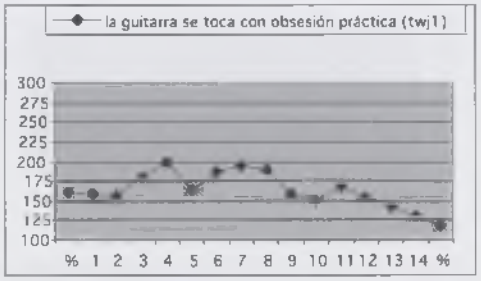
NÚCLEO AGUDO "... con obsesión..."

- (GR): En todos los casos menos uno ha habido un pico tonal muy definido en sincronización con la sílaba tónica (p.ej.: pwj, twj), posición que coincide también con el final de palabra.
- (P): En todos los casos ha habido un pico tonal muy definido en sincronización con la sílaba tónica (p.ej.: pwj, twj).



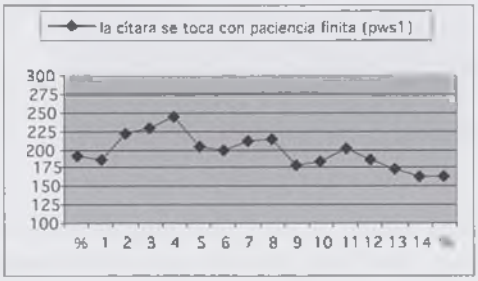


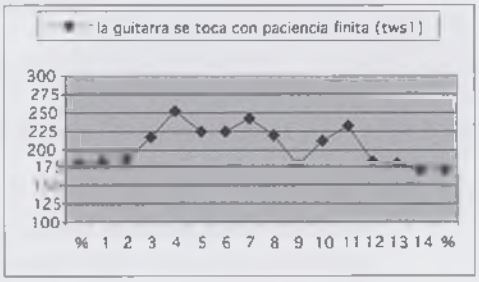


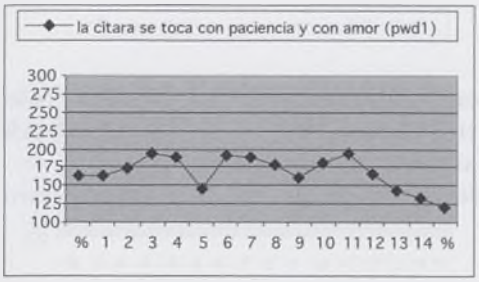


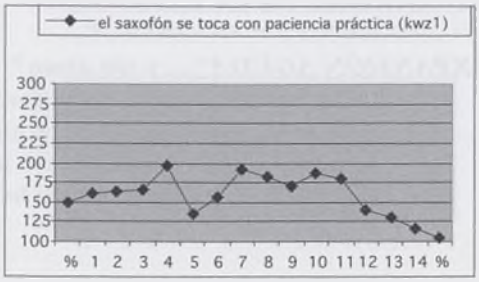
NÚCLEO LLANO "... con paciencia..."

- (GR): En todos los casos el tono ha ascendido en escalera de forma progresiva hasta realizar un pico en final de palabra (inicia la bajada tonal cuando empieza la expansión); ejemplos: pws1, tws1.
- (P): En 7/9 casos el tono ha ascendido en escalera de forma progresiva hasta realizar un pico en postónica final de palabra (inicia la bajada tonal cuando empieza la expansión); ejemplo: pwd. En 2/9 casos el tono de la tónica ha sido más alto que el de la pretónica (y el de la postónica) realizándose un pico en "cien" (ejemplo kwz).



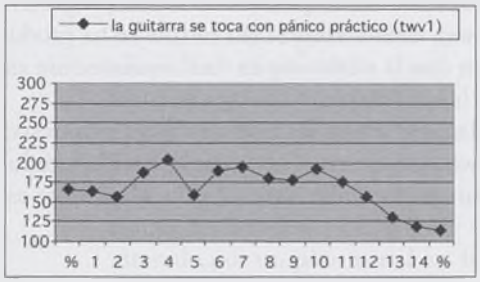


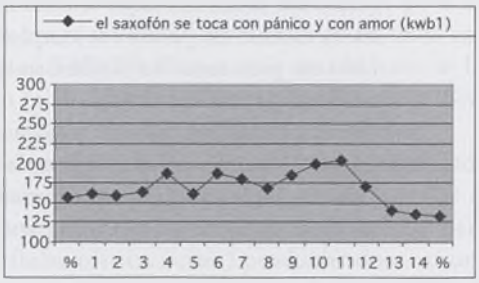


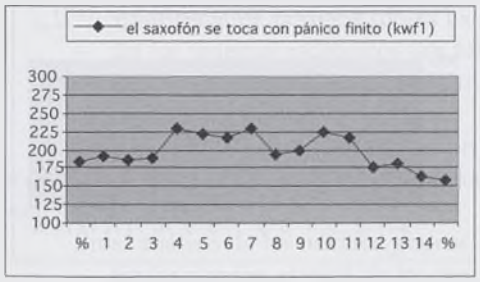


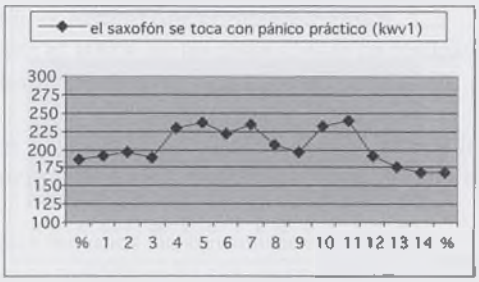
NÚCLEO ESDRÚJULO "... con pánico..."

- (GR): En 4/9 casos el tono ha ascendido en escalera de forma progresiva hasta realizar un pico en final de palabra (ejemplo: kwv); en 5/9 casos el pico se ha producido en la primera postónica "ni" y la bajada tonal se ha iniciado en la segunda postónica "co" hasta el final de la frase (ejemplo kwf).
- (P): En 6/9 casos el tono ha ascendido en escalera de forma progresiva hasta realizar un pico en final de palabra (ejemplo kwb); en 3/9 casos el pico se ha producido en la sílaba "ni" y la bajada tonal se ha iniciado en "co" hasta el final de la frase (ejemplo twv).





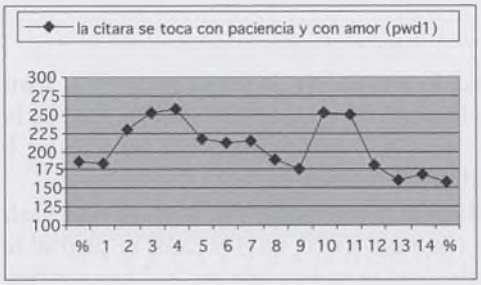


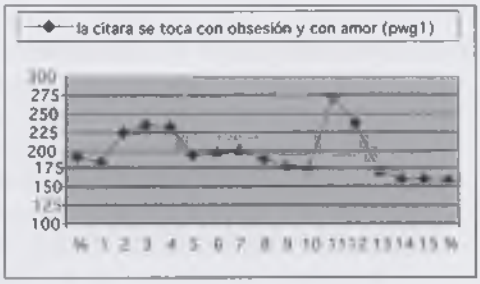


B2- ANÁLISIS DE LA EXPANSIÓN:

EXPANSIÓN AGUDA"... y con amor."

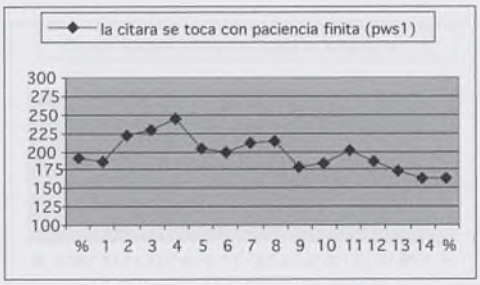
- (GR): En 4/9 casos el tono ha descendido en escalera de forma progresiva hasta el final de palabra (ejemplo pwg); en 5/9 casos se ha producido un moderado pico tonal en sincronización con la sílaba tónica "mor" (ejemplo pwd).
- (P): En todos los casos el tono ha descendido en escalera de forma progresiva hasta el final de palabra.

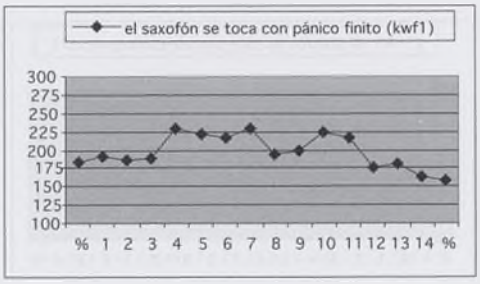


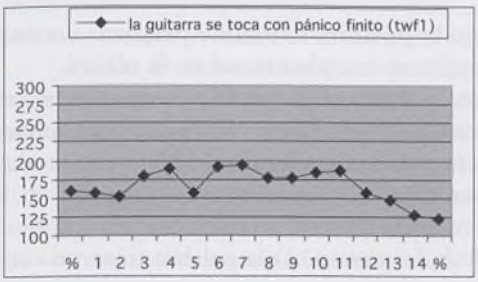


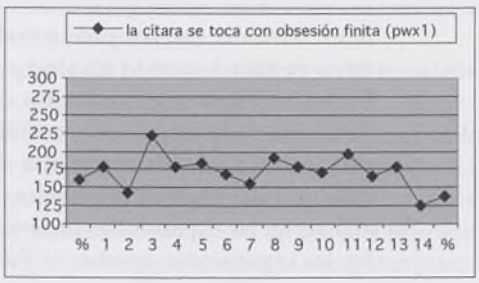
EXPANSIÓN LLANA "... finito."

- (GR): En todos los casos el tono ha descendido en escalera de forma progresiva hasta el final de palabra (ejemplo pws), menos en un caso en que se ha producido un pico tonal en sincronización con la sílaba tónica "ni" realizándose así un final circunflejo (kwf).
- (P): En todos los casos el tono ha descendido en escalera de forma progresiva hasta el final de palabra (ejemplo twf), menos en un caso en que se ha producido un pico tonal en sincronización con la sílaba tónica "ni" realizándose así un final circunflejo (ejemplo pwx).









EXPANSIÓN ESDRÚJULA "... práctico."

- (GR): En todos los casos el tono ha descendido en escalera de forma progresiva hasta el final de palabra.
- (P): En todos los casos el tono ha descendido en escalera de forma progresiva hasta el final de palabra.

4. DISCUSIÓN

Como hemos podido observar, el comportamiento de los tonemas ha presentado unos resultados similares para ambas categorías sintácticas.

- En los tonemas agudos con predicados sin expansión (*con obsesión*): el 59% de los casos (14) se ha realizado con un contorno tonal **descendente** iniciado con un pico en el proclítico "con" y en el 33% (8) el contorno descendente se ha visto alterado por un **pico tonal en la sílaba tónica** (*-sión*). En dos casos (8%) el valor tonal de la sílaba tónica ha sido el mismo que el de la pretónica.
- En los tonemas llanos (*con paciencia*): el 46% de los casos (11) se ha realizado con un contorno tonal **descendente** desde un pico en la sílaba anterior al inicio del tonema y en el 29% (7) el contorno tonal descendente se ha visto alterado por un **pico tonal en la sílaba tónica** (*-cien-*) produciéndose así una sincronización pico tonal – acento y un final circunflejo. En 6 casos (25%) el valor tonal de la sílaba tónica ha sido el mismo que el de la pretónica.
- En los tonemas esdrújulos (*con pánico*): el 100% de los casos (24) se ha realizado **con un contorno tonal descendente** desde un pico en la sílaba anterior al inicio del tonema.

En los finales de frase con expansión en el predicado debemos distinguir entre los núcleos gramaticales en posición inmediatamente previa al tonema; y las expansiones en posición de tonema.

En el análisis de los núcleos gramaticales tenemos las siguientes diferencias:

- En los agudos, el 94% de los contornos melódicos ha sido **ascendente** y ha terminado con un pico tonal en la sílaba tónica: (que es también final de palabra); en 1 caso (6%) el contorno melódico ha sido descendente hasta el final de frase desde el pico tonal en el proclítico "con".

- En los llanos, el 89% (16) de los contornos melódicos ha sido **ascendente** hasta terminar con **un pico tonal en la postónica final de palabra** (*overshooting*); en sólo 2 casos (11%) se ha producido un **pico tonal en la tónica**.
- En los esdrújulos, en todos los casos (100%) se ha producido *overshooting*: el 56% (10) de los contornos melódicos ha sido ascendente hasta terminar con un **pico tonal en la última sílaba de la palabra**; en 8 casos (44%) se ha producido un pico tonal en la primera postónica "ni".

El análisis de las expansiones también muestra diferentes relaciones:

- En las expansiones agudas, el 72% de los casos (13) la palabra tiene un contorno tonal **descendente continuo** iniciado en un pico en la sílaba anterior; en el 28% (5) el contorno tonal descendente se ha visto alterado por un **pico tonal en la sílaba tónica** "mor".
- En las expansiones llanas, el 89% de los casos (16) hay un **contorno tonal descendente** iniciado desde un pico en la sílaba anterior al tonema; en el 11% (2) el contorno tonal descendente se ha visto alterado por **un pico tonal en la sílaba tónica** produciéndose así una sincronización pico tonal - acento.
- En expansiones esdrújulas el 100% de los casos (18) se ha realizado con un **contorno tonal descendente** desde un pico anterior al tonema.

5. CONCLUSIONES

El presente trabajo parece confirmar que:

A. El contorno tonal en el pretonema de las palabras agudas es ascendente con el pico en la tónica final en el 94% de los casos (esta posición también es frontera final de palabra). En las palabras llanas el contorno tonal es ascendente y el pico está en la sílaba postónica (89% de los casos) que también es final de palabra. En las palabras esdrújulas en el 100% de los casos el contorno tonal es ascendente y el pico recae en la sílaba postónica (56% en postónica final de palabra y 44% en la primera postónica).

B. En el tonema, en cambio, la posición del acento en la palabra sí parece haber influenciado la curva tonal. El patrón más común ha sido una caída tonal desde un pico en la sílaba anterior al inicio del tonema, descendiendo hasta el final de frase, pero hemos observado 8 casos (33%) de un pico tonal en sincronización con la sílaba tónica en las palabras agudas y 7 casos (29%) de un pico tonal en sincronización con la sílaba tónica en las palabras llanas. En las palabras esdrújulas no ha habido ningún caso de sincronización, pues en todas ellas el tono cae desde un pico anterior, situado en la sílaba anterior al inicio del tonema.

C. Este mismo comportamiento se ha repetido en los tonemas de las frases con expansión del predicado. La entonación de dichas expansiones en el tonema también se muestra sensible al lugar del acento. En general todas las expansiones se han realizado con un contorno tonal descendente, salvo 5 casos (28%), en que hay un pico posterior en sincronización con la sílaba tónica en las palabras agudas, y 2 casos (11%) de un pico posterior en sincronización con la sílaba tónica en las palabras

llanas[2]. En las palabras esdrújulas no ha habido ningún caso de sincronización: el tonema siempre es descendente a partir de un pico anterior.

Dicho de otro modo: en el pretonema, sólo la posición sintáctica de la palabra analizada parece influir en el contorno melódico de la misma. En el tonema, el acento sí parece influir en el contorno tonal, aunque de una manera sólo parcialmente predecible: si el tonema es agudo tenemos más o menos 30% de posibilidades de pico tonal en la sílaba tónica; si el tonema es llano tenemos cerca de un 20% de posibilidades de que se produzca un pico tonal en la sílaba tónica; si el tonema es esdrújulo: raramente se producirá sincronización entre pico tonal y sílaba tónica (parece que la cadencia empieza justo a la izquierda de la palabra nuclear y el tono ya no vuelve a subir). En cambio, si la misma palabra pasa al pretonema porque se añade una expansión a la frase, el contorno melódico de esta palabra se vuelve ascendente y el pico tonal se realiza en la sílaba tónica si la palabra es aguda, en la sílaba postónica si la palabra es llana, y en segunda postónica si la palabra es esdrújula, coincidiendo con la frontera entre el núcleo del sintagma y su expasión.

BIBLIOGRAFÍA

AMORÓS M.C. (2003): *La relación entre entonación y acento en español*. Tesis doctoral (inédita) Dept. de Lingüística General, Universidad de Granada.

AMORÓS, M.C. (2004): "Sincronización entre pico tonal y acento: Resultados según posición métrica y morfológica" en *Estudios de Fonética Experimental*, XIII, Barcelona, pp. 203-223.

CALLEJA AZPIAZU, N. (2004) "Alineamiento fonético de acentos tonales", *Estudios de Fonética Experimental*, XIII, pp. 39-64.

DORTA, J & HERNÁNDEZ, B. (2004) "Prosodia de las oraciones SVO declarativas e interrogativas en el español de Tenerife", *Estudios de Fonética Experimental*, XIII, pp. 225-274.

PAMIES, A. & AMORÓS, M.C. (2003): "The relation between stress and tonal peaks" en Solé, M.J., Recasens, D. & Romero, J. (eds.) *15th ICPhS (Proceedings of the 15th International Congress of Phonetic Sciences)*; International Phonetic Association & Universitat Autónoma de Barcelona. 3-6 de Agosto de 2003.

PAMIES, A. & AMORÓS, M.C. (2005): "Pico tonal, acento y fronteras morfosemánticas: experimento con hablantes granadinos", en *Estudios de Fonética Experimental*, XIV, Barcelona, pp. 201-223.

PAMIES BERTRÁN, A.; FERNÁNDEZ PLANAS A.M.; MARTÍNEZ CELDRÁN, E.; ORTEGA ESCANDELL A.; AMORÓS CÉSPEDES, M.C.: 2002 "Umbrales tonales en español peninsular", en Díaz García, J. *Actas del II Congreso Nacional de Fonética Experimental*. Sevilla: Universidad, pp. 272-278.

2 Sin embargo hay que admitir que estos picos han sido muy débiles y no han llegado al margen de +9% que se establece en Pamies *et al.* 2002. La media está en tan sólo +4hz de diferencia.

APROXIMACIÓN Á ENTOACIÓN DOS ENUNCIADOS DECLARATIVOS E INTERROGATIVOS EN TRES ÁREAS DIALECTAIS DA ITALIA CENTRO-MERIDIONAL (ABRUZZO, BASILICATA E CAMPANIA)

FRANCESCO AVOLIO
Dip. di Storia e metodologie comparate - Università dell'Aquila, Italia avolio@cc.univaq.it
ANTONIO ROMANO
Dip. di Scienze del Linguaggio - Università di Torino, Italia antonio.romano@unito.it

Esta contribución realizouse principalmente ao abeiro do proxecto denominado *AMPER*, iniciado recentemente. No seo dos estudos prosódicos, descritivos e dialectais, realizáronse algúns traballos sobre a entoación dos dialectos de Italia e das variedades rexionais do Italiano (*AMPER-ITA*). Así mesmo, neste proxecto tamén se pretende proceder á elaboración de *corpora* comparables doutras variedades románicas nunha orientación teórica unitaria, para o que se recolleron gravacións e se realizou a dixitalización dalgúns ficheiros.

A investigación en curso enmárcase no seo dunha interesante experiencia de colaboración que leva a cabo unha análise das configuracións dos enunciados declarativos e interrogativos nunhas áreas dialectais da Italia centro-meridional. As gravacións fixéronse de maneira individual polos dous autores nas distintas rexións, dentro da tradición dos estudos dialectais e co obxectivo dunha descrición xeral da entoación destas áreas. Os enunciados recollidos cumpriron varios requisitos: ter a mesma estrutura sintáctica, a mesma modalidade oracional, o mesmo número de sílabas e as mesmas posicións acentuais nas palabras.

Aínda que tamén se presenten datos sobre a duración, neste estudo de carácter fonético e de pretensións eminentemente descritivas, préstase especial atención á frecuencia fundamental que parece seguir contornos distintos dentro dun *continuum* no que se apunta a diferenza dialectal entre unha rexión do Appennino Lucano (Aliano, San Mauro Forte) e o resto dos dialectos da área meridional xa considerados (Ravello, na área do napolitano meridional, e Tèramo, no Abruzzo setentrional).

PRESENTACIÓN DAS VARIEDADES CONSIDERADAS

Este estudo presenta os primeiros resultados dunha interesante experiencia de colaboración na que se leva a cabo unha análise das configuracións dos enunciados declarativos e interrogativos nunhas áreas dialectais da Italia centro-meridional.

Nunha primeira enquisa de ensaio, as variedades foron escollidas de xeito que representaran tendencias prosódicas probablemente diferentes segundo as descricións tradicionais dos estudos dialectais destas rexións[1].

As variedades consideradas seguen a codificación definida no marco de *AMPER-ITA* que aquí presentamos por primeira vez (v. fig. 1). Desde o punto de vista da análise dialectolóxica dos trazos xerais, poden ser definidas como segue.

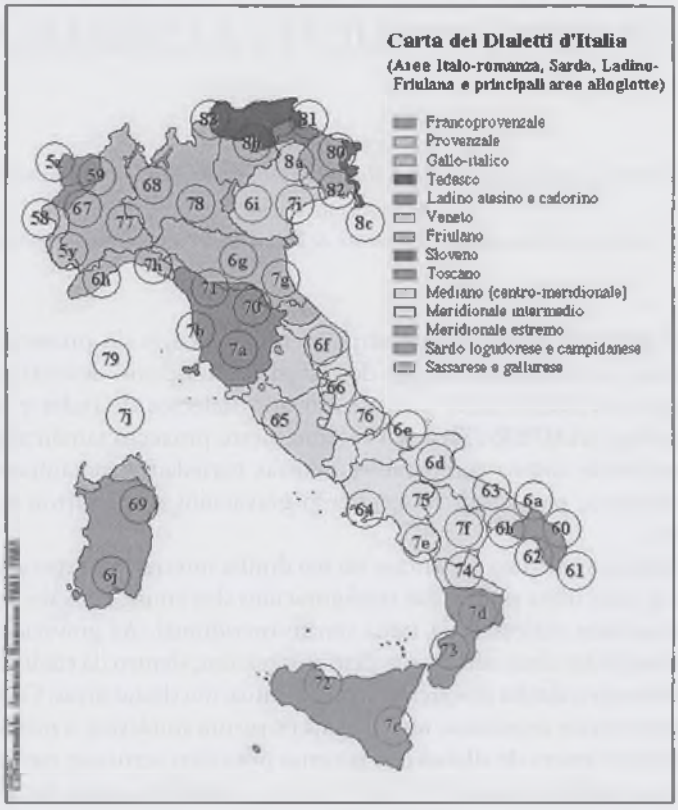


Fig. 1. Codificación das áreas da enquisa de *AMPER-ITA*.

1 AVOLIO (1995). Unha selección dos títulos máis directamente implicados debería ter en conta ademais AVOLIO (2000, 2002 e 2003).

Aliano (ALN). Variedade lucana extremadamente conservadora, con vocalismo tónico de tipo "sardo", localizada no límite set. da área máis conservadora e illada nunha zona de vocalismo distinto (Missanello e Alianello teñen un vocalismo de tipo balcánico, Stigliano ten un vocalismo napolitano con isocronismo de sílaba tónica, como San Mauro e case toda a provincia de Matera). Conserva tamén -T e -S finais (este con oscilacións) nunhas formas verbais (*càccete* 'sae (a algo)', *cantese* 'cantas', e polo menos hoxe, *vaj* 'vas', *faj* 'fas'), J- orixinario (*jettà* 'tirar' vs. *scettà* de Stigliano, San Mauro, etc.) e o neutro nos demostrativos (condicións sanmaureses: *quist* vs. *chist*). Coma o dialecto de Tursi, non presenta ditongos; pódese pensar que o pobo asimilara a influencia dunha corrente de metafonía sabina (*statte bbone* 'que esteas ben', con *o* tendencialmente pechado) que quizais non se estabilizou por mor das historias particulares do vocalismo tónico (fusión de E longo con E curto, O longo con O curto). De tódolos xeitos, profundar na súa posición xeolingüística pode ser revelador para comprender mellor a posición lingüística e histórica real da Lucania meridional no contexto da Italia centro-meridional.

Ravello (RAV). Variedade periférica do napolitano, definida na clasificación de AVOLIO (2000) como "napoletano meridionale", e caracterizada, no plano morfolóxico, por unha extensión do "redobramento fonosintáctico" (*RF*), típico do neutro, a moitos nomes orixinariamente de xénero feminino (*o llane* 'a lá', *o nnotte* 'a escuridade'); pero tamén, no plano fonético, pola asignación das formas adxectivas ditongantes metafonéticas a nomes femininos plurais (*e mméle cuotte* 'as mazás cocidas', *e scarpe viécchie* 'os zapatos vellos', en contraposición ao nap. *e mmele cotte, e scarpe vecchie*, etc.).

San Mauro Forte (SMF). Variedade lucana oriental de influencia *barese*, con isocronismo de sílaba tónica[2], forte ditongación (*ruptura*) de vogais tónicas en sílaba aberta (*matéure* 'madura'), incluídas as de orixe metafonética (*cugghièune* 'collóns'), retraemento dos ditongos metafonéticos napolitanos (*pide* 'pés', *nuve* 'novo, -os') e inserción de *u*, sobre todo cando a tónica é *a* ou *e* (*le cuane* 'o can', *le puépe* 'a pementa'). Aparece o neutro románico nos demostrativos, pero non nos artigos determinados (*cust* 'esta persoa' vs. *chist* 'esta cousa').

Tèramo (TER). Variedade do Abruzzo adriático ou oriental, caracterizada por un claro predominio de vogais abertas, de calquera etimoloxía e estrutura silábica (*pale* 'pelo', *frande* 'fronte'), determinada tamén pola fusión entre vogais orixinarias diferentes, metafonía só por -I (*tatte* 'tellado', *titt* 'tellados', *pòrch* 'porco', ma *purc'* 'porcos'), ditongos napolitanos monotongados (*pide* 'pés'), desenvolvementos particulares do grupo L + cons. (con frecuencia hai asimilación, e tamén africación ou lenición: *pòzze* 'pulso', *faggiòne* 'falcione'), neutro románico só nos demostrativos (*qualle* 'aquela persoa', *culle* 'aquela cousa'), construcións impersoais interesantes (*ôme dice* 'se diz'), etc.

---

2 Esta isocronía reflíctese tamén na selección dos timbres das vogais medias, os cales, independentemente das oportunas excepcións, se presentan con distribución complementaria: (extra)aberto en sílaba pechada, alto e sostido en sílaba aberta.

MATERIAIS ANALIZADOS

As gravacións realizáronse de maneira individual polos dous autores nas distintas rexións. Os enunciados recollidos debían cumprir varios requisitos: ter a mesma estrutura sintáctica, a mesma modalidade oracional, o mesmo número de sílabas e as mesmas posicións acentuais na estrutura silábica.

Os locutores son dous homes PD37 (Aliano) e EC41 (San Mauro Forte), e dúas mulleres EB78 (Ravello) e MF60 (Tèramo).

O corpus de proba, gravado en cintas magnéticas e despois dixitalizado en formato.wav, está baseado principalmente en expansións adxectivas dos grupos de palabras das catro estruturas seguintes (18 oracións nas dúas modalidades):

Saverio era un bambino strano./?
Antonio canta una canzone./? (exp. SN_2 -> antica)
Donato faceva il caffe./? (exp. SN_2 -> dolce, torbido)
Donato mangia una pesca./? (exp. SN_2 -> matura, acerba)

No proxecto inicial pretendíase recoller ata cinco repeticións da mesma oración por locutor. Ademais dunha variabilidade na escolma dos elementos lexicais, varios locutores escolleron tamén diferentes opcións de orde (por exemplo, inversión nas preguntas) e de focalización, causando así a presenza de datos incomparables.

Os resultados obtidos están fundados na análise dunha ou dúas repeticións (en poucos casos, tres). Os que aquí presentamos corresponden ás oracións máis repetidas[3].

UNHA PRIMEIRA LECTURA DOS DATOS ANALIZADOS

Desde o punto de vista da estruturación temporal e melódica, os nosos datos concordan con aqueles das descricións dialectolóxicas tradicionais (v. anteriormente): aínda que son escasos, podemos describir os materiais recollidos para as variedades consideradas co recurso a algúns parámetros indicadores empregados habitualmente. En particular, como elementos de discriminación resultaron importantes os fenómenos de ditongación e *ruptura* dos núcleos vocálicos e as características de isocronismo (acentual vs. silábico)[4].

---

3 Os *corpora* de *AMPER* son xeralmente redundantes (para permitir unha verificación da estabilidade e da persistencia dos fenómenos observados, mesmo en condicións de variación segmental). Neste caso, tratándose de materiais dunha enquisa preliminar, non se atopou esta característica. Todos os contextos investigados nos *corpora* de referencia seguen presentes na mostra, de xeito que, se tivésemos a certeza dun bo control das variables estilísticas e interpretativas, poderíamos tirar conclusións fiables dos fenómenos máis xerais.

4 En relación a este tema, véxase BERTINETTO (1989) e ROMITO & TRUMPER (1993). A lista dos indicadores experimentados para a clasificación tipolóxica corrixida das variedades segundo estes aspectos está en SCHMID (1996) e foi reutilizada para a confrontación entre variedades meridionais da fronteira ápulo-salentina por ROMANO (2003): pode destacarse a tendencia ao isocronismo acentual (agora reanalizado cos índices de RAMUS *et alii*, 1999), na que podería ser determinante a forte tendencia á elisión das vogais átonas ou a súa frecuente redución nun *schwa*.

As producións de Aliano (ALN) caracterízanse por unha estruturación prosódica que podemos definir como de transición, con algúns matices do *calabrese* setentrional. No conxunto das súas características, existe tamén unha maior aspiración de oclusivas e africadas. Os fenómenos entoacionais estudados ata aquí amosan certa orixinalidade en comparación coas outras variedades consideradas.

As características entoacionais de Ravello (RAV) confirmaron, grazas tamén á presenza do típico *RF* (v. anteriormente), a súa clasificación como *napoletano* meridional, que se presenta cheo de *ictus* e relacións dinámicas, xunto coa presenza dunha redución en *schwa* das vogais átonas (e redución enerxética das vogais tónicas en lugares de pouca prominencia). Algunhas curvas melódicas características parecen coincidir a primeira vista coas posicións onde a ditongación das vogais medias forma núcleos con retraemento (onde o ditongo ascendente tende ao hiato).

San Mauro Forte (SMF) confirma a súa pertenza a unha área con características que nos lembran nalgúns versos aos modelos entoacionais e acentuais da rexión de Bari. A presenza de *schwa* asóciase a unha preferencia pola cancelación das vogais non acentuadas fronte a unha maior predisposición ao alongamento dos núcleos afectados polo acento de frase ou de oración, os cales se presentan xeralmente quebrados.

Tèramo (TER) caracterízase por un modelo que pertence a unha área dialectolóxica distinta (cfr. co *romanesco*, 65, xa presente no proxecto *AMPER* grazas aos resultados dun ensaio preliminar)[5]. Os perfís melódicos observados achégannos definitivamente a algúns dialectos da área central (Lazio central, Umbria meridional, Marca picena) con particular atención ás propiedades entoacionais dos dialectos do Ascolano[6].

METODOLOXÍA DA ANÁLISE

O método de análise, xa experimentado e adoptado nos estudos prosódicos doutras áreas románicas incluídas no proxecto *AMPER*, está resumido no artigo recente de ROMANO *et*

---

5 En ROMANO (2005: 43) atópanse comentarios sobre as dúas solucións prosódicas utilizadas para as preguntas totais nestas dúas variedades da Italia central (tratábase naquel caso do italiano rexional, de Roma e de Tèramo). Sendo os dous locutores considerados "assez différents sur le plan de leur profil sociolinguistique", as curvas analizadas presentaban configuracións globais moi parecidas: a pesar dalgunhas diferenzas evidentes no aliñamento de obxectivos prosódicos específicos (e na extensión da gama de valores de F0 explorados), trazos distintos foron considerados tributables á realización de matices expresivos e dunha distinta forza elocutiva. A comparación realizouse naquela ocasión para mostrar como, contrariamente ao que se comproba nalgunhas áreas lingüísticas (onde os índices prosódicos poden ser a orixe dunha diverxencia entre variedades tradicionalmente consideradas partes dun mesmo espazo dialectal), noutros casos as mesmas variables poden determinar condicións de escasa variación, tamén en áreas longas e diferenciadas.

6 Esta área, que aínda queda prosodicamente inexplorada, rexístrase como unha área de fenomenoloxía meridional, ao longo da costa, polo menos ata o curso do Aso (cfr. VIGNUZZI, 1988). Tamén sería interesante a variación entoacional no Lazio central, onde conviven dialectos tipoloxicamente diferentes: *romanesco*, na banda intermedia ao oeste do Tevere, dialectos centrais (Sabina e Ciociaria) ao leste do mesmo. A Umbria meridional, e en particular a Valnerina e o Ternano, presentan tamén unha fenomenoloxía central, mais con gran coherencia sobre todo coas áreas intermedias próximas ao Lazio e do Abruzzo. Isto podería explicar, dalgún xeito, as súas semellanzas prosódicas co Teramano, que queda dialectoloxicamente meridional, e cos seus numerosos trazos idiosincráticos (cfr. AVOLIO, 2002).

alii (2005). Como xa se subliñou nese traballo, a análise destas estruturas necesita dunha pluralidade de aproximacións que deben centrarse en distintos niveis 'entonéticos' e 'entonolóxicos'. Mentres para estes parece necesaria unha maior reflexión abstracta de falantes nativos que centren a súa atención en investigacións de tipo endofásico das regras prosódicas da súa lingua, os primeiros deberían en lugar diso insistir tamén na importancia das simulacións e no recurso a un aparato de análise non soamente instrumental, senón tamén experimental.

En particular, no noso caso as tentativas de síntese da súa prosodia prototípica, aínda que limitada aos tipos de produción observados, permítennos centrar a atención naquelas relacións melódicas, temporais e dinámicas que se revelan máis interesantes para o obxectivo da caracterización xeoprosódica. As numerosas avaliacións que fixemos para definir as condicións de reprodución dos trazos típicos das áreas estudadas permitíronnos chegar a unha mellor descrición dos estereotipos entoacionais, coa inclusión de elementos habitualmente ignorados ou excluídos dos estudos neste sector (por ex. as contribucións de arrastramentos, explosións, etc.)[7].

DISCUSIÓN DOS RESULTADOS

De forma global podemos observar que o material recollido presenta características interesantes na oposición entre as declarativas e as interrogativas, e na variación entre as solucións entoacionais dialectais representadas.

Resumimos aquí brevemente algunhas das características máis importantes observadas (en relación coas curvas e os histogramas que se recollen ao final do artigo), concentrando a nosa atención nunha selección de elementos contrapostos. Esperamos os resultados de estudos máis completos para unha *prototipización* definitiva, que non foi posible ata agora por mor da variabilidade observada nos datos recollidos.

A dificultade na comparación pódese achacar, en parte, a unha inconstante elisión de vogais, á presenza de fenómenos de énfase sobre elementos distintos e de estratexias de *topicalización* que diverxen entre unha repetición e outra, mais sobre todo entre unha oración e outra[8]. Mesmo na realización dalgunhas preguntas "si/non", os enunciados foron moi diferentes pola presenza dunha distinta 'forza interrogativa', probablemente atribuíble a varias posibilidades de contextualización imaxinadas polo locutor.

Todo isto conduce á obxectiva imposibilidade de definir con precisión os efectos asociados á presenza das expansións e aos cambios de modalidade.

Aliano (ALN)

En particular, en canto á configuración global, ALN presenta uns esquemas interrogativos pouco diferentes dos correspondentes enunciativos, con percorridos e curvas

---

7 Desexamos que esta metodoloxía de avaliación, xa definida no curso da investigación en ROMANO (2001), continúe a ser o centro do método *AMPER* (v. tamén a presentación en ROMANO, 2005).

8 Algúns destes elementos de variabilidade xa existentes nos materiais de *AMPER-POR*, sinálanse tamén en DE CASTRO MOUTINHO *et alii* (2004).

melódicas que corresponden á dúas macroestruturas entoacionais presentes nos enunciados: o pretonema e o tonema. A distinción da modalidade parece asegurada principalmente pola velocidade local dos movementos melódicos (caídas máis acentuadas, perfís con maior velocidade sobre vogais máis breves) e pola diferente organización temporal que conduce á redución e a unha maior tendencia á cancelación de vogais non acentuadas (ca. 15%)[9].

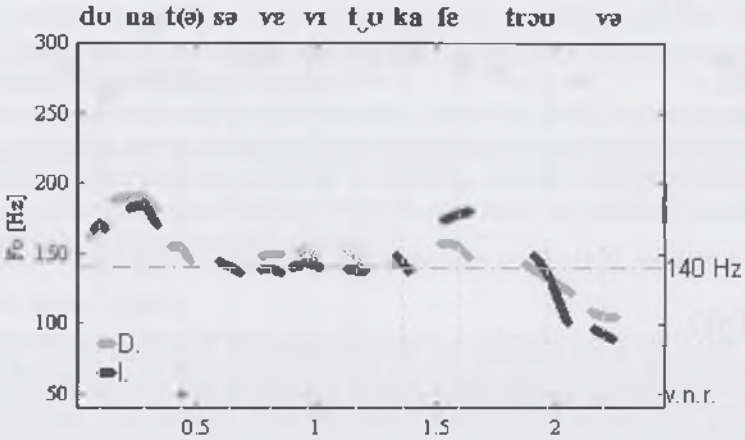


Fig. 2. ALN (loc. PD37): Curva esquemática de F_0 de "Donato si beveva il caffè tòrbido./?"

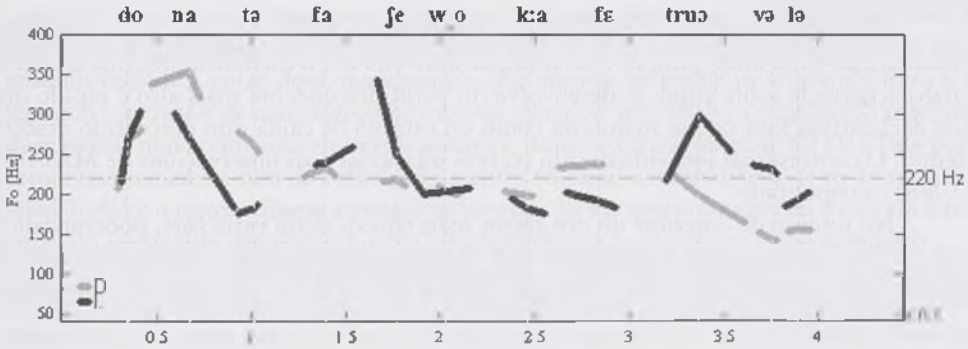


Fig. 3. RAV (loc. EB78): Curva esquemática de F_0 de "Donato faceva il caffè tòrbido./?"

9 Na comparación entre o *timing* de declarativas (D) e interrogativas (I), pódese observar que a única vogal de I que pode presentar unha duración maior cá vogal correspondente de D é a tónica do verbo.

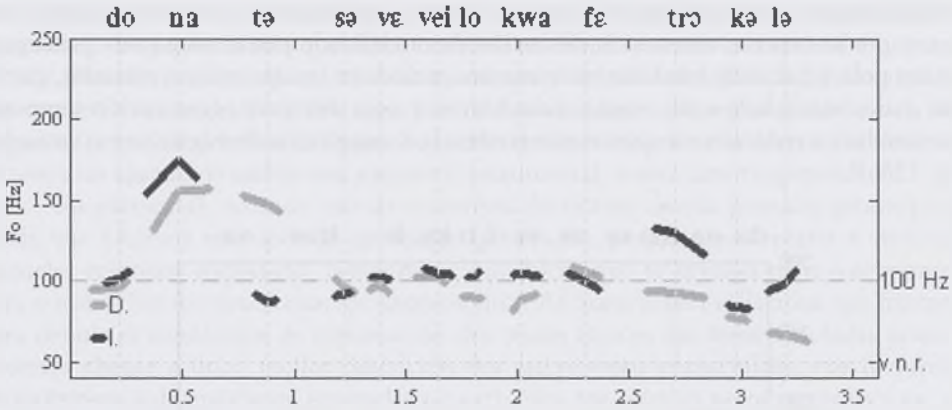


Fig. 4. SMF (loc. EC41): Curva esquemática de F_0 de "Donato si beveva il caffè tòrbido./?"

Ravello (RAV)

Os datos de RAV parecen ser os caracterizados por unha maior estabilidade. Presentan tres movementos melódicos porque a locutora deu relevo tamén ao verbo dos enunciados[10]. Este movemento resulta especialmente pronunciado nas interrogativas, mentres que se mostra atenuado (ou ausente) nas declarativas, nas que pola contra, é evidente unha xerarquización entre dous cumes que aparecen no contorno terminal de modalidade que se corresponden coas dúas últimas posicións prominentes presentes nos casos de expansión.

O tonema (asociado ao perfil final de modalidade, PFM) das preguntas totais aquí consideradas preséntase sempre cunha importante elevación melódica próxima á última sílaba acentuada sobre a que se desenvolve un perfil descendente máis alto e rápido que nas declarativas (nas que se manifesta como un cambio de caída nun percorrido descendente). O contraste así individualizado vólvese relacionar coas observacións de MATURI (1988) co napolitano.

No intento de describir un dos trazos máis típicos desta variedade, podemos dicir que o PFM interrogativo caracterízase por un pico principalmente aliñado co inicio da última vogal acentuada (UVA)[11], que vai seguida por un perfil descendente cun cambio

10 No caso desta localidade, en todos os exemplos de I do modelo obsérvase sobre a tónica do verbo, e ademais do aumento da duración, un pico e unha excursión melódica moi significativas á que segue unha caída improvisada de aproximadamente 150 Hz aliñada entre a tónica e a postónica.

11 Cfr. co "pico sobre a última tónica característico das si/non" de MATURI (1988: 24). Compre observar que se a última preacentuada da pregunta non está tampouco acentuada, a súa altura típica se inscribe de forma bastante estable, no caso da locutora considerada, nun nivel medio-baixo, con valores aproximadamente de 20 Hz por debaixo da frecuencia media das frases do modelo utilizado. Constitúe un dos datos máis estables observados.

antes da última vogal non acentuada (que se realiza nun nivel baixo). A falta desta (neste caso o último elemento léxico das frases é tónico) e en presenza dun fenómeno de alongamento, a solución observada é a do truncamento.

Reproducindo con estes materiais o xuízo de MATURI (1988) cando di que "a oposición entre os contornos terminais nos dous tipos de frases, que ás veces é cualitativa (contorno terminal ascendente vs. contorno terminal descendente), aquí se basea só na súa diferenza cuantitativa (contorno terminal máis vs. menos ascendente)", non podemos non ter en conta as importantes distincións en termos de configuración do PFM.

En canto á cancelación das vogais non acentuadas, rexistramos para esta variedade unha taxa do 13% aproximadamente.

Pero é sobre todo nas producións desta variedade onde a presenza de fricativas asume unha función particularmente destacable (ou de oclusivas non seguidas de vogais, cuxa explosión parece adquirir unha cor "vocálica") e invita a considerar definicións máis xerais de prosodia (como a de FIRTH, 1948) na que non se exclúen as achegas de elementos consonánticos e elementos de transición[12].

San Mauro Forte (SMF)

Neste exemplo, rexistramos ante todo unha tendencia á economía das vogais finais das frases declarativas, sobre todo se non se acentúan. Na segunda das pausas realizadas (ou das posicións de maior probabilidade) observamos tamén unha notable redución das vogais non acentuadas internas. Nalgunhas series de repeticións observamos a aparición de musicalidade e entoación escolásticas (típicas dunha recitación histriónica) que prevalecen na valoración dos perfís en determinadas condicións. Utilizáronse tres estratexias diferentes para resolver a frase: diferentes modelos de creación de temas do suxeito que provocan o cambio do perfil inicial e que puideron influír no desenvolvemento de toda a entoación da frase. Nas declarativas, cando domina a estrutura, na frase con expansión establécese unha relación melódica entre dous movementos que afectan ao primeiro acento da frase e ao segundo (de frase e de oración), en función da distancia intersilábica (separación mediante dúas sílabas non acentuadas en *perkoke mateure* _ dminución melódica de 80 Hz). Este perfil consérvase tamén en caso de colisión de acentos (distancia intersilábica=0 en *kwafe dolce* e *kwafe trokele*): o resto redúcese a aproximadamente 30 Hz nestas condicións e proporciona á curva un percorrido descendente xa en correspondencia coa primeira vogal acentuada.

Respecto ao PFM de RAV, neste modelo a última preacentuada sitúase nun nivel medio, o salto melódico e o descenso conseguinte sobre a UVA son moito máis suaves, preanunciando unha subida sobre a última vogal ausente nas dúas variedades descritas anteriormente.

Tamén neste caso, unha organización temporal considerablemente diferente conduce á redución e á cancelación de aproximadamente o 10% das vogais non acentuadas.

---

12 Moitos, por exemplo, tamén os diferentes perfís asociados ás secuencias CVCVCV con consoantes sonoras e ás secuencias con xordas: observamos, en especial nestas realizacións, como a presenza de nasais en posicións salientes leva frecuentemente á fusión dos núcleos e á compresión dos perfís.

Tèramo (TER)

A locutora deste lugar só produciu a segunda serie de frases, con realizacións espontáneas e coherentes. Como trazo característico presentou ao final de cada realización interrogativa un cambio brusco final de F_0 mal controlado (en termos de rexistros de fonación), quizais típico seu ou quizais imputable á tensión da enquisa.

Aínda que auditivamente se manteña moi afastado dos outros, o PFM interrogativo desta rexión non se afasta moito dos de RAV e de SMF en canto á súa forma.

O que parece evidencialo máis en particular é unha progresión de subidas melódicas que comeza bastante antes en relación á UVA, levando a unha preacentuada case sempre medio-alta, marcada por un perfil ascendente e abrupto (como en Roma, onde ocorre nun nivel medio), seguido na UVA dun descenso máis gradual cara aos mesmos valores medios da remonta da preacentuada e dunha posterior (feble) subida na última vogal da frase.

Aínda que recordemos que para esta variedade tamén se describiron habitualmente fenómenos de elisión e de cancelación vocálica máis xerais, ningunha manifestación deste tipo estaba presente no noso modelo. Pode considerarse un primeiro elemento importante de diferenciación entre os modelos, aínda que sexan diferentes das outras áreas.

CONCLUSIÓNS E PERSPECTIVAS

Sobre a base do modelo observado, tratamos parcialmente algunhas conclusións sobre os elementos de diferenciación das tres áreas xeoprosódicas indagadas. As principais tendencias xurdidas esperan agora a confirmación por parte de investigacións máis profundas no programa ou na fase de execución dalgunhas delas.

Aínda que no estado actual queda efectuar unha confrontación dos resultados que dispoñemos grazas aos estudos da entoación xa publicados (sobre todo para a área 64 [*napoletano*], 63 [*barese*] e 76 [*abruzzese* centro-meridional]), os nosos datos, polo momento parciais e provisionais, ofrecen un punto importante para demostrar tamén en Italia a presenza da variedade cuxa prosodia non pareza estar limitada ao estudo dos perfís melódicos e esquemas rítmicos delimitados por unha organización temporal rixidamente vocálica. Como xa mostraran as análises dos datos portugueses e en menor medida aqueles da área galorrománica, estes materiais confirman a importancia mesmo dos elementos non sonoros na caracterización prosódica dalgunha variedade[13].

Estas consideracións indúcennos a reflexionar con máis profundidade na definición dun método para a valoración e a confrontación de solucións prosódicas características ligadas á presenza deste fenómeno. En particular, unha interesante hipótese de traballo é aquela que nos leva a estudar a modalidade de inclusión nunha concepción xeral de prosodia, das axudas das reducións e das rupturas, así como a consideración das fricciónes e daqueles ruídos con cor vocálica que ás veces compensan á perda de elementos vocálicos non realizados.

---

13 *Cela va sans dire* que as implicacións desta hipótese poden levar a reflexións máis xerais sobre a prosodia das linguas (cfr. DE CASTRO MOUTINHO *et alii*, 2004; ROMANO, 2005).

BIBLIOGRAFÍA

AVOLIO, Francesco: *Bommèspr'. Profilo linguistico dell'Italia centro-meridionale*, San Severo, Gerni, 1995.

AVOLIO, Francesco: "Ma nuje cómme parlamme? Problemi di descrizione e classificazione dello spazio dialettale «campano»", *Romance Philology*, 54, 1-28, 2000.

AVOLIO, Francesco: "L'Abruzzo", In G. Clivio, M. Cortelazzo, N. De Blasi, C. Marcato (a cura di), *Dialetti italiani: storia, struttura, uso*, Torino, UTET, 568-607, 2002.

AVOLIO, Francesco: "Gli «indicatori geografici»" come fonte per gli studi dialettali: esempi lucani", In M.T. Greco (a cura di), *Gli "indicatori geografici" della Basilicata nord-occidentale nei territori delle Comunità montane del Marmo e del Melandro*, Napoli, Università degli Studi "L'Orientale", Quaderni di *AION* (nuova serie - 6), 27-44, 2003.

BERTINETTO, Pier Marco: "Reflections on the dichotomy "stress" vs. "sillable-timing"", *Revue de Phonétique Appliquée*, 91-93, Université de Mons, 99-130, 1989.

CUTUGNO, Francesco: "Il tempo della voce". In R. Delmonte e A. Bristot (a cura di), *Aspetti computazionali in fonetica, linguistica e didattica delle lingue: modelli e algoritmi. Atti delle IX Giornate di Studio del "Gruppo di Fonetica Sperimentale" dell'Associazione Italiana di Acustica* (Venezia, 17-19 Dic. 1998), Roma, Esagrafica, 191-204, 1999.

DE CASTRO MOUTINHO, Lurdes, *et alii*: "Estudo comparativo da variação prosódica em duas línguas românicas: o Português e o Italiano", *Actas do XIX Encontro Nacional da APL - Associação de Linguística Portuguesa* (Lisboa, outubro 2003), Lisboa, APL, 719-723, 2004.

DI FRANCESCO, Alessandro: "Variabilità degli schemi intonativi dialettali: analisi di un corpus scelto di alcune varietà abruzzesi", Tesi di Laurea (rel. Prof.ssa L. Mucciante, Fac. di Lettere - Univ. "G. D'Annunzio", Chieti-Pescara, a.a. 2002-2003.

FIRTH, John R.: "Sounds and Prosodies", *Transactions of the Philological Society* (rep. in W.E. Jones & J. Laver (ed.), *Phonetics in Linguistics. A Book of Readings*, London, Longman, 1973), 127-152, 1948.

GIAMMARCO, E.: *Abruzzo dialettale*, Pescara, Istituto di Studi Abruzzesi, 1973.

HOLTUS, Günter, METZELTIN, Michael, & SCHMITT, Christian (Hrsgg.): *Lexikon der Romanistischen Linguistik (LRL)*, vol. IV (Italiano, Corso, Sardo), Tübingen, Niemeyer, 1988.

LÜDTKE, H.: *Lucania*, Pisa, Pacini, 1979.

MATURI, Pietro: "L'intonazione delle frasi dichiarative ed interrogative nella varietà napoletana dell'italiano", *Rivista Italiana di Acustica*, 12, 13-30, 1988.

RAMUS, Frank, NESPOR, Marina, & MEHLER, Jacques: "Correlates of linguistic rhythm in the speech signal", *Cognition*, 73(3), 265-292, 1999.

ROMANO, Antonio: *Analyse des structures prosodiques des dialectes et de l'italien régional parlés dans le Salento: approche linguistique et instrumentale*, Lille, Presses Univ. du Septentrion, 2001.

ROMANO, Antonio: "Accento e intonazione in un'area di transizione del Salento centro-meridionale", In P. Radici Colace, G. Falcone & A. Zumbo (a cura di), *"Storia politica e storia linguistica dell'Italia meridionale", Atti del convegno internazionale di studi parlangeliani* (Messina, 22-23 Maggio 2000), Messina-Napoli, Ed. Scientifiche Italiane, 169-181, 2003.

ROMANO, Antonio: "Utilisation des données *AMPER* pour une description de la variation linguistique : tests de perception et contrôles statistiques", *Géolinguistique*, no. 3 (hors série: Projet *AMPER - Atlas Multimédia Prosodique de l'Espace Roman*), 39-64, 2005.

ROMANO, Antonio, LAI, Jean Pierre, & ROULLET, Stefania: "La méthodologie *AMPER*", *Géolinguistique*, no. 3 (hors série: Projet *AMPER - Atlas Multimédia Prosodique de l'Espace Roman*), 1-5, 2005.

ROMITO, Luciano, & TRUMPER, John: "Problemi teorici e sperimentali posti dall'isocronia", *Quaderni del Dipartimento di Linguistica dell'Univ. della Calabria*, Serie Ling. 4, 10, 89-118, 1993.

RUSSO, Domenico: "Sulla prosodia dell'italiano regionale abruzzese (IRA)", In F. Albano-Leoni, F. Cutugno, M. Pettorino & R. Savy (a cura di), *Il Parlato Italiano* (Atti del Convegno Naz. di Napoli, 13-15 Febbraio 2003), Napoli, D'Auria (CD-ROM), 2004.

SCHMID, Stephan: "A typological view of syllable structure in some Italian dialects", In P.M. Bertinetto *et alii* (a cura di), *Certamen Phonologicum, Atti del 3° Colloquio di Fonologia* (Cortona, 1996), Torino, Rosenberg & Sellier, 247-265, 1996.

VIGNUZZI Ugo: "Italienisch: Areallinguistik VII. Marche, Umbrien, Lazio", In HOLTUS, METZELTIN & SCHMITT (1988), 606-42.

LA REDUNDANCIA DEL LENGUAJE PREDICE LA DURACION SILABICA Y LAS CARACTERISTICAS ESPECTRALES DEL NUCLEO SILABICO VOCALICO

MATTHEW AYLETT
Departamento de Lingüística Teórica y Aplicada, Universidad de Edimburgo, UK

INTRODUCCION

En el habla espontánea existe una gran variación en la articulación de un mismo fonema por el mismo hablante. La mayoría de la investigación en el campo del habla se ha dedicado a la explicación y descripción de esta variación. Una observación ya mencionada por otros investigadores en los años 60 (Bolinger 1963; Lieberman 1963; Sharp 1960) es el hecho de que cuanto más redundante algo es, más corto y más reducido tiende a ser. Cuando hablamos de algo redundante queremos decir predecible, bien gracias a su contexto o gracias a la frecuencia de la unidad léxica, sílaba o fonema. Estudios más recientes han confirmado estas observaciones (e.g. Aylett 2000, Bybee 2000, Bell et al. 2003, Wright 2003, Munson 2004). Además de la reducción en la duración, también se han encontrado efectos espectrales en los primeros dos formantes medidos en la posición central de la vocal. Hasta que punto la redundancia en el lenguaje se relaciona con la reducción acústica tiene una serie de implicaciones importantes para las teorías de la percepción y producción del lenguaje. En la teoría H&H de Lindblom (1990) este sostiene, que la reducción y prominencia, en la forma de un continuum de articulación hypo-hypa, permite a los hablantes conservar el esfuerzo mientras producen un habla lo suficientemente distintiva. Sin embargo, tanto la redundancia del lenguaje como el cuidado de la articulación afectan lo que a oidos del oyente es suficiente acusticamente hablando. El cuidado de la articulación hace que la acústica sea más distintiva y la redundancia del lenguaje aumenta la probabilidad de reconocer correctamente la palabra gracias al contexto o a su frecuencia. Por esta razón la reducción acústica y la reducción del lenguaje pueden ser integradas dentro de las explicaciones probabilísticas de la producción y la percepción. Algunos ejemplos de estudios recientes en este campo incluyen The Probabilistic Reduction Hypothesis

(Jurafsky, 2001), y The Smooth Signal Redundancy Hypothesis (Aylett, 2000; Aylett and Turk, 2004).

En el estudio que presentamos aquí analizamos un extenso corpora de voz (aproximadamente 500,000 sílabas). Demostramos que 1) el efecto de la redundancia/duración está definitivamente presente en estos datos, 2) el efecto esprectal es también significativo, 3) que gran parte del poder de predicción de los factores prosódicos y de redundancia es compartido, y 4) la dirección del efecto es de redundancia que conyeba una reducción en forma de centralización vocal. Estos resultados respaldan la Probabilistic Reduction Hypothesis y hasta cierto punto la Smooth Signal Redundancy Hypothesis en el ámbito espectral.

DISEÑO EXPERIMENTAL

Con el fin de examinar la redundancia de la señal suave de los factores espectrales, utlizamos un extenso corpus de habla leída en inglés americano (el corpus de Rhetorical Sys.). Escogimos este corpus debido a la calidad de las grabaciones y el cuidado en la articulación que nos permitió hacer un seguimiento automático de los formantes y una segmentación fiables.

Cada sílaba del corpus fue puntuada siguiendo un modelo de redundancia acústica de F1/F2, un modelo sencillo de redundancia del lenguaje basado en la probabilidad n-gram de la sílaba, y en un modelo prosódico conservador. También realizamos un análisis de ANOVA con posterior análisis de post hoc y análisis de regresión para investigar la relación entre los valores de los formantes F1/F2, la redundancia del lenguaje y la prosodia.

Los resultados que presentamos aquí se limitan a un grupo de vocales de nuestro corpus. Nuestro objetivo era concentrarnos en un grupo de vocales que minimizarían el ruido causado por los errores en el análisis de los formantes. Las vocales con duraciones cortas como /ɹ̩, ɪ, ʊ, ʌ/ son frecuentemente demasiado cortas para poder realizar un análisis espectral automático con total seguridad. También evitamos los diptongos debido a sus objetivos espectrales cambiantes que hacen una sola medición poco fiable. Finalmente, la vocal /ɔ/ (como en 'lawn') fue descartada debido a la gran variación existente en la pronunciación de esta vocal en el acento general Americano (Wells 1882). El grupo definitivo de vocales fue el siguiente: /ɑ/ (pass), /æ/ (tap), /ɛ/ (less), /i/ (lease), /u/ (goose).

En el análisis aquí presentado ultilizaremos los valores de los formantes F1/F2 del núcleo de la vocal como medida de la calidad espectral de la sílaba. Utilizaremos la centralización como modelo para relacionar estos valores F1/F2 con la redundancia acústica. Los resultados seran presentados individualmente para cada vocal (mostrando diferencias F1/F2 reales) para evitar el efecto de una posibible existencia de una conducta de centralización que sería marcadamente diferente entre las vocales.

El modelo de la redundancia del lenguaje

El modelo del lenguaje se basó en la probabilidad silábica construida a partir de un contexto de unigrama, bigrama, trigrama, i.e. que probable es una sílaba sin contexto, cual es la probabilidad si se tiene en cuenta la sílaba previa, y cual es la probabilidad si se tienen en cuenta las dos sílabas anteriores. Una transformación log fue aplicada a estas probabilidades ya que asocia los valores al contenido de la información (Pierce 1961). Utilizamos el 'language toolkit' de CMU (Clarkson y Rosenfield 1997) para poder calcular estas probabilidades con estadística n-gram basada en la transcripción silábica de 187 millones de palabras provinientes de recursos de noticias en internet. Las transcripciones se generaron automáticamente utilizando el sintetizador de voz de Rhetorical Systems. Realizamos un análisis de factores para producir dos factores que llamamos medidas de redundancia de *contexto abierto* y *contexto reducido* donde:

- contexto abierto era aproximadamente una media de las tres probabilidades (ver Tabla I).
- contexto reducido era la probabilidad unigrama con la eliminación de probabilidades bigrama y trigrama (ver Tabla I).

	Factor Analysis Component		Mean	Standard Deviation
	Wide	Narrow		
UNIGRAM	0.315	1.090	-6.924	1.956
BIGRAM	0.421	-0.314	-5.480	2.931
TRIGRAM	0.407	-0.520	-5.063	3.411

Tabla I: Contribución de las probabilidades log del n-gram original a los factores abierto y reducido. Normalizamos los valores del unigrama, bigrama y trigrama utilizando los valores de la media y desviación estándar y multiplicándolos por los componentes del análisis factorial. Esta transformación produjo los valores para los factores abierto y reducido.

Los factores resultantes tenían las siguientes ditribuciones: media del factor abierto 0.0, mediana 0.132, desviación estándar (sd) 1.00, media del factor reducido 0.0, mediana 0.1130, sd 1.00. Estas distribuciones explicaron el 96.8% de la variación de las probabilidades del unigrama/bigrama/trigrama. La variable de redundancia del lenguaje utilizada en el análisis ANOVA fue construída de la forma siguiente:

Agrupamos los datos dentro de tres grupos distintos de redundancia:

1. Redundancia del lenguaje alta: tanto el factor de redundancia abierto como el reducido estan por encima del valor de la mediana (+,+).
2. Redundancia del lenguaje mediana: redundancia reducida más alta que la mediana y redundancia abierta más baja que la mediana o redundancia reducida más baja que la mediana y redundancia abierta más alta que la mediana (+,-) o (-,+).

3. Redundancia del lenguaje baja: tanto el factor de redundancia reducido como el abierto estan más bajos que el valor de la mediana (-,-).

El modelo prosódico

Los efectos derivados de las fronteras prosódicas fueron controlados eliminando todas las sílabas seguidas de una frontera de frase o no seguidas de una frontera de palabra (i.e. consideramos sílabas seguidas de un index de interrupción de 1). Cada sílaba fue marcada con tres niveles de prominencia:

- Ninguno.
- Acento léxico primario.
- Alta probabilidad de acento de frase (acento léxico primario + clase abierta).

RESULTADOS

Duración silábica

Efectuamos un análisis de varianza ANOVA del grupo de redundancia del lenguaje, y del grupo de prominencia que mostraron un efecto significativo marcado, (Grupo de redundancia del lenguaje – alta, media, baja, $F(2, 313941) = 67726$, $p < 0.001$, Grupo de prominencia – ausencia de acento, acento léxico, acento de frase, $F(2, 313941) = 72118$, $p < 0.001$). La Figura 1 muestra que la media de los valores de cada grupo fueron todas significativamente diferentes entre ellas (post hoc t-test con corrección Bonferroni, $p < 0.001$). A medida que la redundancia del lenguaje aumenta las sílabas se acortan y viceversa, a medida que el nivel de prominencia prosódica aumenta las sílabas se alargan.

Un análisis de regresión lineal, con fronteras controladas, mostró que los modelos de regresión que hacen uso de estos factores predicen una proporción significativa de los datos del Corpus Rhetorical ($r= 0.7219$, $r^2= 0.5211$ – 52% de la variación). La contribución independiente del modelo prosódico y de redundancia y el poder predictivo compartido de los dos modelos se presenta en la Figura 2. Cerca de un 20% del poder predictivo fue compartido. La relación entre la redundancia de la duración y el lenguaje en todas las vocales fuertes que analizamos mostraron el mismo comportamiento, i.e. más larga = menos redundancia del lenguaje. Sin embargo, la relación entre el grupo de redundancia del lenguaje medio y la duración si mostró una gran variación dependiendo de la vocal, ver Figura 3. (e.g. en /i/ vemos un ligero descenso pero en otras vocales como /u/ el resultado no es tan fácil de interpretar). Todas las diferencias fueron significativas $p < 0.001$ (post hoc t-test con corrección Bonferroni).

La Figura 4 muestra la cantidad de la variación en la duración silábica (log ms) que fue predecida en su totalidad por los modelos de regresión que incluyen tanto el factor de prominencia prosódica como el factor de redundancia del lenguaje. El modelo combinado regresión/prosodia fue capaz de predecir significativamente ($p < 0.001$) entre un 18-50% de la variación en la duración silábica de las vocales analizadas. La

mayoría del poder predictivo fue compartido entre la redundacia del lenguaje y la prominencia prosódica, aunque la cantidad de esta contribución compartida fue muy diferente dependiendo del tipo de vocal.

RESULTADOS ESPECTRALES

Resultados espectrales: t-tests

Siguiendo un modelo sencillo de centralización de la redundancia acústica fue posible predecir la variación que se muestra en la Tabla II. Un aumento en la redundancia del lenguaje conlleva una centralización más marcada en las vocales.

F1/F2 Prediction with centralization model of acoustic redundancy		
Vowel	*F1 change*	*F2 change*
ɑ	minus	plus
æ	minus	same
ɛ	minus	same
i	plus	minus
u	plus	plus

Tabla II: Predicciones de F1/F2 para cada una de las vocales al disminuir la redundancia acústica debido a la centralización.

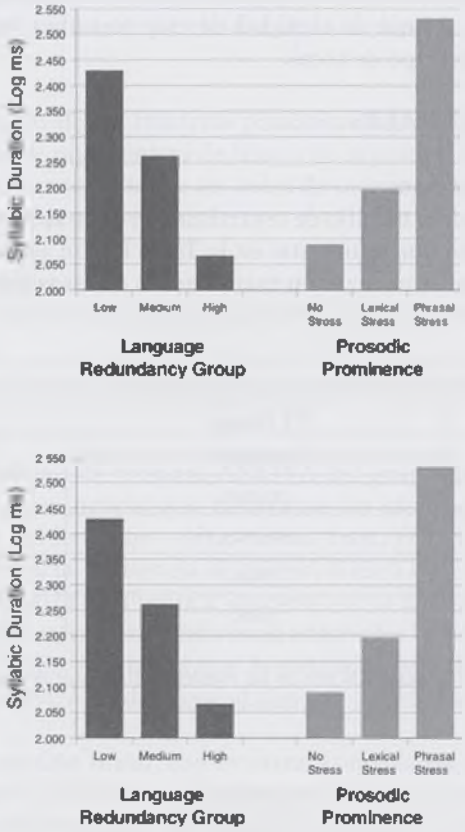
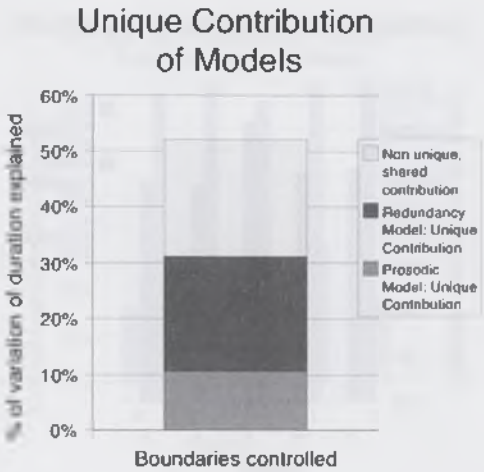


Figura 1. Medias de la duración silábica (log ms), según el grupo de redundancia y prominencia prosódica. (Las fronteras prosódicas estan controladas para todas las medias). Todas las medias son significativas. Como referencia: 2.00 log ms = 100 ms, 2.20 log ms = 158.5 ms, 2.40 log ms = 251.2 ms, 2.60 log ms = 398.1 ms.



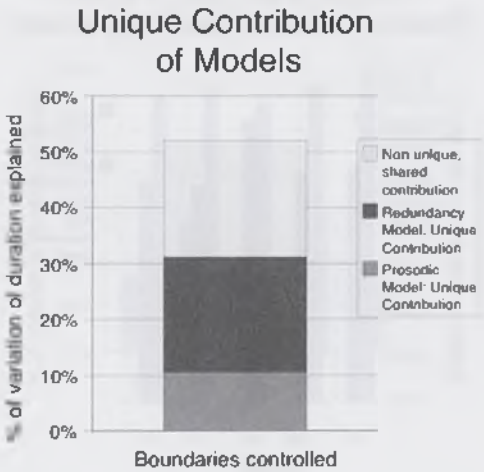
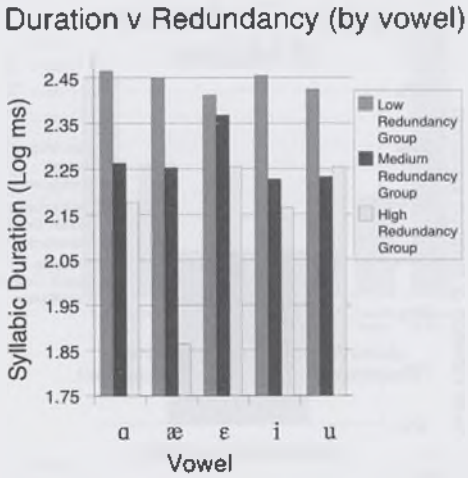


Figura 2. La contribución única del modelo de prominencia prosódica y del modelo de redundancia a la correlación linear.



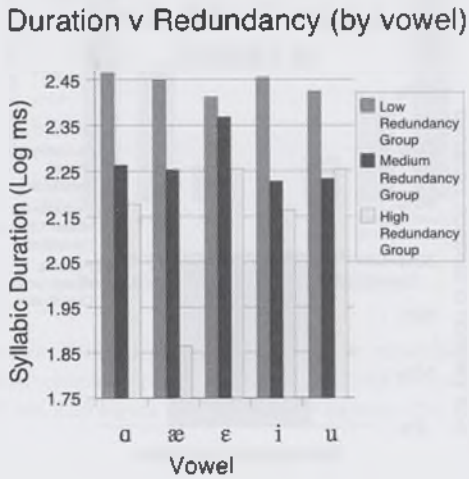
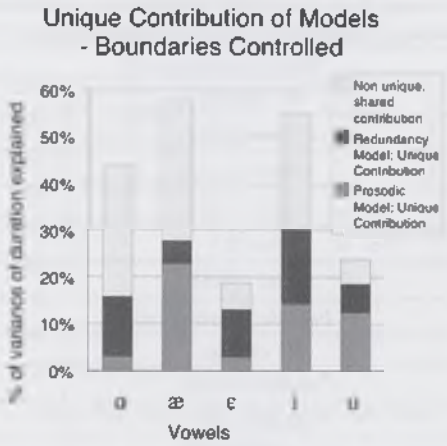


Figura 3. Medias de duración silábica (log ms), por grupo de redundancia y prominencia prosódica, por cada vocal. Todas las medias son significativas. Como referencia: 2.00 log ms = 100 ms, 2.20 log ms = 158.5 ms, 2.40 log ms = 251.2 ms, 2.60 log ms = 398.1 ms.



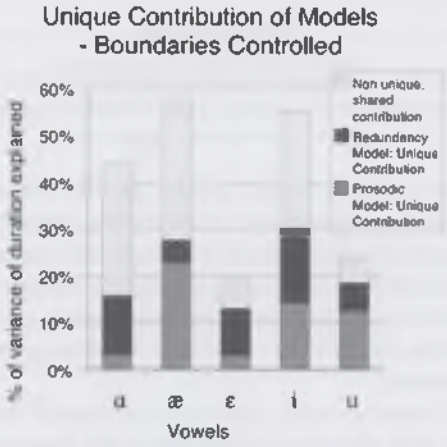


Figura 4. La contribución única del modelo de prominencia prosódica y del modelo de redundancia a la correlación linear por cada una de las vocales.

Los resultados de F1/F2 de las cinco vocales fuertes seleccionadas fueron analizados controlando las fronteras prosódicas. La redundancia del lenguaje resultó estar significativamente relacionada con los valores F1/F2 de las vocales pero el grado de significación y la dirección de este efecto fue muy variable dependiendo del tipo de vocal, ver Tabla III. Los resultados se presentan por separado para los hablantes masculinos y los femeninos.

F1/F2 change from low to high redundancy prediction/result low redundancy group F1/F2 - high redundancy group F1/F2 (Hz)		
Male Speakers (n=3)		
Vowel	F1 change	F2 change
ɑ	minus/ -42.0	plus/ +22.2
æ	minus/ -21.0	same/ +15.1
ɛ	minus/ -72.2	same/ -36.3
i	plus/ 8.4	minus/ -48.8
u	**plus/ -18.43**	plus/ 31.72
Female Speakers (n=5)		
Vowel	F1 change	F2 change
ɑ	minus/ -188.6	**plus/ -13.2**
æ	minus/ -44.1	same/ 37.65
ɛ	minus/ -119.06	same/ 12.93
i	plus/ 8.76	minus/ -37.0
u	**plus/ -5.0**	plus/ 82.2

Tabla III. Cambios en las medias de F1/F2 con una escala de sílabas de redundancia baja a redundancia alta.

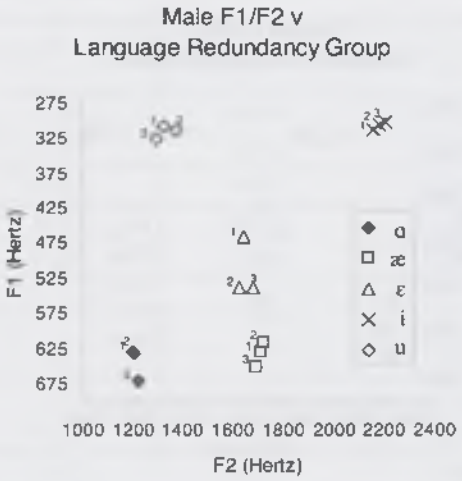
Los resultados que contradicen una predicción de centralización se presentan en negrita. Los resultados que esperabamos se mantuviesen constantes necesitaron un anális por separado. Podríamos decir que, siempre y cuando los resultados no cambien dramáticamente, un pequeño cambio en los resultados podría explicarse por el modelo de centralización. En general, los resultados fueron razonablemente consistentes con una simple hipótesis de centralización excepto por /u/ donde F1 disminuye y /ɑ/ en los hablantes femeninos donde F2 aumenta.

Las Figuras 5 y 6 muestran estos cambios en el espacio vocálico para los hablantes femeninos y masculinos. Los grupos 1 (redundancia del lenguaje alta) y 3 (redundancia del lenguaje baja) muestran un comportamiento de centralización. Sin embargo, para la mayoría de las vocales esto ocurre debido a una disminución general de F1. El grupo 2, formado por una mezcla de redundancia lingüística alta/baja y baja/alta es mas volátil, lo que sugiere que sus miembros son menos homogéneos.

Resultados espectrales: regresión lineal

Mediante la aplicación de una regresión lineal a los resultados espectrales fuimos capaces de mostrar:

1. La cantidad de variación predecida por el modelo prosódico y de redundancia.
2. Cuanto poder de predicción es compartido entre el modelo prosódico y el modelo de redundancia del lenguaje en cada una de las vocales.



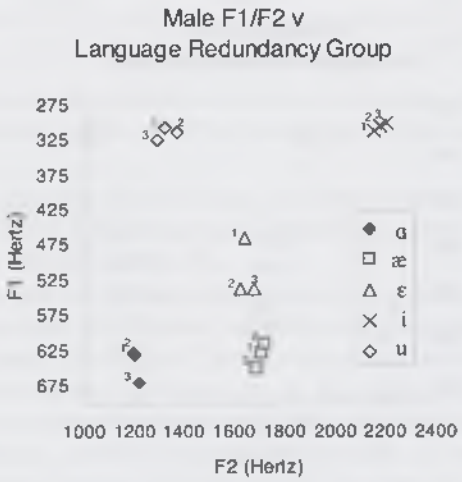
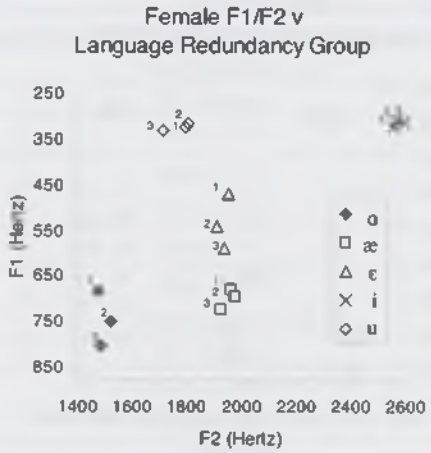


Figura 5. Resultados espectrales de los hablantes masculinos por vocal (n=3). 1 = Redundancia del lenguaje alta, 2 = Redundancia del lenguaje mediana, 3 = Redundancia del lenguaje baja. Todas las diferencias entre 1&2, 2&3, 1&3 son significativas ($p<0.005$) para los formantes F1 y F2 (a menos que se indique lo contrario) basándonos en los resultados de un test post hoc con corrección Bonferroni. No se encontraron diferencias significativas entre los grupos: /ɑ/ F1 1&2, /ɑ/ F2 1&2&3, /ɛ/ F1 2&3, /ɛ/ F2 1&2.



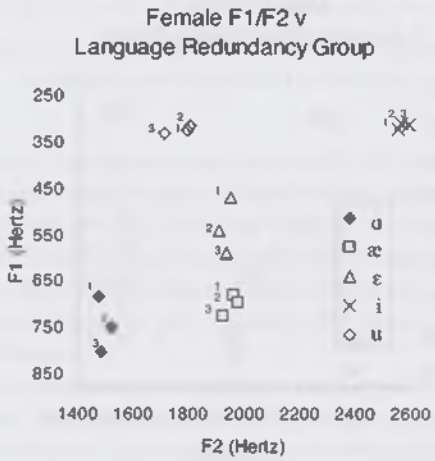


Figura 6. Resultados espectrales de los hablantes femeninos por vocal (n=5). 1 = Redundancia del lenguaje alta, 2 = Redundancia del lenguaje mediana, 3 = Redundancia del lenguaje baja. Todas las diferencias entre 1&2, 2&3, 1&3 son significativas ($p<0.005$) para los formantes F1 y F2 (a menos que se indique lo contrario) basándonos en los resultados de un test post hoc con corrección Bonferroni. No se encontraron diferencias significativas entre los grupos: /æ/ F1 1&2, /æ/ F2 1&2, /i/ F1 2&3, /u/ F2 1&2.

Male F1/F2				
	F1		F2	
	r	r^2	r	r^2
ɑ	0.261	0.068	0.113	0.013
æ	0.215	0.046	0.190	0.036
ɛ	0.281	0.079	0.081	0.007
i	0.214	0.046	0.243	0.059
u	0.228	0.052	0.205	0.042
Female F1/F2				
	F1		F2	
	r	r^2	r	r^2
ɑ	0.228	0.052	0.071	0.005
æ	0.071	0.005	0.189	0.036
ɛ	0.266	0.071	0.040	0.002
i	0.115	0.013	0.223	0.050
u	0.075	0.006	0.096	0.009

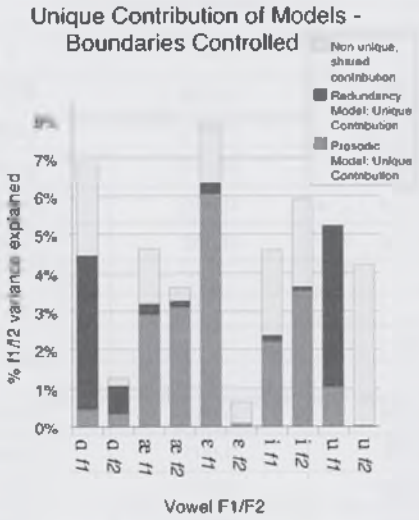
Tabla IV. Resultados de la regresión lineal de la combinación del modelo de redundancia con el modelo de prominencia prosódica para F1 y F2, por cada vocal y sexo del hablante.

El análisis de regresión demostró que el modelo prosódico y de redundancia pueden predecir entre el 0.5%-7.9% de la variación de F1 ($p<0.001$), y un 0.2%-5.9% de la variación de F2 ($p<0.001$), ver Tabla IV.

Las diferencias entre los hablantes masculinos y femeninos podría haber sido causada por el pobre funcionamiento del extractor automático de formantes en el análisis de los hablantes femeninos (Figura 7). La Figura 7 muestra hasta que punto la relación entre el modelo prosódico y de redundancia varió por cada una de las vocales. Algunas vocales (e.g. /u/-f1) contienen un componente de redundancia muy fuerte que es ortogonal al modelo prosódico, y otras (e.g. /i/) tienen una contribución única más pequeña del modelo de redundancia. En todos los casos (excepto en la condición /ɛ/-f2 y /u/-f2) la regresión del modelo de redundancia fue significativa ($p<0.001$). Al contrario que en los resultados de la duración de la sílaba, en algunas vocales la contribución compartida fue bastante pequeña o inexistente.

CONCLUSIONES

El objetivo de este estudio fue examinar si la hipótesis de la redundancia de la señal suave podría ser generalizada para explicar los resultados espectrales. En este estudio hemos encontrado efectos significativos para los valores de los formantes F1/F2. Cuanto más redundante y menos prominente es la sílaba, más centralizado y acústicamente re-



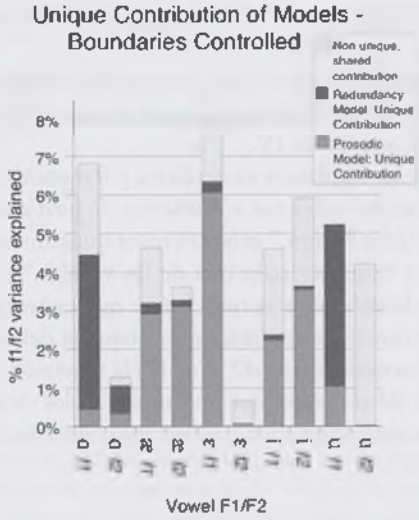


Figura 7. Relación entre el poder de predicción de los modelos de regresión lineal de la prominencia prosódica y el de redundancia en los hablantes masculinos.

dundante es el núcleo de la vocal. También hemos demostrado que el efecto de redundacia y prominencia prosódica son similares. Este hallazgo respalda el concepto de que la estructura prosódica podría servir como método para alcanzar la redundancia de la señal suave y así fortalecer la señal de voz en un entorno ruidoso. En general, creemos que este hallazgo respalda en gran medida nuestra hipótesis.

Sin embargo, es importante destacar que nuestros resultados varían considerablemente dependiendo del tipo de vocal. Algunas vocales, como por ejemplo /i/, se comporta tal y como se esperaba, mientras otras vocales, como por ejemplo /ɛ/, parecen mostrar una relación diferente entre el modelo prosódico y los factores de redundancia. Una complicación que debemos de tener en cuenta es que ni los factores de redundancia ni los factores prosódicos tienen variaciones similares por cada una de las diferentes vocales (ver Tabla III). Por ejemplo, /i/ es una vocal muy común que se utiliza muy a menudo en contextos inacentuados (como en la última sílaba de la palabra 'spongy'). Así pues, el modelo de redundancia contiene más variación capaz de predecir un cambio en la duración de /i/ que el modelo prosódico. Esto lo podemos comparar con /ɛ/ donde menos variación en la redundancia abierta y una media baja sugieren que esta vocal nunca se encuentra en una sílaba muy predecible en ninguno de los contextos posibles. Además, con las fronteras controladas, casi la totalidad de los elementos /ɛ/ estaban acentuados y en palabras de clase abierta. Esta ausencia de variación disponible para el modelo de prominencia y el modelo de redundancia explica el bajo rendimiento de estos modelos en la vocal /ɛ/.

Esta cuestión pone de relieve la pregunta de qué modelo perceptual es más apropiado para construir un modelo de redundancia acústica. Al contrario de la duración por sí misma, donde segmentos más largos han demostrado ser más inteligibles, todavía no está claro como factores del tipo F1/F2, amplitud, alteraciones de f0 e inclinación espectral se podrían combinar con indices acústicos y dinámicos para aumentar o reducir la inteligibilidad. La investigación de estos factores y su relación con la percepción es una labor que continúa y creemos que un marco de redundancia de la señal puede ofrecer un enfoque cuantitativo formal para su estudio.

REFERENCIAS

AYLETT, M. and A. TURK (2004) "The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech.," *Language and Speech*, 47, pp. 31-56.

AYLETT, M.P. (2000) *Stochastic Suprasegmentals: Relationships between Redundancy, Prosodic Structure and Care of Articulation in Spontaneous Speech (http://www.cogsci.ed.ac.uk/matthewa/thesis_sum.html)*. Ph.D. thesis, University of Edinburgh.

BELL, A., D. JURAFSKY, E. FOSLER-LUSSIER, C. GIRAND, M. GREGORY, and D. GILDEA (2003) "Effects of disfluencies, predictability, and utterance position on word form variation in english conversation.," *The Journal of the Acoustical Society of America*, 113, pp. 1001-1024.

BOLINGER, D. (1963) "Length, vowel, juncture," *Linguistics*, 1, pp. 5-29.

BYBEE, J. (2000) "Lexicalization of sound change and alternating environments," in *Papers in Laboratory Phonology V: Acquisition and the Lexicon*, edited by M. Broe and J. Pierrehumbert. (Cambridge University Press, Cambridge)

CLARKSON, P. and R. ROSENFELD (1997) "Statistical language modeling using the CMU-Cambridge toolkit," in *Proceedings of Eurospeech* 97, pp. 2707-10.

JURAFSKY, D., A. BELL, M. GREGORY, and W. RAYMOND (2001) "Probabilistic relations between words: Evidence from reduction in lexical production," in *Frequency and the Emergence of Linguistic Structure*, edited by J. Bybee and P. Hopper. (John Benjamins, Amsterdam)

LIEBERMAN, P. (1963) "Some effects of semantic and grammatical context on the production and perception of speech," *Language and Speech*, 6, pp. 172-187.

LINDBLOM, B. (1990) "Explaining phonetic variation: a sketch of the H & H theory," in *Speech Production and Speech Modeling*, edited by W.J. Hardcastle and A. Marchal, pp. 403-439. (Kluwer Academic Publishers, Dordrecht, The Netherlands)

MUNSON, B. and N. SOLOMAN (2004) "The influence of phonological neighborhood density on vowel articulation," *Journal of Speech, Language and Hearing Research*, 47, pp. 1048-58.

PIERCE, J.R. (1961) *Symbols. Signals and Noise: The Nature and Process of Communication.* (Harper, New York)

SHARP, A. E. (1960) "The analysis of stress and juncture in English," *Transactions of the Philological Society*, pp. 104--135.

WELLS, J.C. (1982) *Accents of English*, vol.3. (Cambridge University Press, Cambridge)

WRIGHT, R. (2003) "Factors of lexical competition in vowel articulation," in *Papers in Laboratory Phonology VI: Phonetic Interpretation*, edited by J. Local, R. Ogden, and R. Temple,. (Cambridge University Press, Cambridge)

INTONATIONAL PATTERNING IN THE PAULISTA BRAZILIAN PORTUGUESE VARIETY: A TWO CASE STUDY IN THE FRAMEWORK OF AMPER

PLÍNIO A. BARBOSA
University of Campinas plinio@iel.unicamp.br
LUCIANA LUCENTE, LEANDRO S. SILVEIRA
University of Campinas
SANDRA MADUREIRA
Catholic University of São Paulo, Brazil

INTRODUCTION

1.1.The AMPER project

AMPER (*Atlas Multimédia Prosodique de l'Espace Roman*) is a project aiming at mapping the prosody of Romance by means of paired comparisons of melodic, temporal, and energy patterns for vowels in sentences with similar structure. The general co-ordination of AMPER is ensured by Michel Contini and Antonio Romano from the Centre de Dialectologie de l'Université de Grenoble 3, France. As regards Portuguese, Lurdes de Castro Moutinho, from the Centro de Investigação de Línguas e Culturas da Universidade de Aveiro, is in charge of the description of the European varieties, whereas João Antônio de Moraes (UFRJ, Rio de Janeiro), Jussara Abraçado (UFF, Rio de Janeiro) and Plínio A. Barbosa (Unicamp, São Paulo) are in charge of the description of the Brazilian varieties.

In order to make the comparison with other Brazilian Portuguese (henceforth BP) varieties easier, the two teams of researchers from the States of São Paulo and Rio de Janeiro created a common corpus of sentences whose lexical items differ from those used in European Portuguese (henceforth EP). Since naturalness of utterances was aimed at, frequency of lexical usage was taken into account in building up both EP and BP corpora and that determined the aforementioned differences concerning lexical items. The SVO type is common for all sentences in order to follow the AMPER guidelines enabling cross-linguistic comparisons in Romance. Another common feature of the Bra-

zilian corpus is the consonant/vowel alternation along the sentences, making the segmentation of the corresponding utterances easier. The BP corpus is listed in section 1.2. The guidelines for building comparable corpora can be read at the URL: http://www.ii.ua.pt/cidlc/gcl/atlas/.

1.2. The AMPER BP corpus

The following table lists the 24 sentences of the BP corpus, recorded both as a statement and as a yes-no question (henceforth YNQ). Note that both the NP-subject and the NP-object are expanded by including modifiers to the right of the noun. The verb "gostar" (to like) is always at the present tense, 3rd person, singular.

O Renato gosta do pássaro	O bisavô gosta do pássaro
O Renato nadador gosta do pássaro	O bisavô nadador gosta do pássaro
O Renato pateta gosta do pássaro	O bisavô pateta gosta do pássaro
O Renato bêbado gosta do pássaro	O bisavô bêbado gosta do pássaro
O Renato de Salvador gosta do pássaro	O pássaro nadador gosta do Renato
O Renato de Veneza gosta do pássaro	O pássaro pateta gosta do Renato
O Renato de Monaco gosta do pássaro	O pássaro bêbado gosta do Renato
O pássaro gosta de Renato	**O pássaro gosta do bisavô**
O pássaro gosta do Renato nadador	O pássaro gosta do bisavô nadador
O pássaro gosta do Renato pateta	O pássaro gosta do bisavô pateta
O pássaro gosta do Renato bêbado	O pássaro gosta do bisavô bêbado
O pássaro gosta do Renato de Salvador	O Renato gosta do pássaro nadador
O pássaro gosta do Renato de Veneza	O Renato gosta do pássaro pateta
O pássaro gosta do Renato de Monaco	O Renato gosta do pássaro bêbado

2. OBJECTIVE

This paper presents an analysis of the Paulista (São Paulo State) variety of BP by illustrating it with the data of two speakers (one male and one female) from the city of Campinas (Northeastern São Paulo). A 20-year male student from the University of Campinas and a 25-year female, high school graduate, recorded ten repetitions of the 48 sentences of the corpus, which contrasts statement and YNQ in the same word order. The patterns of duration and fundamental frequency (f_0) for the vowels of two statements and the corresponding YNQs are presented here for both speakers. Beginning with a concise presentation of previous findings of BP intonation and rhythm, the sections that follow compare intermodality and intersubject differences, and point to differences in tonal alignment between BP and EP varieties.

3. INTONATION AND RHYTHM IN BRAZILIAN PORTUGUESE

Both Madureira (1994) and Moraes (1998) presented descriptions of BP intonation. As regards the description of default statements and YNQs, both authors summarise the findings as follows. Default statements start with a rising f_0 pattern on the first phonological word and end with a downstepped high tone on the last word, followed by a low boundary tone. On the other hand, default YNQs start with a rising of f_0 that sometimes reach a level higher than that for the initial high tone of a statement, according to both authors. The YNQs end with a rising-falling tune during the last word of the utterance, where the high tone has the highest value of f_0 for the whole utterance.

As regards rhythm, Moraes (1998) point to three main features: (1) the distinct durational patterns of words contrasting primary stress position (antepenultimate, penultimate and final stress), stressed syllables showing duration maximum; (2) the tendency to alternate stressed and unstressed syllables, and (3) the presence of aspects of syllable- and stress-timing depending on speaking style and dialect. Based entirely on a dynamical systems perspective, Barbosa (1999, 2003) e Barbosa and colleagues (2004) favour an analysis within stress group domains, pointing to a general rising of V-to-V duration toward the phrase stresses of each utterance. For both Moraes and Barbosa, syllable-size units are the building blocks of the rhythmic patterns, even though the vowel has a crucial role in this respect.

In order to illustrate and revisit the main findings on BP intonation and rhythm, the duration and f_0 patterns for the vowels of two statements and two YNQs from the BP AMPER corpus are presented in the following section.

3.1. Temporal and melodic patterns of the BP AMPER utterances

In the sentences of the corpus the word "pássaro" (bird) is used as the head of the noun phrase in subject and object positions. The sentences, pronounced both as a statement and a YNQ utterance, are: "O pássaro gosta do bisavô nadador" (The bird likes the swimmer great-grand-father) and "O Renato pateta gosta do pássaro" (Goofy Renato likes the bird). Mean durations were calculated for every vowel in the ten repetitions of the sentences. This was done for the sake of consistency in spite of the fact that productions were found to be similar. As regards f_0, the ten repetitions of each sentence were analysed and produced the same phonological pattern for tones and tunes. In order to illustrate them and to avoid averaging patterns that would require a rigid control of tonal alignment, the f_0 pattern of only one repetition is presented here for both speakers and both modalities.

Figure 1 contrasts vowel duration means along the utterance "O pássaro gosta do bisavô nadador" for the male speaker, while Figure 2 does the same for the female speaker. The only statistically distinct vowel duration when comparing statement and YNQ (t-test: $p < 0.006$) is the penultimate vowel of the male speaker, corresponding to the second [a] of "nadador". This vowel is also longer

than stressed vowel [o]. Such an occurrence is related to the realisation of a early pitch accent on that word (with the high tone aligned with the second pre-stressed vowel) as seen in Figure 5. No other duration differences between statement and YNQ are found for this speaker. A similar result is found for the female speaker but, in that case, the narrow focus on "nadador" does not modify the word duration pattern. Also for this speaker, there is a statistically significant greater duration for the second pre-stressed vowel of "nadador" in the statement (t-test: $p < 0.008$). Observe that the vowel durations within the words "pássaro", "gosta" and "nadador" (putting aside the difference pointed out for the statement of the male speaker), follow the expected pattern: stressed vowels are longer than unstressed ones. On the other hand, the second pre-stressed vowel is the longer one in the final stress word "bisavô", an anticipation that is possible in final stress words and could indicate that the phonological word is not the relevant domain for rhythmic stress placement, but the stress group (Barbosa 1999).

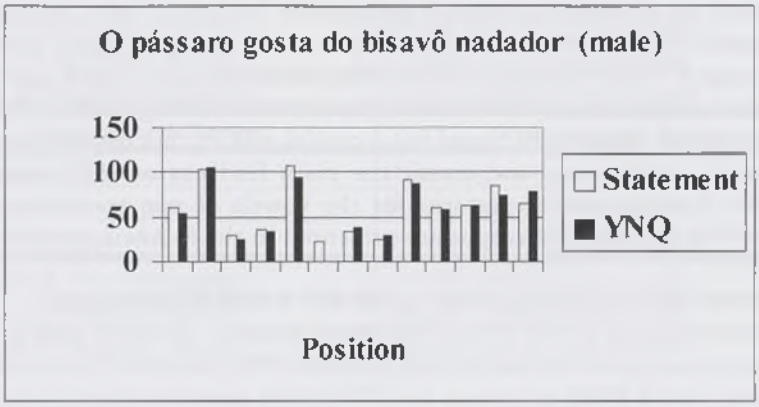


Figure 1: Duration patterns for statement (white) and YNQ (black) for the male speaker. Sentence: "O pássaro gosta do bisavô nadador"

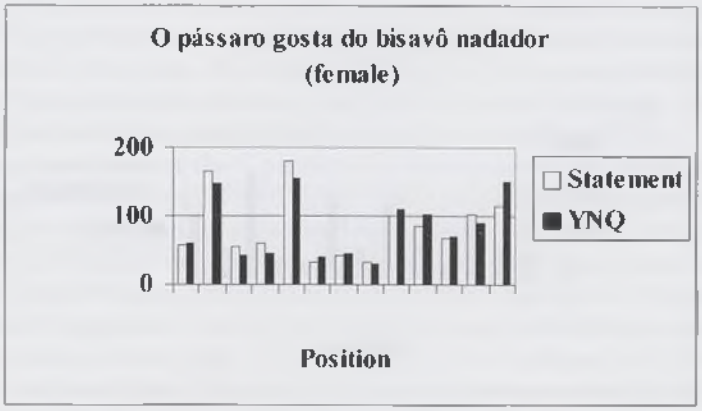


Figure 2: Duration patterns for statement (white) and YNQ (black) for the female speaker. Sentence: "O pássaro gosta do bisavô nadador"

As regards the second sentence, Figures 3 and 4 show the duration patterns for the male and female speakers, respectively. These patterns confirm the longer duration of stressed vowels within the four lexical words. As regards the male speaker, the post-stressed vowel [ʊ] of "Renato" is statistically longer (t-test, $p < 0.004$) in the YNQ than in the statement, whereas the stressed vowel [ɛ] of "pateta" is statistically longer in the statement ($p < 0.009$). This distinction is related to differences of degree of emphasis in the implementation of narrow focus in the two modalities.

As regards the female speaker, the stressed vowel [ɛ] of "pateta" is also statistically longer in the statement ($p < 0.002$), whereas the stressed vowel [a] in "pássaro" is longer

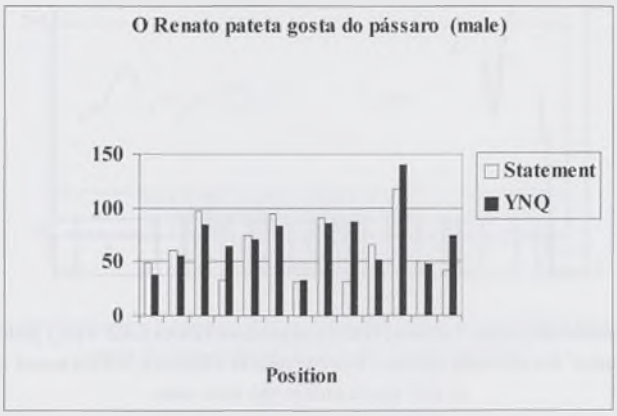


Figure 3: Duration patterns for statement (white) and YNQ (black) for the male speaker. Sentence: "O Renato pateta gosta do pássaro"

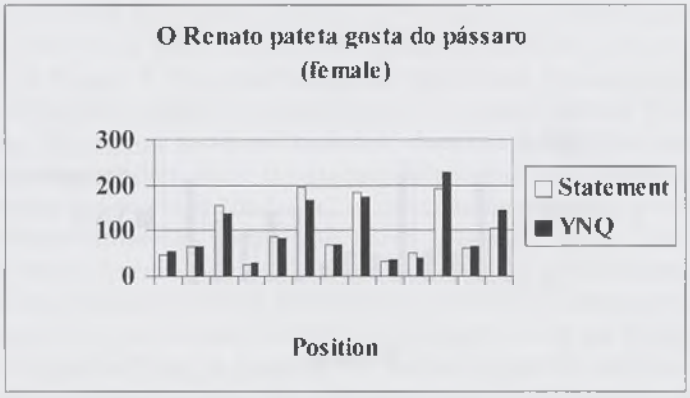


Figure 4: Duration patterns for statement (white) and YNQ (black) for the female speaker. Sentence: "O Renato pateta gosta do pássaro"

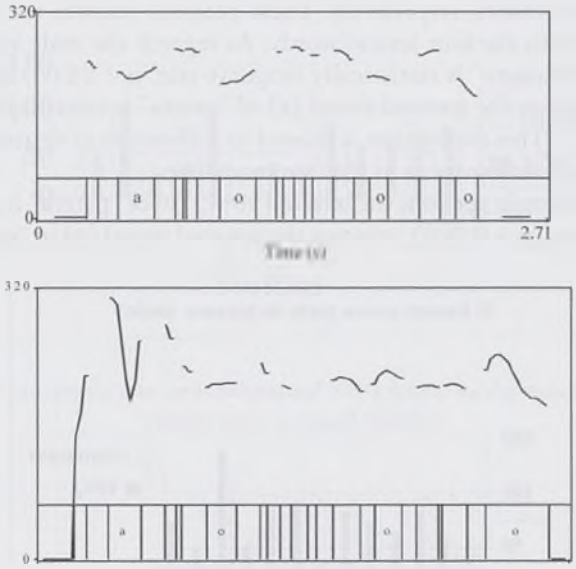


Figure 5: Fundamental frequency patterns (Hz) in statement (above) and YNQ (below) for "O pássaro gosta do bisavô nadador" for the male speaker. For the sake of visibility, only stressed vowels are indicated in this figure and in the next ones.

in the YNQ. The fact that the word "pateta" has narrow focus in the statement and not in the YNQ (Figure 8 and corresponding text) explains the longer duration of [ɛ] in the statement. On the other hand, the duration differences in the stressed vowel of "pássaro" across the two modalities seem to be a consequence of pitch accent range, since there is a high tone aligned with that vowel for both the statement and the YNQ.

Figures 5 to 8 present the f_0 patterns for both speakers and modalities, with the statement presented in the top of each figure. These patterns were obtained with a script implemented by the first author by using Praat version 4.3.24. The script was written in such a way as to respect the AMPER guidelines. Figures 5 and 6 illustrate one of the repetitions for each modality of the first sentence uttered by the male (Figure 5) and the female (Figure 6) speakers. It can be seen that the main tonal events occur during the first and last words, across genre. A rising pitch accent is aligned with the first word "pássaro" in both modalities. The crucial difference between statement and YNQ is related to the opposition between a falling tone starting at the second pre-stressed vowel of "nadador" in the statement vs a rising-falling tone during the same word but rather in the last vowel of the YNQ.

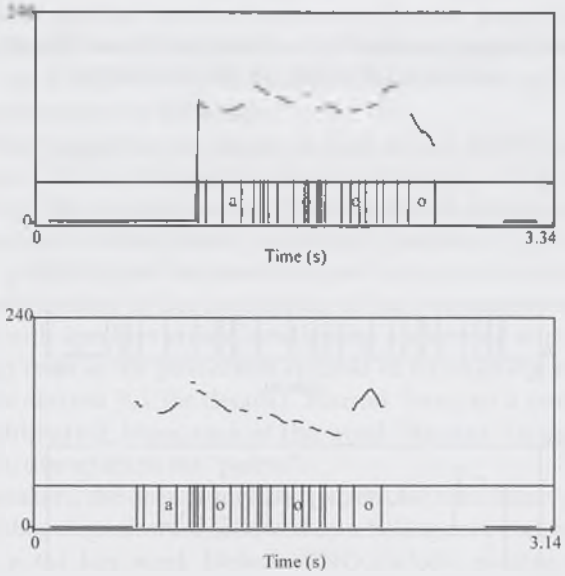


Figure 6: Fundamental frequency patterns (Hz) in statement (above) and YNQ (below) for "O pássaro gosta do bisavô nadador" for the female speaker.

Both speakers put a narrow focus on the verb "gosta" in the statement condition, with the female speaker realising it with a greater range.

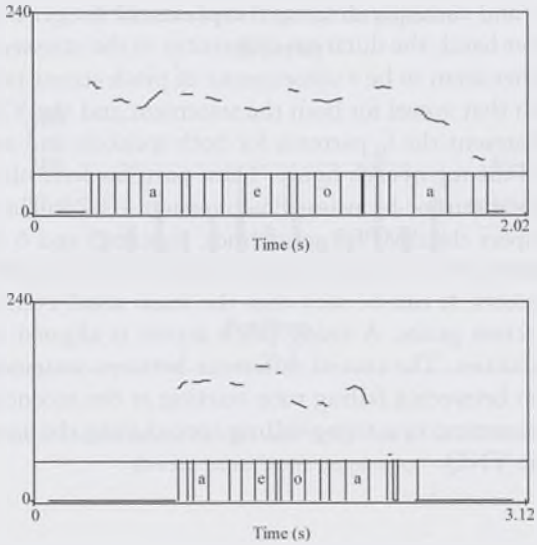


Figure 7: Fundamental frequency patterns (Hz) in statement (above) and YNQ (below) for "O Renato pateta gosta do pássaro" for the male speaker

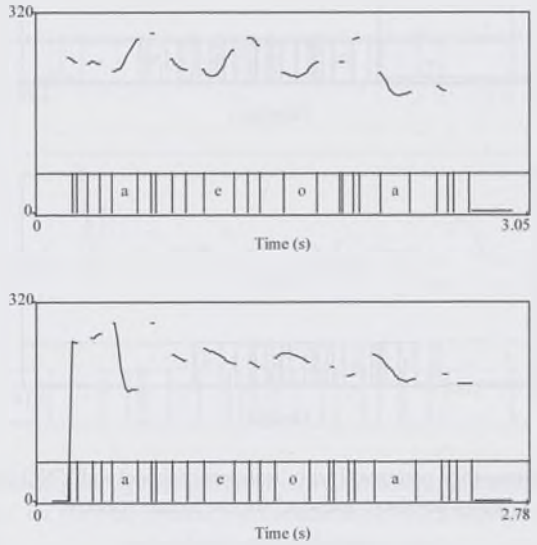


Figure 8: Fundamental frequency patterns (Hz) in statement (above) and YNQ (below) for "O Renato pateta gosta do pássaro" for the female speaker

Figures 7 and 8 illustrate one of the repetitions for each modality of the second sentence uttered by the male (Figure 7) and the female (Figure 8) speakers. The male speaker puts a narrow focus on the qualifier "pateta", whereas the female speaker prefers to maintain the default pitch accent in the first word, even though she also highlights the word "pateta" in the statement. The presence of a antepenultimate stress word at the end of the utterances allow us to confirm that the crucial difference between both modalities is the opposition of a falling (statement) vs a rising-falling (YNQ) tune, with both movements aligned with the stressed vowel. This very fact helped us to interpret the movement in the last stressed word "nadador" (first sentence) as an anticipation of this general pattern.

Observe in Figure 8 that the female speaker puts additional narrow focus in the verb "gosta", strategy that distinguish her from the male speaker.

4. DISCUSSION: INTERSUBJECT AND INTERMODALITY COMPARISONS

The patterns presented in the sections above reveal that both speakers use duration for signalling the lexical stress pattern in words whose penultimate and antepenultimate syllables are stressed. Words ending in a stressed syllable admit, however, the anticipation of the long-short duration pattern to the two final syllables, even though this is not a rule (see the word "nadador" in the female speaker in both modalities and the same word in the YNQ of the male speaker). This can be an indication of the importance of the trochaic pattern in BP words.

The fact that a sequence of images is used in the AMPER project to elicitate the utterances could explain the production of a prosodic pattern whose constituents are closely related to phonological words. These certainly correspond to the individual images in the sequence (nouns, verbs, verbs and qualifiers). Such an occurrence can explain a greater probability of narrowed focused items such as those exhibited especially by the female speaker in her production of the two statements. Narrow focus is implemented by both speakers either by aligning a low tone with the stressed vowel followed by a high tone in the post-stress syllable or by aligning a high tone with the stressed vowel (see section 3.1 for details). Narrow focus on a particular lexical item can deaccent neighbouring items such as the word "Renato" in the YNQ of the male speaker (Figure 7), due to focus on "pateta".

For both speakers, the default melodic pattern for statements is a rising tone associated to the first phonological word followed by a falling tone couched in a strong-weak duration pattern in the last word. Default YNQ melodic patterns start with a rising tune, as for statements, but ends with a steep rising/falling tune in the last word. The great range of pitch accent in the first words of both speakers' YNQs could pragmatically function as an early mark of interrogativeness.

Another distinction between the tonal events of statements and YNQ is, however, worth noting. There is a phonetic difference in the alignment of the falling tune: in YNQs the fall reaches the bottom of the speaker range by the end of the stressed vowel

of the last words; in statements, on the other hand, the f_0 drops slowly during the whole domain defined by the last word.

4.1. Towards a phonological transcription system for BP intonation

Since May 2005, 6 workshops on BP intonation transcription have been conducted by Luciana Lucente, Plínio A. Barbosa and Sandra Madureira in order to propose a notation for BP tones following the general principles of Intonational Phonology (Ladd, 1996) and the set of tones proposed by the ToBI system (Silverman et al., al., 1992). Based on the analysis carried out on these workshops, however, some of ToBI assumptions such as the need for phrase accent tones have been discarded. That is why the phrase accents L- and H- do not appear in the transcriptions that follow. Also from these meetings a consensus default notation for statements and YNQ has emerged and can be summarised as follows, taking the sentences illustrated here as examples. A summary of the discussions held in the workshops can be found in the URL http://geocities.yahoo.com.br/bp_tobi/.

The phonological pattern for the first statement can be represented for both speakers as follows. Tones between parentheses are optional (in this case, realised by the female speaker).

O	pássaro	gosta	do	bisavô	nadador.
	|	|			|
	L*+H	(L*+H)			(H+) !H* L%

The phonological pattern for the first YNQ can be represented for both speakers as follows.

O	pássaro	gosta	do	bisavô	nadador?
	|				|
	L*+H				L+H* L%

The second sentence entails more variability, which can be related to the semantic strength and the syntactic role of a qualifier such as pateta". The phonological pattern for the statement can be represented for both speakers as follows. Tones between parentheses are optional.

O	Renato	pateta	gosta	do	pássaro.
	|	|	|		|
	(L*+)H	L*+H	(L*+H)		H+ !H* L%

The phonological pattern of the second YNQ differs for both speakers in the choice of narrow focus: "pateta" for the male speaker (preceded by the deaccenting of the default initial pattern), or default pitch accent on "Renato" followed by narrow focus on "gosta", for the female speaker. The crucial distinction between the two modalities is transcribed by a rising-falling pattern represented by a bitonal early rising (L+H*) followed by a low boundary tone (L%) in the YNQ. For statements the falling pattern is represented by a bitonal or monotonal falling tune (H+!H* or H*) followed by a low boundary tone (L%).

O	Renato	pateta	gosta	do	pássaro?
		|			|
		H*			L+H* L%

O	Renato	pateta	gosta	do	pássaro?
	|		|		|
	L*+H		!H*		L+H* L%

4.2. Comparing Brazilian and European Portuguese intonation

As far as the analysis of intonation characteristics of declarative sentences in EP and BP are concerned, differences in peak alignment (f_0 peak alignment further right in the prosodic phrase in BP, usually in the post-stressed syllable in BP and within the stressed syllable in EP), initial boundary marking (more frequent in EP) and tone domain (larger in BP) have been found in a *corpus*, recorded with four subjects from Lisbon and four from São Paulo, under investigation by Madureira and colleagues (2004). Analysis of the data revealed interspaced peaks of f_0 and duration in EP corresponding to the stressed syllables of accented words contrasting with an increase of duration culminating at the stressed vowel of the word bearing the phrasal accent and f0 values culminating on the phrasally accented, post-stressed syllable in BP. Such differences in peak alignment can be visualised in Figure 9.

In sum, EP and BP duration contours of stress groups in declarative sentences have been found to display more similarity (duration culminates at the phrasally accented syllable) than f0 contours, since peak delay tends to occur in BP. Other differences are highlighted in the context of the AMPER project, such as the implementation of YNQs. Even though Frota (2002) has shown that YNQs in EP end in a rising tune, with the falling part seemingly restricted to narrowly focussed items, the careful examination of the EP data from the AMPER corpus[1] reveals that there is indeed a falling tune by the end of YNQ, at least in the varieties recorded. Figure 10 illustrates the f_0 patterns during the

1 We thank Lurdes Moutinho for having provided us the EP utterances.

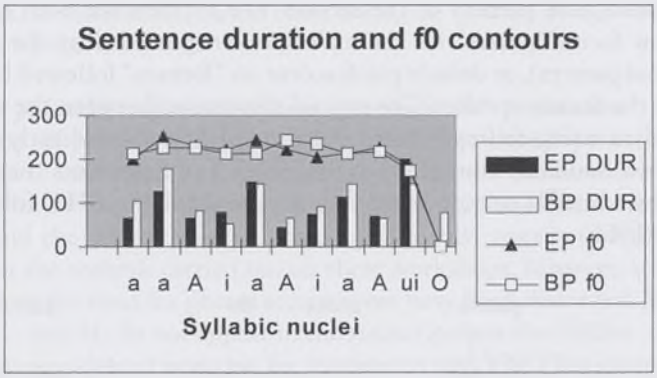


Figure 9. Duration (ms) and f_0 contour (Hz) of the sentence "A gaga ricaça viaja muito". The f_0 values were taken at middle of the syllabic nuclei.

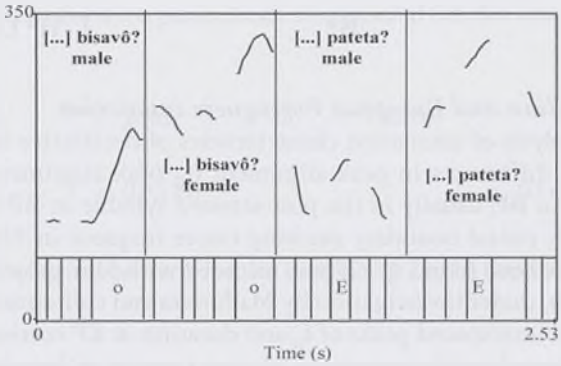


Figure 10: F_0 patterns for words at the end of two YNQs in EP, for two speakers

last words in two YNQs ("O pássaro toca no Toneca bisavô/pateta?") from two speakers from the Algarve region (one male and one female). A similar pattern is found in sentences produced by speakers from Beira Alta. For the sake of comparison, two sentences ending in words bearing the same stress patterns of words in sentence final position in the BP examples (the finally stressed word "bisavô" and the penultimately stressed word "pateta") were chosen. Note that both sentences end in a falling tone. The main difference between BP and EP (as far as intonation of speakers from Algarve and Beira Alta is concerned) seems rather to be related to tonal alignment: EP subjects align the rising tune late in the utterance. Note the opposition between a rising contour during the final stressed vowel in EP versus a falling contour during the stressed vowel in BP.

BIBLIOGRAPHY

BARBOSA, Plínio A.: "Revelar a estrutura rítmica de uma língua construindo máquinas falantes: pela integração de ciência e tecnologia de fala". In: *Estudos de Prosódia*. Scarpa, Ester (Org.). Campinas, Editora da Unicamp, 1999. Pp. 21-52.

BARBOSA, Plínio A.: "O Lugar do Pé Métrico e do Acento no Modelamento Dinâmico do Ritmo". *Letras de Hoje*, 38 (4), 2003. Pp. 135-145.

BARBOSA, P. A., ARANTES, P., SILVEIRA, L. S. "Unifying stress shift and secondary stress phenomena with a dynamical systems rhythm rule". *Proceedings of the 2nd International Conference on Speech Prosody*, March 23-26, 2004, Nara, Japan. Pp. 49-52.

FROTA, Sónia: "Nuclear falls and rises in European Portuguese: a phonological analysis of declarative and question intonation" *Probus*, 14, 2002. Pp. 113-146.

LADD, D. R. *Intonational Phonology*. Cambridge, Cambridge University Press, 1996.

MADUREIRA, Sandra: "Pitch Patterns in Brazilian Portuguese: an Acoustic-Phonetic Analysis", *Proceedings of the Fifth Australian International Conference on Speech Science and Technology*, Sidney, Austrália, 1994. Pp. 156-159.

MADUREIRA, Sandra; GAMA-ROSSI, Aglael; BARBOSA, Plínio A.; FONTES, M. A. S. "Brazilian and European Portuguese contrasted: an experimental acoustic study of speech segments in calsh and non-clash conditions". *Tone and Intonation in Europe. Santorini Conference Abstracts*. September 9-11, 2004, Massaria, Santorini, Greece.

MORAES, João A.: "Intonation in Brazilian Portuguese". In HIRST, Daniel, DI CRISTO, Alberto: *Intonational Systems: a survey of twenty languages*, Cambridge, Cambridge University Press, 1998. Pp.179-194.

SILVERMAN, K. et al.: "ToBI: a Standard for Labeling English Prosody". *Proceedings of the 2nd International Conference on Spoken Language Processing*, Banff, Canada, 1992, Pp. 867-870.

ACOUSTIC CHARACTERIZATION OF RECITING AND SINGING IN A NON-NATIVE LANGUAGE[1]

Silvia C. Barreiro, Eva Estebas-Vilaplana, Isabel Soto
Universidad Nacional de Educación a Distancia

1. INTRODUCTION

It is generally accepted that speakers of a second language find it difficult to achieve the same level of pronunciation accuracy as a native speaker. However, recent research on accentedness (Barreiro, Estebas-Vilaplana and Soto, 2005) indicates that a foreign accent is stronger in certain voice conditions than in others. Pronunciation in "singing" and "reciting" conditions seems to be more natural than in a "speaking" condition.

In this paper we present the results of an acoustic analysis carried out to determine the segmental and suprasegmental features in three voice conditions (spontaneous, reciting and singing) so as to identify the main cues that trigger the different perceptual effect of a more/less natural English pronunciation by Spanish speakers as perceived by native English listeners.

2. EXPERIMENTAL PROCEDURES

The data used in the acoustic analysis was gathered in a prior perception test (Barreiro et al. 2005) that evaluated the degree of accentedness of five Spanish learners of English with different levels of proficiency in three voice conditions (spontaneous speech, reciting and singing) on a five-point scale (1 = "heavy foreign accent", 2 = "heavy to moderate foreign accent", 3 = "moderate foreign accent", 4 = "slight foreign accent" and 5 = "no foreign accent").

---

1 This paper forms part of a research project financed by the Madrid regional government (ref. no. 06/HSE/0123/2004).

The speakers who showed greater degree of improvement in the verse and song conditions as opposed to spontaneous speech were those with intermediate and advanced levels of English, namely, female speaker RP and male speaker PT. The data analysed in the present study included the same samples used in the perception test for these two non-native speakers.

The program *Praat 4.3* (Boersma and Weenik, 2001) was used to analyse the data, enabling us to carry out a time-aligned inspection of the speech waveform, the spectrogram and pitch contours.

3. RESULTS

The information provided by the analysis of the acoustic cues to the perception of the speech segments was first divided according to manner, place and voicing distinctions. Therefore, to identify "manner of articulation" we determined whether the sound was harmonically structured with no noise (vowels -including the vocalic sequences such as diphthongs- approximants and nasals) or whether the sound contained noise (stops, fricatives and affricates). The acoustic cues to "place of articulation" of the periodic speech sounds depended on the formant relationships, whereas for the aperiodic speech sounds, they were mainly based on the frequency of the noise components and the formant transitions, in particular the transitions of the second and third formants (F2 and F3). Finally, the acoustic cues for "voicing" were mostly determined by the presence or absence of the voice bar and relative durations (vowel vs. consonant, frication vs. silence, etc.) rather than frequency or intensity differences.

The acoustic analysis of prosody was based on the study of fundamental frequency (F_0), pitch patterns and range, pauses, tone units and sentence stresses.

We also considered the speech processes that usually occur in connected speech, such as the weakening of unstressed vowels, pre-fortis clipping, weak forms, elisions or assimilations.

3.1. Results of the speaker with an intermediate level of English (RP)[2]

3.1.1. Spontaneous Speech

> "I have two children, one the elder one, is studying for a lawyer, and the second one, is studying for to be a medicine, a doctor."

RP's spontaneous speech was judged as having a "heavy to moderate foreign accent". We cannot tell for sure from our study how the listeners made the accentedness judgements. Presumably, they assessed the pronunciation by comparing the samples with their own native-like version. But what is obvious is that their judgments were

2 Due to space restrictions we will present here the results obtained in the spontaneous and singing conditions only.

influenced by non-phonological variables such as grammatical errors. RP's speech clearly included inappropriate use of prepositions, incorrect words and wrong connectors as well as hesitations and self-corrections.

The grammatical errors were accompanied by pronunciation errors, both at the segmental and suprasegmental levels. The segmental errors were either phonemic or phonetic. Following Munro & Derwing (1999), by phonemic errors we considered the elision or addition of a segment, or the substitution of one segment by another English one. By phonetic errors we meant the production of a segment in such a way that was noticeably non-native or not appropriate for that phonetic context. The suprasegmental errors were stated in relation to fundamental frequency, pitch (patterns and range), pauses, intonation contours and sentence stress, weakening of sounds, as in Llisterri (1992).

3.1.1.1. Segmentals

The phonemic errors mostly affected the production of certain consonants and they were reflected in the substitution of one segment by another English one, as in the following examples: a) the final consonant in the word '*have*' was voiceless [f], as shown by its long duration and the absence of the voice bar on the spectrogram; b) the initial consonant in '*the*' was always a plosive [d] instead of a fricative [ð], as reflected in the transient duration of the noise component; and c) the final nasal in '*studying*' was always alveolar instead of velar, as seen by the positive transitions of the F3.

The acoustic analysis of all the segments showed that a great number were not perceived as being either English or Spanish. Phonetic errors were made both in consonants and vocalic sequences, with a noticeably non-native production of segments.

The most prominent vocalic errors were the following: a) the vowel in the word '*have*' was not an open front English vowel [æ] but a Spanish [a], with formant values (F1: 770, F2: 1493) matching those of the Spanish sound (Quilis, 1988, 1993); b) the weak vowels in '*children*' and '*and*' were not a *schwa*, but an English vowel [e] (F1: 607, F2: 1891), while in '*second*' this weak vowel was a mixture of a Spanish sound [o] and the English /ɒ/, according to its formant values (F1: 550, F2: 808); and c) the vowel in '*lawyer*' was short and it had formant values similar to an English vowel /ɒ/ (F1: 596, 1253).

The most prominent consonant errors were the following: a) the production of the initial consonant in the word '*have*' was clearly a mixture of velar and glottal fricatives [h] and [x] -loser to the Spanish [x]- as shown in the overall high intensity, slight vertical striations and its resistance to voicing (Marrero, 1990, Quilis, 1993); b) the initial plosive in '*two*' was aspirated with an over-marked release, making the sound unnatural for a native speaker; c) the lateral approximant in '*children*' and '*elder*' was not velarised, in accordance with the context, as its F2 values were not in the range of 900-1600 Hz; d) the voiced plosive in '*studying*' was produced as a fricative [ð] -as normally happens in that phonetic context in Spanish- as shown on the spectrogram by a lack of silence and a clear frication; and e) the second consonant in '*lawyer*' showed a certain amount of frication on its spectrogram, indicating the presence of a fricative [ʝ] instead of approximant [j].

3.1.1.2. Suprasegmentals

The analysis of prosody found errors in relation to rhythm, sentence stress and pitch range, with an absence of the stress-timed rhythm used in English.

RP's pitch range varied in a range of 190 Hz, between a minimum pitch value of 177 Hz and a maximum value of 367 Hz, a pitch range too narrow for English. RP made use of clearly non-native pitch patterns, which were almost flat with a rise on the final word of each sentence. Our results were consistent with those obtained in previous studies (Jilka, 2000; Munro, 1995), showing that the sentence-level intonation seems to be significantly correlated with the degree of perceived foreign accent.

There was a wrong placement of sentence stress. The main stress was always placed at the end of each sentence -mainly marked by a rising of the F_0- and some of the other stresses were wrongly located, as in: *'I have two children'* instead of saying *'I have two children'*.

Finally, RP's speech lasted 15,69 seconds, uttered at a speaking rate slower than optimal for a non-native speaker (MacKay & Flege, 2004; Munro & Derwing, 1998 and 2001).

3.1.1.3. Connected speech processes

There was an overall poor usage of English connected speech processes. Firstly, a lack of weak forms was observed in functional words, such as *'a'*, *'the'*, or *'and'*. Secondly, there were no weak vowels in unstressed syllables, since they were pronounced as shown orthographically: *'elder'* or *'children'* as a kind of /e/, *'second'* as a kind of /o/. Their duration was not reduced either. Finally, the speaker did not make use of elisions between words, stretching at the end of a sentence or pre-fortis clipping. Nevertheless, RP produced assimilations, as in the final consonant in *'is'* in the sequence *'is studying'*, which was generated without voicing.

3.1.2. Singing

> "Oh, I come from Alabama with my banjo on my knee. Oh, I'm going to Louisiana, my true love for to see."

The listeners judged RP's singing as having a "moderate-to-slight foreign accent". There were no grammatical errors that might have influenced the listeners' evaluations.

The acoustic analysis of her speech showed pronunciation errors mainly at the segmental level.

3.1.2.1. Segmentals

As with spontaneous speech, a certain amount of sounds were not perceived as being either English or Spanish. Phonemic errors affected both consonants and vowels resulting in elisions or substitutions of segments, such as: a) the vowel in the word *'to'* was pronounced as [u] (F1: 360, F2: 1489) instead of a *schwa*; b) there was no presence of an [i]-type vowel in *'Louisiana'*; or c) the final nasal in *'going'* was always alveolar instead of velar, as reflected by the positive transitions of the F3.

The phonetic errors made were of the same type as in spontaneous speech, affecting both consonants and vocalic sequences, and nearly always producing non-native sounds.

The most prominent vocalic errors were the following: a) the diphthongs in '*going*' and in the sequence '*Oh I*' suffered from smoothing, as there was a reduction of articulatory movements: [ʊ] ended in [o] sound, similar to the English one [ɒ] (F1: 626, F2: 1014); b) the vowel in '*come*' showed a formant structure (F1: 769, F2: 1957) which was a mixture of [æ] and [ʌ], while the vowel in '*love*' showed a formant structure (F1: 580, F2: 1446) similar to a Spanish sound [a]; c) the vowels in '*from*' and '*to*', were not pronounced as a *schwa* but as a vowel similar to the Spanish sound [o] (F1: 548, F2: 1149); and d) the vowel in "*banjo*' was a mixture of the Spanish [a] and the English [æ], according to its formant values (F1: 649 and F2: 1827), while the vowel in '*Alabama*' was a mixture of the Spanish [a] and the English [ɑ], with formant values: F1: 665, F2: 1293.

The most prominent consonant errors were the following: a) the initial plosives in the words '*come*' and '*true*' were over aspirated, making the sound unnatural for a native speaker; and b) the plosive in '*Alabama*' was produced as a fricative [ð] -shown on the spectrogram by a lack of silence and a certain amount of frication- as is usually the case in that phonetic context in Spanish.

3.1.2.2. Suprasegmentals

The analysis of prosody showed barely noticeable suprasegmental errors as the song prompted the prosodic features.

RP used native pitch patterns with great variation and intonation patterns associated with the English language. The pitch range varied around 352 Hz between a minimum pitch value of 117 Hz and a maximum value of 469 Hz, a pitch range more consistent with the English language.

The duration of the two paragraphs were alike (4,77 and 4,47 seconds, respectively), and there was an accurate placement of sentence stress on the syllables marked by the rhythmic pattern, that is, on '*come*', '*Alabama*', '*with*', '*banjo*','*on*', '*knee*', '*going*', '*my*', '*true*', '*for*', '*see*'. The main stress was clearly marked by a rising of the fundamental frequency (F_0), although RP made use of duration and intensity as well. There were regular rhythmic beats at an average of 0,6 seconds, showing an English stressed-timed rhythm. Finally, the speaking rate was perceived as 'appropriate' for the song.

3.1.2.3. Connected speech processes

There was a slight improvement in connected speech processes, as compared to spontaneous speech.

RP did use stretching at the end of each sentence and also pre-fortis clipping, for example as happened in '*banjo*', in which the affricate seemed to have lengthened the previous vowel ([æ]: 0,10 s., [n]: 0,8 s., [dʒ]: 0,10 s.). A similar situation was observed in the final fricative of '*with*', in that it was produced without voicing but with a lengthening of the previous vowel ([ɪ]: 0,27 s. [ð]: 0,11 s.).

Nevertheless, a lack of weak forms remained noticeable in functional words such as '*from*' or '*to*'. Also, there was no weakening of vowels in unstressed syllables, since they were pronounced as shown orthographically: '*Alabama*', '*Louisiana*', as a kind of [a]. Finally, the speaker did not make use of assimilations between words either, as in the consonant in '*on*' in the sequence '*on my*', which was alveolar instead of bilabial.

3.2. Results of the speaker with an advanced level of English (PT)[3]

3.2.1. Spontaneous Speech

> "... and it's four years since I am living here and I miss them a lot, and I try to go there to see them once a month if I can, if not, once every two months."

The listeners judged PT's spontaneous speech as having a "heavy to moderate foreign accent" -closer to "moderate"-. His speech contained grammatical errors (for example, he said '*it's four years since I'm living here*' instead of '*I've been living here for four years*') that possibly influenced the judges. These grammatical errors were accompanied by pronunciation errors.

3.2.1.1. Segmentals

The analysis of all the segments showed that the main errors consisted of inappropriate use of certain allophones. These phonetic errors affected both consonants and vocalic sequences. Thus, the vowel in '*once*' showed a formant structure (F1: 715, F2: 1248) that was more similar to the sound [ɑ] rather than a vowel [ʌ]. Other examples of vocalic errors were the shortening of vowels and diphthongs in the words '*four*', '*if*' and '*can*'; also, the diphthong in '*I'm*' was reduced to a single sound with the elimination of the second element, together with the shortening of the first one. As far as the consonants are concerned, there was an over-aspiration in the plosives in '*tray*', '*two*' and '*to*'.

The phonemic errors were made both in consonants and vocalic sequences, and resulted in the substitution of one segment by another English one, or elisions, such as, a) the production of the final nasal in '*living*' as alveolar instead of velar, as shown by the positive transitions of the F3; and b) the omissions of consonants in the words '*month*', '*months*' and '*living*'.

3.2.1.2. Suprasegmentals

The analysis of suprasegmentals showed that the speaker made few errors in rhythm and sentence stress, although overall there was an appropriate use of prosodic features: there was a suitable location of sentence stress -mostly marked by a change in the fundamental frequency (F_0)-, an appropriate speaking rate and an English-like use of

3 Due to space restrictions we will present here the results obtained in the spontaneous and reciting conditions only.

pitch patterns and a wide pitch range, changing between a minimum pitch value of 90 Hz and a maximum value of 465 Hz.

3.2.1.3. Connected speech processes

The speaker produced an unexpected reduction in the duration of unstressed content words, such as *'it's'*, *'living'*, *'months'*, *'go'*, *'every'*, with the consequent omission of vocalic and/or consonant phonemes. This phenomenon also affected the diphthongs of the words *'I'm'*, *'here'*, *'try,'* *'go'* and *'there'*, with the elimination of the first element, instead of the second one. However, overall there was an appropriate of sentence stress and a suitable use of weak forms in functional words such as *'a'*, *'the'*, or *'and'* in most cases. PT also made use of assimilations, such as the devoicing of the final consonant in *'years'* in the sequence *'years since'*.

3.2.2. Reciting

"It rained all night the day I left
The weather it was dry
The sun so hot
I froze to death,
Susanna, don't you cry!"

PT's reciting voice was judged as having "moderate to slight" foreign accent. There were no grammatical errors that could have influenced the listeners' evaluations.

The acoustic analysis of his speech showed pronunciation errors mostly at the segmental level.

3.2.2.1. Segmentals

The analysis of all the segments showed an overall reduction in the number of segmental errors that nearly always consisted in producing non-native sounds or inappropriate use of allophones. There was a shortening of vowels and diphthongs, such as the vocalic sequences in the words *'if'* or *'dry'*. Furthermore, the second element of the diphthong in *'dry'* was not a close vowel but a more open sound (F1: 387, F2: 1928), similar to a *schwa*. The most prominent consonant error was the excessive amount of aspiration of plosives, as in *'cry'*.

Few phonemic errors occurred in both consonants and vocalic sequences, as reflected in the addition of epenthetic elements or elision of sounds: the final consonant in *'rained'* was elided, and a very short vowel [ˇ] was added between the words *'night'* and *'the'*.

3.2.2.2. Suprasegmentals

Few prosodic errors were perceived in relation to sentence stress and pitch patterns. The sentence stresses were mainly marked by a change in the fundamental frequency (F_0), although PT made use of duration and intensity contrasts as well. There

were suitable sentence stresses on the syllables marked by the rhythmic patterns, that is, on *'rained'*, *'night'*, *'left'*, *'weather'*, *'dry'*, *'sun'*, *'hot'*, *'froze'* and *'Susanna'*. However, PT did not place stresses on the words *day, death. don't*, as one would have been expected.

However, the pitch range was wider than in spontaneous speech. It varied in a range of 407 Hz, among a minimum pitch value of 89 Hz and a maximum value of 496 Hz. This pitch range sounded quite natural for English. The pitch patterns used throughout the whole text had a considerable variation and an intonation acceptable for the English language.

PT made use of a stress-timed English rhythm with regular rhythmic beats at an average of every 0,4 seconds, reciting at an appropriate speaking rate. There was a suitable use of pauses. PT even added interpretive pauses, as in the sequence *'the weather (...) it was dry'*.

3.2.2.3. Connected speech processes

The number of processes in connected speech increased as compared to the spontaneous speech. Thus, an appropriate use of weak forms was observed in words such as *'to'*, *'was'*, *'the'* or *'and'*. Also, the speaker PT made use of assimilations between words, such as the devoicing of the final consonant of *'froze'* in the sequence *froze to.* PT even produced a fusion of phonemes or coalescence in the sequence *'don't you'* where the stop and approximant became an affricate with the obscuration of the final weak vowel to a *schwa.*

Furthermore, PT used pre-fortis clipping, as happened in the words *'weather'*, *'froze'* or *'Susanna'*, or in the sequences *'was dry'* or *'to death'*, in which the consonants were also devoiced. Finally, PT also made use of stretching at the end of each sentence, noticeably at the end of the whole paragraph.

On the other hand, PT had a tendency to reduce the duration of the first element in diphthongs, not very suitable in the English language, even in the case of smoothing. As examples: *'day'* ([e] 0,04, [ɪ] 0,05), *'dry'* ([a] 0,07, [ɪ] 0,10), *'so'* ([ˈ] 0,07, [ʊ] 0,10), *'froze'* ([ˈ] 0,05, [ʊ] 0,06), *'don't'* ([ˈ] 0,04, [ʊ] 0,16) and *'cry'* ([a] 0,09, [ɪ] 0,12).

4. CONCLUSIONS

Both speakers produced segmental errors in the three voice conditions. However, the acoustic analysis showed an improvement in the pronunciation of reciting and singing as opposed to spontaneous speech. This was due to a reduction of prosodic errors in these two voice conditions as compared to spontaneous speech.

Therefore, though, as in previous studies (Magen, 1998; Major, 1986), we cannot tell for sure from our study which specific aspects of pronunciation were more crucial for the perception of a greater or lesser degree of foreign accent in the different types of voice conditions uttered by the two non-native speakers, it seems that segmental errors had less of a detrimental effect than suprasegmental errors when it came to perceiving and judging non-native pronunciation.

The results presented in this paper are in accordance with those studies that argue that prosodic errors are more serious than segmental errors (Anderson-Hsieh, Johnson & Koehler, 1992; Johansson, 1978; Palmer, 1976) and in disagreement with those that show evidence of the greater detrimental effect of segmental errors on comprehension (Koster & Koet, 1993, and Fayer & Krasinski, 1987).

Furthermore, it reinforces our view that phonetics teaching needs to be reappraised and a top-down approach adopted, that is, from a suprasegmental to the segmental level, in which the prosodic features play an important role in the teaching and learning of the pronunciation of English to non-native speakers. In the near future, we intend to apply this approach to different groups of Spanish learners of English to determine its effect on pronunciation and incorporate songs and poems as teaching materials.

A great deal of additional research is needed to better identify those aspects of pronunciation (mainly at the suprasegmental level) that most affect the perception of accent in non-native speech, and under different voice conditions, in order to bring new insights to bear on the discussion on hierarchies of pronunciation errors.

BIBLIOGRAPHY

ANDERSON-HSIEH, J., JOHNSON, R. & KOEHLER, K. (1992): The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody and syllable structure. *Language Learning*, 42 (4), 529-555.

BARREIRO, S., ESTEBAS-VILAPLANA, E. & SOTO, I. (2005): "Teaching Phonetics through singing and reciting", *Proceedings of the Phonetics Teaching and Learning Conference 2005*. London: University College London.

BOERSMA, P. & D. WEENINK (2001): *Praat: a system for doing phonetics by computer*. available at http://www.praat.org.

ANDERSON-HSIEH, J., JOHNSON, R. & KOEHLER, K. (1992): "The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody and syllable structure", *Language Learning*, 42 (4), 529-555.

FAYER, J.M., & KRASINSKI, E. (1987): "Native and nonnative judgments of intelligibility and irritation", *Language Learning*, 37, 313-326.

JILKA, M., (2000): "Testing the Contribution of Prosody to the Perception of Foreign Accent", *New Sounds*, 4: 199-207.

JOHANSSON, S. (1978): "Studies in error gravity: Native reactions to errors produced by Swedish learners of English", *Gothenburg Studies in English*, 44. Gothenburg, Sweden: University of Gothenburg, Department of English.

KOSTER, C.J. & KOET, T. (1993): "The evaluation of accent in the English of Dutchmen", *Language Learning*, 43, 69-92.

LLISTERRI, J. (1992): "Speaking styles in speech research" [en línea] *ELSNET/ESCA/SALT Workshop on Integrated Speech and Natural Language*. Dublin, Ireland, 15-

17 July 1992. Available: http://liceu.uab.es/~joaquim/ publications/Speaking-Styles_92.pdf

MACKAY, I. R.A. & FLEGE, J.E. (2004): "Effects of the Age of Second-Language (L2) Learning on the Duration of L1 and L2 Sentences: The Role of Suppression" *Applied Psycholinguistics*, 25: 373-396.

MAGEN, H. (1998): "The Perception of Foreign-Accented Speech", *Journal of Phonetics*, 26: 381-400.

MAJOR, R.C. (1986): "Paragoge and Degree of Foreign Accent in Brazilian English", *Second Language Research*, 2(1): 53-71.

MARRERO, V. (1990): "Estudio acústico de la aspiración en español", *R.F.E.*: 345-397.

MUNRO, M.J. (1995): "Nonsegmental Factors in Foreign Accent", *Studies in Second Language Acquisition*, 17(1): 17-34.

MUNRO, M.J. & DERWING, T.M. (1998): "The Effects of Speaking Rate on Listener Evaluations of Native and Foreign-Accented Speech", *Language Learning*, 48(2): 159-182.

MUNRO, M.J. & DERWING, T.M. (1999): "Foreign Accent, Comprehensibility, and Intelligibility in the Speech of Second Language Learners", *Language Learning*, 49(1): 285-310.

MUNRO, M.J. & DERWING, T.M. (2001): "Modelling Perceptions of Accentedness and Comprehensibility of L2 Speech", *Studies in Second Language Acquisition*, 23: 451-468.

PALMER, J. (1976): "Linguistic accuracy and intelligibility", *Proceedings of the 4th International Congress of Applied Linguistics* (pp. 505-513). Stuttgart, Germany: Hockschul Verlag.

QUILIS, A. (1988): *Fonética Acústica de la Lengua Española*. Madrid: Gredos, Biblioteca Románica Hispánica.

QUILIS, A. (1993): *Tratado de Fonología y Fonética Españolas*. Madrid: Gredos, Biblioteca Románica Hispánica.

EL TECNEFÓN: UNA MÁQUINA PARLANTE EN LA ESPAÑA DEL XIX. APROXIMACIÓN FONÉTICA Y ACÚSTICA

ELENA BATTANER MORO
Universidad Rey Juan Carlos (Madrid)

INTRODUCCIÓN

El *Tecnefón* es una máquina parlante diseñada por Severino Pérez y Vázquez durante la segunda mitad del siglo XIX; se trata, en principio, de la primera máquina diseñada con el propósito de la síntesis de voz en España. Pérez dio noticia de su tarea investigadora a través de dos obras: *El Tecnefón. Memoria presentada al Excmo. Ministro de Fomento en virtud de Real Orden*, publicada en 1865; y *El Tecnefón y los medios representativos de la palabra,* publicada en 1868. En ellas expone, no sin ciertas pretensiones literarias, un esbozo de la máquina, sus posibles aplicaciones y la descripción de su propia investigación.

En este trabajo se ofrece una primera noticia de esta máquina, centrándonos no tanto en la historia de su creación (Battaner Moro, en prensa[1]) como en la teoría acústica que hay detrás de ella. Este artículo se divide en tres secciones. La primera repasa la historia moderna de las máquinas parlantes y el desarrollo de la física acústica, con objeto de contextualizar la tecnología del Tecnefón. La segunda presenta una descripción de las obras de Pérez a través de los aspectos más interesantes a este respecto: las alusiones a autores e investigaciones del punto anterior y cómo aplica o no estos conocimientos en su propia investigación. La tercera sección es una descripción del funcionamiento de la máquina y de la teoría acústica que propone el autor para la lengua española.

---

1 En ese trabajo se ofrece un análisis crítico y genealógico acerca del Tecnefón y de su autor, así como del lugar teórico e institucional que poseen en la lingüística española del siglo XIX. Debido a los intereses de este congreso, se ha aligerado la carga historiográfica del presente trabajo; de esta forma, se ha dado prioridad a los contenidos fonéticos y acústicos.

1. PRIMEROS DESARROLLOS DE LA SÍNTESES DE VOZ

A lo largo de los siglos -incluso hoy en día- y a través de muy distintos lugares, la creación de objetos (o sujetos) que pueden hablar recorre el ideario literario y mítico de las tradiciones orientales y occidentales. Autómatas androides, cabezas encantadas, máquinas parlantes... son los ejemplos más conocidos de la aspiración humana por crear lenguaje artificial. En esta larga historia podemos viajar, por ejemplo, desde los trabajos de Herón de Alejandría, la cabeza parlante de Bacon, los autómatas de Alberto Magno o Descartes, la cabeza encantada en el Quijote o la máquina de Kempelen, hasta el VODER, los sintetizadores de voz o los *robots* más modernos. Ya fueran objetos de feria, fraudes o androides de leyenda, ya verdadera ingeniería de la síntesis de voz, todos ellos recogen esa aspiración por reproducir de manera artificial una de las cualidades de lo humano; esto es, el lenguaje.

El desarrollo científico y tecnológico del siglo XVIII comienza a dejar atrás la leyenda y nos sitúa en el umbral de la ciencia actual, por lo que la creación de sonidos humanos artificiales comenzará a sostenerse con investigaciones que, hoy en día, consideramos claramente *científicas*. En el proceso de consecución de habla humana artificial conviven y se entremezclan (o se confunden) dos puntos de vista: un punto de vista articulatorio -las máquinas diseñadas reproducen el aparato fonador-, y un punto de vista *acústico* -las máquinas imitan el sonido del lenguaje humano-.

Dentro del vasto panorama de androides y cabezas encantadas, la historia moderna de la síntesis de voz puede empezar con la obra del abad Mical, que presentó en 1783 dos cabezas que mantenían un pequeño diálogo entre ellas. La obra de Mical fue objeto de un informe de la Academia de Ciencias francesa –firmado, entre otros, por el mismo Lavoisier–; el informe fue favorable pese a que la calidad de las voces no era buena y a que la *coarticulación* no parecía del todo lograda. En cualquier caso, las cabezas parlantes de Mical, así como otros autómatas parlantes atribuidos a autores de la época, no eran exactamente máquinas *productivas*; esto es, sólo reproducen lenguaje. Y así, tienen más de autómata que de máquina parlante. Sin embargo, la disección de las cabezas de Mical permite intuir, al menos, dos partes: una cámara de aire, alimentada por fuelles, en la que diferentes válvulas llevan el aire hacia ciertas cavidades, donde el aire se transforma en sonido; y un cilindro que mueve algunas lengüetas que, a su vez, moldean el sonido que procede de la cámara anterior y lo convierten en lenguaje articulado. Pese al automatismo de las cabezas de Mical, los elementos de que consta se repetirán en las máquinas sucesivas: fuelles, lengüetas, tubos... dispuestos a modo de pulmones, glotis y boca, o con los que se imitará el sonido de la voz.

Otro de los ejemplos clásicos en esta historia de la síntesis de voz es el de Wolfgang von Kempelen (1734-1804). Aunque la bibliografía suele denominar al diseño de Kempelen como 'máquina parlante', en realidad interesa apuntar que se trata de un aparato que representaba de una manera aproximada los órganos de la fonación; ésta es la razón por la que algunos autores hablan de 'diseño antropomórfico' (Lyénard 1991). Kempelen estudió el mecanismo de la producción del lenguaje humano durante cerca de

veinte años, y escribió y publicó en 1791 el libro titulado *Mechanismus der menschlichen Sprache nebst Beschreibung einer sprechenden Maschine*; en esta obra, además de otras cuestiones relativas al lenguaje, Kempelen dedica interesantes capítulos a la descripción fisiológica del habla y al diseño de esta máquina. Gracias a las indicaciones ofrecidas en este libro, la máquina de Kempelen se ha podido reproducir hace sólo unos años.

Para Kempelen, 'For a speaking machine no more is required than a lung, a glottis, and a mouth' (cit. por Asher (ed.) 1994: 3083). Su máquina, entonces, consistía en un fuelle (los pulmones) que hacía pasar el aire por una serie de pequeños conductos que poseían membranas o lengüetas (glotis), y que, de esta forma, convertían el aire en sonido. Dicha máquina poseía además otro conjunto de elementos que daba forma al sonido y creaba, por consiguiente, sonidos articulados del lenguaje: por ejemplo, una boquilla de cuero con la que podía articular las vocales, o unas palancas para conseguir sonidos fricativos. Según las descripciones del propio Kempelen, de algunos de sus contemporáneos y de autores que han conseguido construir esta máquina, su diseño era capaz de producir diecinueve consonantes y cinco vocales[2].

La idea de Kempelen representa prototípicamente el punto de vista de reproducción casi literal, por así decirlo, del habla humana. Esto se conseguía, como hemos visto, a través de pretender producir artificialmente el habla mediante la representación o reproducción –o mejor, de la abstracción– del aparato fonador, aunque también es cierto que algunas partes de la máquina de Kempelen están directamente relacionadas con la percepción, como por ejemplo la realización de las fricativas con palancas, y no dentro de la boca[3].

Unos sesenta años después de la publicación de Kempelen, el conocido investigador inglés Charles Wheatstone (1802-1875) rediseñó su máquina (una reproducción que conoció, ni más ni menos, Alexander Graham Bell (1847-1922)). Es interesante apuntar que a Wheatstone se le atribuye el *descubrimiento* de que la cualidad de algunas vocales tiene que ver con la longitud del tracto vocálico y no exactamente con su diámetro. En realidad, lo que subyace es la idea de que el tracto vocálico es un filtro acústico y que, como tal, selecciona ciertas frecuencias y no otras. Este es el lugar que marca un antes y un después en la historia de estas investigaciones. De esta forma, y para conectar con el punto de vista acústico de la producción artificial de voz, también podemos señalar que el conocido físico y fisiólogo H.L. Helmholtz (1821-1894) se basó en algunos de los trabajos de Wheatstone; así, del británico también es la idea de que el tono de la cavidad bucal se ve excitado (o reforzado) por ciertas frecuencias de la fundamental (la fuente).

---

2 Kempelen caracterizó las vocales en función de dos factores: la apertura de los labios y la apertura de lo que denominaba el 'canal de la lengua'. Para cada factor estableció cinco grados, del más estrecho al más ancho: el primer caso daba como resultado {u, o, i, e, a}; y el segundo, {i, e, a, o, u} -parece que Kempelen era capaz de distinguir el *pitch* de las vocales-. Haciendo un poco de abstracción, podemos afirmar que, en la medida de la época, Kempelen *anticipó* la presencia de al menos el primer formante de las vocales (el que correlaciona con el lugar de articulación: anterior-posterior).

3 En cualquier caso quedan muchas cuestiones fonéticas sobre el lenguaje humano que no fueron representadas en la máquina de Kempelen, básicamente porque en la época no eran del todo conocidas. Más información sobre este inventor puede encontrarse en, entre muchos otros, Brocke (1983), Lyénard (1991) o Schroeder (1993).

En este momento comenzamos a mezclar descripción articulatoria y descripción acústica de los sonidos del lenguaje, lo que conduce hacia ese segundo punto de vista en el desarrollo de la producción artificial del habla, y que es aquel en el que podemos situar el trabajo de Pérez. Atrás dejamos a personajes como Kempelen, aunque es necesario que nos detengamos de nuevo en su época para hablar del hombre que inaugura los estudios fundamentalmente acústicos.

El danés Christian Gottlieb Kratzenstein (1723-1795) abre paso al siglo XIX y a toda la investigación acústica del sonido en general y a la del sonido humano en particular –entre otros; por ejemplo, Ernst Chladni (1756-1827)–. Hacia 1770, la Academia Imperial de San Petersburgo convocó un concurso para intentar explicar las diferencias fisiológicas del habla con relación a las vocales. Kratzenstein ganó dicho concurso gracias al diseño y producción de los denominados 'resonadores de Kratzenstein', un conjunto de tubos –en teoría modelados a propósito del tracto vocálico– que imitaban los sonidos de estas cinco vocales. La cuestión estriba en el diferente diseño de los tubos y en la presencia de lengüetas, unas lengüetas que, de aquí en adelante, van a adquirir cierto protagonismo en las restantes máquinas parlantes.

Por último, y antes de entrar en el Tecnefón, es necesario detenerse en Joseph Faber y su máquina *Euphonia* (terminada hacia 1835; Liénard 1991). Poco se sabe de Joseph Faber, un oscuro profesor de matemáticas vienés. A diferencia de Kempelen y otros autores, Faber no dejó obra escrita alguna, y lo que conocemos de su máquina lo sabemos a través de testimonios de contemporáneos que pudieron verla en funcionamiento. *Euphonia*, pese a toda la base científica con la que Faber construyó su máquina, no pasó de ser, hasta cierto punto, un artículo de feria. De hecho, los testimonios a los que antes hacía referencia tienen que ver con este tipo de visitas: previo pago de un chelín, por ejemplo, se podía acceder a una sala donde se encontraban la máquina y su inventor y se asistía a su funcionamiento

La máquina de Faber podía hablar con voz normal o murmurada y, gracias a un segundo teclado, también podía cantar. *Euphonia* consistía en un conjunto de fuelles accionados por pedales que, de nuevo, era la fuente del viento; poseía asimismo seis diafragmas que podían modificar la forma y la sección de una especie de tracto vocálico, formado a su vez por unas mandíbulas articuladas y una lengua flexible. El teclado estaba dividido en catorce llaves o claves; presionándolos, se conseguían los siguientes sonidos: [a, o, u, i, e, l, r, v, f, s, ʃ, b, d, g]. Como afirma Lyénard (1991: 23), "the control of *Euphonia* was practically phonemic in nature, even tough some controls remained physiological or acoustical (nasality, voicing)". Es interesante apuntar que una de las lengüetas estaba dispuesta de tal manera que podía generar y controlar la entonación[4]. Para una pequeña historia de la síntesis de voz, puede consultarse la siguiente página web: [http://haskins.yale.edu/haskins/HEADS/simulacra.html].

4 La verdad es que no se trata exactamente de una novedad, porque algo de esto ya apuntó el mismo Kempelen; sin embargo, bien es cierto que se trata de la primera máquina que concibe específicamente los denominados *rasgos suprasegmentales*.

El autor español del que aquí damos noticia desarrolló el prototipo de una máquina que recoge muchas de las aportaciones anteriores, sobre todo aquellas referidas a la acústica y al uso de membranas, lengüetas, etcétera. Sin embargo, la tarea historiográfica es, en este punto, compleja; no tanto en lo que a Severino Pérez se refiere –no poseemos apenas datos sobre el autor–, sino a la investigación en sí que llevó a cabo. Es verdad que existen dos obras de este profesor y que en ellas hay indicaciones precisas sobre algunos aspectos; pero, en general, lo ofrecido por Pérez en lo que tiene que ver con la descripción de la constitución y del funcionamiento interno de la máquina es ciertamente escaso. Todo esto lo veremos más adelante.

2. SEVERINO PÉREZ Y EL *TECNEFÓN*

Además de su tarea docente –que, según comenta el propio autor, le quitaba bastante tiempo y no le reportaba muchos beneficios económicos (de ahí las solicitudes de ayuda al Ministerio de Fomento a instancias de la Sociedad Económica Matritense)–, parece que Severino Pérez invertía parte de su tiempo en el aprendizaje de la física acústica con la intención de desarrollar una máquina parlante. Hay pocos datos acerca esta máquina, como apuntábamos antes, y en su mayoría procede de las mismas obras publicadas por Pérez.

Pérez acudió a la Sociedad Económica Matritense de Amigos del País hacia abril de 1864 en busca de financiación para sus estudios. Tras una presentación del autor y un entusiasta informe elaborado por la sección encargada de estudiar la solicitud de Pérez, se elevó al Ministerio de Fomento una petición de ayuda económica para el inventor. El Ministerio, por su parte, requirió un informe de la Real Academia de Ciencias; este informe, sin embargo, fue poco favorable. Gracias a la transcripción de dicho informe que aparece en el libro de Pérez de 1868, podemos saber que en poco o nada estimaron su obra; en mi opinión, los académicos encargados del informe no fueron conscientes del alcance teórico de la máquina, probablemente porque estuvieran menos versados que Pérez en física acústica. Pérez apuntó haber cometido "un delito de lesa academia", y es que pueden haber pesado mucho algunos factores institucionales, como el hecho de que Pérez pudiera ser considerado un advenedizo de la ciencia –esto se desprende de algunos comentarios del informe–. En cualquier caso, Pérez terminó abandonando el proyecto, siguiendo así con una especie de maldición que rodea a los inventores de cabezas o máquinas parlantes.

Lo que conocemos del Tecnefón se incluye en los dos libros publicados por Pérez a este respecto: *El Tecnefón. Memoria presentada al Excmo. Ministro de Fomento en virtud de Real Orden* (Madrid, 1865; en adelante *Memoria*) y *El Tecnefón y los medios representativos de la palabra* (Madrid, 1868; en adelante *Tecnefón*). La *Memoria* fue publicada en 1865 a instancias de dar cuenta de lo realizado gracias a la subvención recibida por parte del ministerio de Fomento. No contiene índice ni partes en que se pueda dividir dicha memoria; contiene un grabado del aspecto externo de la máquina, aunque poco aporta dicho grabado para su estudio. La mayor parte de esta obra, por no decir la obra entera, se recoge en el segundo libro de Pérez. Esta obra sí contiene un índice:

I.	La inteligencia y la palabra natural.
II.	Recursos gráficos para dar fijeza al pensamiento haciendo permanente el lenguaje hablado, y perfecciones que añaden la palabra impresa y telegrafiada.
III.	Palabra artificial ó tecnefón teórico: Etimología.
IV.	Análisis literal.
V.	Síntesis conexiva.
VI.	Tecnefón aplicado: Perennidad de un idioma.
VII.	Adelantos recientes.
VIII.	Pronunciación eléctrica.
IX.	Resumen comparativo.
	Apéndice.
	Oda.

La novedad con respecto a la *Memoria* estriba, básicamente, en los puntos I-III, en los VI-IX, en el Apéndice y en la *Oda* que cierra la obra y que Pérez escribió para inaugurar el Tecnefón.

Lo que llegó a existir del *Tecnefón* sólo se conoce por las indicaciones del mismo Pérez y de algún contemporáneo; parece que lo que este profesor llegó a construir fue un prototipo con el *nexo parcial* de las cinco vocales y las siguientes consonantes: [m, p, b, k, s, f]:

> Con escasos medios para proveerme de lo mejor en la parte material, mi mano nada versada en el manejo del formón no pudo hacer más que lo que ha hecho: poner de bulto los caracteres microscópicos de cada sonido oral, estudiando con la profundidad que es dable el enmarañado fenómeno de la fonación, y llegar a obtener, después de una larga serie de combinaciones, el nexo parcial de las cinco vocales simples que entran en la composición del castellano, unidas a las consonantes m, p, b, k, gue, s, f, con que se forman diferentes frases. (Pérez 1868: 19)

Carlos Nebreda, miembro de la Matritense y Director del Colegio Nacional de Sordomudos y Ciegos de Madrid, alude al Tecnefón en 1870 en los siguientes términos:

> Nuestro compatriota D. Severino Pérez, se ocupa actualmente en perfeccionar una máquina de este género inventada por él hace algunos años, y que ha dado a conocer con el nombre de *Tecnefon*, la cual hemos tenido el gusto de ver funcionar sorprendiéndonos agradablemente los resultados obtenidos por su autor. El *Tecnefon* pronuncia con muchísima claridad gran número de palabras y habla en sentido interrogativo, afirmativo y admirativo, lo que constituye el mayor elogio que pudiéramos hacer de los estudios del Sr. Pérez. (Nebreda, 1870: 19)

En el siguiente apartado se ofrece un estudio puntual de la máquina de Pérez y de la teoría acústica y fonética que el autor despliega a propósito de su diseño. En cualquier caso y por razones de espacio, esta descripción no es del todo completa, aunque sí sirve para dar una impresión general del alcance teórico del Tecnefón.

3. LA MÁQUINA: LA TEORÍA Y LA PRÁCTICA

3.1. Descripción del Tecnefón

El Tecnefón, parecido a un órgano o piano portátil, era una caja de madera que medía aproximadamente ciento treinta centímetros de alto y setenta centímetros de ancho. La máquina está formada, teóricamente, por tres elementos principales: (a) el denominado "nexo oral", (b) un sistema de fuelles, y (c) un teclado:

> Comunican con el nexo un depósito neumático en forma de fuelle destinado a establecer la corriente que agita la laringe, y un teclado en que están seccionadas las letras debajo del modulador o registro prosódico que maneja la mano izquierda, al mismo tiempo que las vocales, para dar a la frase los tonos de que es susceptible. (Pérez 1868: 20)

a) El nexo oral está formado por un *centro sonoro permanente* (el generador de sonido) y un conjunto de tubos de distinta configuración (resonadores) con los que se consigue un número variable de *timbres*. Este centro sonoro consiste en una cuerda elástica de caucho adherida a la membrana de un pequeño tambor.
b) El sistema de fuelles, que funciona mediante pedales, permite producir y sostener el sonido generado por el centro sonoro.
c) A través del teclado se da cuenta de los sonidos (cada tecla representa una letra), a los que se añade un registro prosódico (acento y entonación) gracias a un *modulador*, una palanca situada a la izquierda del teclado[5].

La mecánica del Tecnefón es sencilla: manejando los pedales, las teclas y el modulador pueden reproducirse los sonidos de la lengua española. El Tecnefón busca la unión de los sonidos a través de la pulsación de teclas (lo que su autor llama "síntesis conexiva"), junto con ciertos ajustes prosódicos que añaden *acento* a las palabras y *entonación* a las oraciones que va formando. Consiste, como podemos suponer, en dar la correspondiente voz a cada tecla de un teclado y añadir un elemento acentual a cada palabra (convirtiéndola en aguda, llana o esdrújula) y a cada oración (resultando en una oración enunciativa, exclamativa o interrogativa).

---

5 Pérez apunta: "De este modo se consiguen todos los sentidos de que es capaz la oración, y se puede transformar cualquier palabra en aguda, llana o esdrújula, comunicando al conjunto silábico el ritmo propio del elemento literal" (Pérez 1868: 20-21).

> Para poner en acción su mecanismo y pronunciar por ejemplo *Sofía me amaba*, basta herir sucesivamente las teclas correspondientes a estas letras, y oprimir el pedal durante el tiempo que se tarde en recorrerlas. Mas para que la frase no resulte monótona e incorrecta en la medida, es preciso dar a cada sonido oral el valor métrico y la nota prosódica que tenga en la dicción natural, inflexionando adecuadamente el modulador, y deteniéndose en los timbres solamente lo preciso para no hacerlos más breves ni más largos de lo que son. (Pérez 1868: 20)

Pese a la simplicidad de la idea, la teoría acústica que subyace al Tecnefón es compleja. La construcción de la máquina consiste, básicamente, en el diseño de artilugios que *suenen* como las vocales o como las consonantes del español, ya que la máquina no es exactamente una laringe artificial. La investigación de Pérez parecía centrarse más en la física acústica que en la fonética articulatoria, aunque parecía conocer bien el mecanismo de la fonación; de su formación en estas materias dan buena cuenta las autoridades citadas en su obra y sus intuiciones acerca de los sonidos: "No hay, pues, en el aparato más que efectos puramente físicos, debidos a la separación de lo fisiológico y lo mecánico tan íntimamente hermanados en el instrumento de que sirve la naturaleza." (Pérez 1868: 20).

La lista de autores que cita Pérez en la obra de 1868 es relativamente larga, y algunos de ellos han aparecido en la sección anterior. Por ejemplo, Pérez menciona numerosas veces, entre los físicos, a Helmholtz, aunque los libros de acústica que parece manejar son los de Chladni (París, 1809) y Rodolphe Radau (1835-1911) (París, 1867). Alude también a algunos estudios de fisiología, como los de Auguste Laugel -a través de Pedro Felipe Monlau (1808-1871), que parece que le proporcionó una laringe artificial-. También hay indicaciones acerca de Kempelen y Kratzenstein, así como un repaso por el estudio de la fonación y por el estudio de las membranas vibrantes. Aunque se trata de un informe realizado por la Sociedad Económica Matritense incluido en la obra de Pérez, también aparecen algunos de los más importantes autores dentro la *mítica* del automatismo: desde San Alberto Magno a Mical, pasando por Athanasius Kircher (1602-1680), Jacques de Vaucanson (1709-1782) o Antoine Rivarol (1753-1801). Sin embargo, es interesante traer la siguiente cita debido a que en ella podemos leer el alcance de las investigaciones de diversos autores en el trabajo de Pérez (la cursiva es mía):

> Los Sres. Laugel y Radau, el primero en su obra *La voix l'oreille et la musique* –París, 1867- y el segundo en la suya ya citada, hablan con grande elogio de los adelantos acústicos debidos a la superior inteligencia del Sr. Helmholtz. Sabido es desde hace mucho tiempo que "cada cuerpo que suena libremente es por sí solo una pequeña orquesta" formada por sonidos más o menos oscuros que se elevan sobre uno fundamental y dominante. Partiendo de este hecho, el Sr. Helmholtz ha ensayado la producción artificial de las vocales, construyendo una especie de piano con ocho diapasones, templados según la nota dicha B y sus siete primeras armónicas, y reforzados por

> otras tantas cajas sonoras que se pueden abrir y cerrar moviendo un teclado. Los diapasones se mantienen constantemente en vibración por medio de un sistema de electro-imanes que, 120 veces por segundo, se imantan y se desimantan. Con esto y un segundo aparato de cuatro armónicas, más agudas que las anteriores, ha conseguido una *u* "sorda, bastante más apagada que la que produce la voz natural", la *o*, la *a* y la *e*; faltando la *i*, porque su timbre particular es debido a una armónica sobreaguda que la corriente no hace vibrar con bastante fuerza.
> *Como se ve, estos experimentos llevan un rumbo diverso del mío; y aunque su tendencia no parece ser la producción total de los sonidos orales, posible es que acaso por esto mismo se alcance con ellos el objeto deseado. Sin embargo, yo no veo que con este piano de armónicas se explique más que una parte secundaria del timbre de las letras vocales.* La cavidad que forma la boca, no es sólo un reforzador de la voz, es más bien un instrumento músico completo, que recorre independientemente todas las notas de su gama, sin que el sonido laríngeo, sea el que fuere, desempeñe en él otro papel que el de tañedor o mero ejecutante. Por eso bajo cada una de las notas cardinales, por decirlo así, u o a e i del tímpano bucal, se deslizan indistintamente y sin grave menoscabo de su intensidad, los tonos que constituyen la extensión de la voz. Así es que los Sres. Radau y Laugel me permitirán que les diga, sin ánimo de prejuzgar ulteriores adelantos, que el entendido alemán no habrá hecho la verdadera síntesis de las vocales, mientras no las obtenga empleando una *sola* caja sonora, sensible indistintamente a la influencia de *cualquier* diapasón. (Pérez 1868: 34)

La generación de la voz es tarea relativamente sencilla, según Pérez, ya que la voz propiamente dicha no es esencial en la locución. Lo que Pérez quiere decir en este punto es que la voz puede sustituirse por cualquier sonido producido por un instrumento que se adapte a las circunstancias de la laringe. Pérez estima que la palabra "cuchicheada" o "la voz del que padece afonía"[6], así como "la brillante voz de un orador o la que brota de una bocina" son, esencialmente, las mismas, "sin otra variación que el grado de intensidad y la distancia[7]" (Pérez 1868: 10). Y apunta, un poco más adelante, que "el sonido emanado

6 Ante esta afirmación, parece inevitable preguntarse si Pérez conocía o intuía los *modos de fonación*. Los modos de fonación hacen referencia al estado de la glotis cuando se produce la voz. Hay diferentes modos: normal (las cuerdas vocales vibran, por lo que se producen sonidos *sonoros*; cuando la glotis está abierta pero las cuerdas no vibran, entonces se producen sonidos *sordos*); susurrado (no hay vibración y el aire pasa sólo a través de un pequeño hueco en la parte posterior de las cuerdas); o ronco (una vibración más lenta al menos en la parte anterior de las cuerdas vocales), entre otros. En español los modos de fonación -excepto, claro está, el normal- no tienen un valor estrictamente fonológico, aunque sí paralingüístico. Parece que Pérez reconocía que existían distintos tipos de voz (no digo que conociera los modos de fonación), pero deja sentado que es irrelevante en lo que al habla se refiere. Este aspecto no sería tan llamativo si se redujera a este punto, pero es que la distinción entre sonidos sordos y sonoros es, en español, indispensable.

7 *Audibilidad*, probablemente

de la laringe, cualesquiera que sean su fuerza, su metal y entonación, es siempre compatible con los demás componentes de la palabra" (Pérez 1868: 10).

El inventor parece redundar en la idea de que no es intrínsecamente necesaria la existencia de *voz* como tal [la fuente], ya que ésta puede ser sustituida por un sonido que, en un momento dado, pueda ser moldeado y dé lugar a los distintos *timbres* de una lengua. A esto parece referirse Pérez cuando señala que este sonido debe adaptarse a las circunstancias de la laringe y que son dichos timbres los verdaderos artífices de la palabra:

> Los elementos acústicos que, en unión del *laríngeo*, constituyen la palabra son indefectiblemente otros tantos timbres que se suceden con más o menos rapidez alrededor de un centro sonoro permanente; pues no puede haber otra cosa que aditamentos de timbres diferentes en un sonido que, sin dejar de ser el mismo, se diversifica a cada momento de su duración. (Pérez 1868: 10)

Por tanto, este parece ser el punto de partida en las reflexiones e investigaciones de Pérez: es necesaria la existencia (mínima) de un centro sonoro que genere el sonido y de un conjunto de *timbres* que efectivamente moldeen tal sonido. El segundo paso, entonces, tiene que ver con cuáles son estos timbres y cuál puede ser su resorte físico.

3.2. Los sonidos del español y su implementación: los timbres

Según Pérez, la primera dificultad del Tecnefón fue diseñar los timbres. Este autor iguala *timbre* a lo que conocemos como filtro y/o resonador, ya que alude con este término al efecto físico de amplificación o atenuación de frecuencias producido en una cámara determinada. El diseño de los timbres, por tanto, consiste en crear o conseguir tubos o cámaras que efectivamente refuercen o no determinadas frecuencias del sonido, de forma que tales sonidos se perciban como los de una lengua. Pérez habla, por tanto, de que se pueden conseguir hasta treinta timbres, y de que la lengua castellana emplea veintitrés.

Los timbres son "las cualidades literales que puede adquirir la emisión aérea en su paso por el órgano oral" (Pérez 1865: 5), y tales cualidades son las que Pérez intenta reconstruir. En un primer paso, reduce a cinco grupos los timbres de la lengua castellana: *timbre timpánico, timbre explosivo, timbre bitimpánico, timbre semiexplosivo* y *timbre sibilante*. En cada grupo incluirá determinados sonidos. Pero antes de entrar en las particularidades de cada timbre, es interesante recordar que Pérez añade, a la importancia del conjunto de timbres, dos matices igual de relevantes para la creación del sonido: lo que denomina "consonancia de dos timbres", y que se acerca a la *coarticulación*; y la "modulación", que permitirá dotar de acento y entonación a lo pronunciado.

3.2.1. Timbre timpánico

Pocas son las características que ofrece Pérez de este primer y fundamental timbre. Pérez denomina *timbre timpánico* "al efecto acústico que produce la cavidad", y que "equivale a producir en un timbal las notas necesarias para la ejecución de una pieza." (Pérez 1865: 5). Su escala consta de tres octavas, de tal forma que "haciéndola subir al mayor

grado de amplitud, se hallan las articulaciones vocales de todas las lenguas, inclusas las compuestas o que resultan de una consonancia como la u francesa" (Pérez 1868: 5). En este sentido, afirma que la u es un *do grave*; o es una octava más alta que *u*; si o sube tres tonos resulta a[8]; si o sube siete se obtiene e; y si e se aumenta una cuarta, resulta i. Por tanto, *u* es un *do grave*; *o* es un *do agudo*; *a* es un *fa agudo*; *e* es un *fa sobreagudo*; e *i* es un *do agudísimo*. Tratando esta diferencia desde un punto de vista acústico, estamos ante una diferenciación que puede basarse en el segundo formante de las vocales que, como sabemos, correlaciona con el punto de articulación.

Con respecto a la descripción de las vocales, es interesante dar cuenta de un comentario de Pérez a propósito de su *rango de variación*. Pérez apunta que las vocales sólo pueden realizarse dentro de ciertos límites que, de forzarse, podrían resultar en la anulación de dicha vocal. Parece que nos encontramos ante la intuición de que existen límites en la variación (vocálica); o, dicho de otra forma, que la variación vocálica no puede ser ilimitada.

3.2.2. Timbre explosivo

La descripción del timbre explosivo resulta algo más completa, quizá debido a su aparente sencillez. Según el autor, este segundo timbre "parece tener alguna semejanza con los sonidos por percusión" y, de este modo, consiste en "producir tres gamas en un tubo cilíndrico por medio de un émbolo que funcione como el corcho de una botella lanzado con fuerza" (Pérez 1865: 5-6). Según Pérez, este orden de sonidos, "caracterizado por el estampido de la emisión aérea", ofrece un "diapasón igual al de las articulaciones vocales". Seis son los sonidos que Pérez caracteriza a través de este *timbre explosivo*:

> p es un golpe explosivo, unísono con el tímpano u
> t es otro golpe explosivo, unísono con o
> k, o llámese q, es una explosión afinada con el tímpano a
> erre = 6t, esto es, seis golpes de t ejecutados rápidamente, producen el movimiento de trepidación característico de la erre
> l consiste en una explosión unísona y semiconsonante con el tímpano e
> ll es también una explosión, pero unísona y semiconsonante con el tímpano i (Pérez 1865: 6)

3.2.3. Timbre bitimpánico

El tercer conjunto de timbres agrupa a los sonidos nasales del español. Según Pérez, "consiste en un tímpano de entonación u, que puede consonar perceptiblemente con los timbres anteriores, menos con el erre" pero "disminuyendo su intensidad y tinturándolos del carácter que los gramáticos llaman nasal". Estos tres sonidos son "m, n y ñ [...] de entonación p, l, ll, respectivamente" (Pérez 1865: 6).

---

8 Para Pérez, el tímpano a "es la base o punto de partida natural, no el científico" (Pérez 1868:11).

3.2.4. Timbre semiexplosivo

El timbre semiexplosivo, como su mismo nombre indica, no causa más que "un remedo de explosión", y consiste en "aproximar los pistones de p, t, k, l a los tímpanos u, o, a, e" para constituir las articulaciones "b, d, gue, ere" (Pérez 1865: 6).

3.2.5. Timbre sibilante

Según Pérez, este timbre tiene "muchísima [relación] con el instrumento que forman los labios al silbar", y lo forman los sonidos *f, g, c, s,* que "son silbatos de entonación ascendente, partiendo de f", y que tienen la rara particularidad "de sonar lo mismo con corriente hacia dentro que hacia fuera". Pérez añade el timbre *ch*, "una semiconsonante con un silbato agudo", "por lo mismo letra compuesta", que se coloca bajo esta denominación "por ser su carácter dominante".

Tras la descripción de los timbres -que, al fin y al cabo, son también los elementos teóricos de que consta el Tecnefón-, Pérez entra en lo que denomina *síntesis conexiva*: esto es, cómo ha de hacerse sonar la máquina y cómo hacerla funcionar. Para ello introduce el concepto de *consonancia*:

> La consonancia de los timbres es de dos maneras: simultánea, si los componentes llenan la unidad entera del tiempo literal, como en los bitímpanos; sucesiva, si a cada uno pertenece una fracción, como sucede en *l, ll, ch*. Y es de notar que en los timbres de consonancia fraccionaria un mismo componente ocupa unas veces la primera mitad del tiempo, y otras la segunda, aunque con diversa afinación. Estos y otros pormenores de sutileza acústica no sólo se pueden presentar a la vista por medio del grabado, sino que también al tacto empleando articuladores descomponibles que hagan de todo ello una verdad rigurosamente *palmaria*. (Pérez 1865: 6)

A todo esto hay que añadir el efecto del modulador, la palanca que, situada a un lado del teclado, permitía dotar de acento a las palabras y de entonación a las frases:

> La modulación, que es el accidente oral que convierte la articulación en canto y presta a la palabra la hermosa facultad de expresar los más delicados matices del sentimiento, se produce por un doble estrangul compuesto de dos placas elásticas que se chocan recíprocamente al vibrar, y se entonan de diversos modos según la tensión que sufren [...] Y esto resuelve de una manera palpable la cuestión gramatical tan debatida, y hasta ahora *sub judice*, de si la cantidad y el acento son una misma cosa en castellano, o si son accidentes separables. Por medio del tecnefón no sólo se demuestra que la cantidad breve es compatible con el acento agudo, sino que existen sílabas acentuadas con el circunflejo, por más que en nuestra usual escritura no haya signo alguno que lo indique; este, sin embargo, es más frecuente en el estilo familiar que en el académico; más marcado en

la entonación de ciertas provincias que la usada en Madrid." (Pérez 1868: 6 y 20-21)

Para terminar, enumeramos algunas observaciones que apunta Pérez con relación a la teoría del Tecnefón, y que enlazan con algunas de las cuestiones tratadas en este trabajo:

> ... la *máquina de hablar* tiene que resolver la difícil cuestión de hacer pasar a voluntad por diversos timbres un sonido dado; o lo que es lo mismo para que se entienda mejor; construir un tecnefón para la lengua castellana, es parecido a formar una trompa de veintitrés tonos pertenecientes a otras tantas especies de instrumentos, sin que por esto pierda su carácter el sonido que le es propio.
> Para conseguirlo, hay que tener muy a la vista las siguientes observaciones, como otras tantas reglas de *estereotomía oral*:
> 1ª. El *nexo* o unión más sencilla es la de dos timbres de igual denominación; si ésta varía, o es otro el número de los que se enlazan, el nexo se complica y se hace cada vez más difícil.
> 2ª. Atendiendo al grado de conexión que entre ellos existe, los cinco grupos de timbres aparecen colocados por este orden: 1º. tímpanos; 2º. explosiones; 3º. bitímpanos; 4º. semiexplosiones; y 5º. silbatos.
> 3ª. Los timbres timpánicos, que constituyen la cavidad o fondo complectivo sobre que se destacan los demás accidentes de la locución, son los menos fáciles de *conexionar* entre sí por su inmediata relación con la voz, cuya entonación se desfigura a poco que aquellos excedan los límites convenientes a su escala [...].
> 4ª. La lengua [...] no puede menos de ser reemplazada con llaves distributivas de la emisión aérea, que sirvan al mismo tiempo de medios mecánicos para tañer los timbres, facilitando en una dirección conveniente la propagación del sonido que procede de la laringe.
> 5ª. Los nexos deben ser binarios, ternarios, cuaternarios, etcétera, primero dentro de un orden de timbres, y luego entre órdenes *afines*, procediendo de esta manera hasta obtener el nexo total.
> 6ª. En el juego de los articuladores, al cambiar de timbre, hay tres partes que es forzoso saber apreciar para la exactitud de la metrificación literal: 1ª. *movimiento de ascenso*, en el cual se expone el timbre a la influencia de la corriente sonora; 2ª. *movimiento estático*, en que suena el timbre claramente durante un tiempo proporcional al *aire* o rapidez con que se pronuncia; y 3ª. *movimiento de descenso*, en el cual se desvanece. Los movimientos envuelven los timbres en la oscuridad causada por el encuentro del que cesa con el que empieza, al paso que el momento estático les da pureza y brillantez.

> 7ª. Las inflexiones de la voz y los timbres con que ella consuenan, deben sujetarse a la acción de un teclado que manejen fácilmente las dos manos, o en defecto un cilindro dentado, y en el cual aparezcan aquellos distribuidos en cinco grupos o categorías, colocadas según el grado de conexión.
> 8ª. Los timbres a, e, s, son los más usados en castellano, y ll, ñ, ch, los menos. Y
> 9ª. Los tímpanos o vocales componen un tercio de nuestra literatura, resultando la lengua, por esta razón, sonora, clara, rotunda y majestuosa, en un grado tal vez que ninguna otra. (Pérez 1868:14)

4. FINAL

En estas líneas hemos llevado a cabo, en la medida del espacio disponible, una primera caracterización de la máquina parlante denominada *Tecnefón*. Su autor fue el español Severino Pérez y Vázquez, que desarrolló el prototipo de esta máquina hacia 1865. En este estudio se ha revisado el contexto tecnológico en el que se sitúa el Tecnefón y se ha dado una primera noticia del funcionamiento de la máquina. Asimismo, este trabajo ofrece una descripción de los *timbres* de la lengua española tal y como fueron pensados y diseñados por Pérez. Es necesario apuntar que a través del estudio de tales timbres se puede dar cuenta no sólo de cuáles son los que propone el autor para el español, sino además de cómo los implementa; esto es, qué y cómo hace para conseguirlos. Con estos datos, también se puede analizar la relación que establece entre la acústica de los sonidos y su percepción, y si tal cosa pudiera correlacionar con algún rasgo articulatorio -dado que existe cierta confusión en los estudios de Pérez entre una perspectiva fisiológica o fonética y una acústica o perceptiva-. En cualquier caso, esta circunstancia suele ser el denominador común de las teorías que existen bajo las máquinas parlantes o los más modernos sintetizadores de voz.

BIBLIOGRAFÍA

ASHER, R.E.: *Encyclopaedia of language and linguistics*. Oxford, Pergamon Press, 1994.

BATTANER MORO, E.: "Un caso de heterodoxia lingüística en la España del XIX: el Tecnefón, una máquina parlante". *Actas del V Congreso de Historiografía Lingüística*, Universidad de Murcia, 7-11 de noviembre de 2005 (en prensa).

BROECKE, M.P.R. van den: "Wolfgang von Kempelen Speaking Machine as a Performer" en M.P.R. VAN DEN BROECKE - V. VAN HEUVEN - W. ZONNEVELD (Eds.) *Sound Structure. Studies for Antonie Cohen*. Dordrecht, Foris Publications, 1983, pp. 9-20.

CHLADNI, E. *Tratado de acústica*. París, 1809.

HELMHOLTZ, H.L.: *On the sensations of Tone as a Physiological Basis for the theory of Music* [= *Die Lehre von den Tonempfindungen als physiolosgische Grundlage für die Theorie der*

Musik]. Traducción y edición de A. J. Ellis. Reproducción facsimilar en *Classics in Psychology*. Bristol, Thoemmes Press, 1862[1875].

KEMPELEN, W. Von: *Mechanismus der menschlichen Sprache nebst Beschreibung Seiner Sprechenden Maschine*. J.V. Degen, Viena, 1791.

KRATZENSTEIN, C.G.: "Sur la naissance de la formation des voyelles", *Journal de Physique*, 21, 1782, pp. 358-380.

LAUGEL, A: *La voix, l'oreille et la musique*. Paris: G. Baillière, 1867

LIÉNARD, J.S.: "From speaking machines to speech synthesis". *Actes du XIIème Congrès International des Sciences Phonétiques*. Aix-en-Provence, Université de Provence, Service des Publications, 1991, vol. 1 pp. 18-27.

NEBREDA Y LÓPEZ, C.: *Tratado teórico-práctico para la enseñanza de la pronunciación de los sordo-mudos*. Madrid, Imprenta del Colegio de Sordo-Mudos y de Ciegos, 1870.

PÉREZ Y VÁZQUEZ, S.: *El Tecnefón. Memoria*. Madrid, Imprenta de El Cascabel, 1865.

PÉREZ Y VÁZQUEZ, S.: *El Tecnefón y los medios representativos de la palabra*. Madrid, Imprenta del Colegio de Sordo-Mudos y de Ciegos, 1868.

RADAU, R.: *L'Acoustique, ou les phénoménes du son*. París, 1867.

SCHROEDER, M.R.: "A brief history of synthetic speech", *Speech Communication* 13, 1993, 1-2: 231-238.

VILE: ESTUDIO ACÚSTICO DE LA VARIACIÓN INTER E INTRA LOCUTOR EN ESPAÑOL

ELENA BATTANER
Departamento de Comunicación II. Universidad Rey Juan Carlos
CARME CARBÓ, JOAQUIM LLISTERRI, MARÍA JESÚS MACHUCA, NATALIA MADRIGAL,
CARME DE LA MOTA, MONTSERRAT RIERA, ANTONIO RÍOS
Departamento de Filología Española. Universidad Autónoma de Barcelona
JUANA GIL, VICTORIA MARRERO
Departamento de Lengua Española y Lingüística General. Universidad Nacional de Educación a Distancia

1. INTRODUCCIÓN

VILE (*Estudio acústico de la variación inter e intralocutor en español*) es un proyecto de investigación básica, financiado por el Ministerio de Ciencia y Tecnología (BFF2001-2551, 2001-2004)[1], centrado en la caracterización acústica de los elementos segmentales y suprasegmentales que contribuyen a establecer la individualidad de un hablante frente a aquellos que son comunes a un estilo de habla, una variedad geográfica o social, o una lengua. Se pretende con ello obtener el conocimiento fonético necesario para la mejora de los sistemas de identificación o verificación automáticas del locutor, así como dotar a los especialistas en fonética forense de nuevos datos acústicos que permitan comparar, con un mayor grado de certeza, locutores dubitados e indubitados.

En las primeras etapas del proyecto se seleccionaron los parámetros acústicos que se toman en consideración (apartado 2) y el corpus a partir del cual se obtienen los datos (apartado 3). A continuación, se ha llevado a cabo un análisis acústico (apartado 4), del cual se presentan los primeros resultados (apartado 5).

1 Para más información puede consultarse http://liceu.uab.es/~joaquim/VILE.html

2. SELECCIÓN DE LOS PARÁMETROS ACÚSTICOS ANALIZADOS

La revisión de la bibliografía acerca de los parámetros relevantes para el estudio de la individualidad de la voz se centró en tres formas de aproximación a los datos fonéticos: el reconocimiento visual de espectrogramas, el reconocimiento perceptivo y el reconocimiento automático del habla. Debido a la subjetividad implícita en el análisis de los datos que se derivan de la primera aproximación, la búsqueda se limitó a las otras dos aproximaciones que, en cualquier caso, ofrecen un volumen de publicaciones mayor y más en consonancia con las tecnologías actuales. De esta forma, nos centramos en aquella literatura que se refiere a los parámetros denominados *de bajo nivel de información* (Doddington 1985) o los que se corresponden con las dimensiones *fisiológicas* (Kuwabara y Sagisaki 1995). Stevens (1971) destaca dos parámetros generales susceptibles de emplearse en la identificación y discriminación entre hablantes: aquellos referidos a la fuente (frecuencia fundamental, en adelante F_0, y forma de la onda glotal) y a los resonadores (frecuencias formánticas, anchura de formantes, sonidos turbulentos y consonantes nasales). Hollien (1990 y 1991) añade otros rasgos que parecen decisivos en relación con dicha identificación y discriminación: el espectro a largo plazo, el vector de los formantes vocálicos, el vector temporal y el vector de F_0. Otros trabajos como Atal (1972), Karlsson (1988), Eskenazi *et al.* (1990), Kuwabara y Takagi (1991) y Kuwabara y Sagisaka (1995) mencionan aproximadamente los mismos parámetros como responsables de la individualidad de la voz.

A la hora de seleccionar los parámetros para el estudio, pueden tenerse en cuenta algunas de las reflexiones que hace Wolf (1972) acerca de cuáles serían los criterios de decisión válidos:

a) Deberían ser parámetros presentes *natural* y frecuentemente en el habla normal. Los efectos coarticulatorios, en cuanto que son, en cierto grado al menos, 'aprendidos', no se tendrían en cuenta (*cf.* para una opinión contraria, Su *et al.* (1974)).
b) Han de ser fáciles de medir.
c) Deben tener la mayor variabilidad posible inter-hablantes y la menor posible intra-hablante. Tanto la forma de la onda glotal como los sonidos turbulentos parecen reunir esas condiciones.
d) No deberían variar mucho con el tiempo ni verse afectados por las condiciones psico- o fisiológicas del hablante. Las nasales quedarían eliminadas de acuerdo con este último criterio (*cf.* sin embargo Wolf (1972) y Glenn y Kleiner (1968), quienes consideran muy informativo el análisis de las nasales). En el habla espontánea, un rasgo especialmente sensible al estrés experimentado por el hablante es la F_0, y uno muy resistente es el espectro medio (LTAS, *Long Term Averaged Spectrum*) (Pitram 1987, Hollien 1990).
e) Han de ser resistentes al posible ruido ambiental y no han de verse afectados por las condiciones de la transmisión. Los valores medios de los formantes vocálicos (del F3 –que proporciona indicios sobre la longitud del tracto vocal del

hablante-, del F2 y del F1) y su anchura son parámetros muy válidos en este sentido, puesto que son muy resistentes a la distorsión y a las interferencias. Se aconseja estudiar las vocales extremas [a i u]. Por otra parte, se señala que los valores del F1 y F2 de las vocales extremas /i,u,a/ son los más estables y menos sensibles al contexto (Stevens y House 1963).

f) Finalmente, no pueden ser fácilmente modificables por la mera voluntad del hablante, es decir, deben ser resistentes a los intentos de disimular la voz. La distancia entre los tres primeros formantes no puede ser alterada a voluntad (F1 / F2, F2 / F3).

Existe un importante desacuerdo en lo que se refiere a la importancia relativa de los distintos parámetros y, por ende, a cuál o cuáles deben ser los parámetros prioritarios:

a) Prioridad del F_0: Compton (1963), Wolf (1972), Matsumoto *et al.* (1973), Brown (1981), van Dommelen (1987).
b) Prioridad del LTAS: Bordone-Sacerdote y Sacerdote (1969), Doherty (1976), Hollien y Majewski (1977), Furui (1986), Pittam (1987).
c) Prioridad de la estructura espectral, bien sea de las frecuencias formánticas absolutas: Shearme y Holmes (1959), Miller (1964), Itoh y Saito (1982), Carrell (1984) Kuwabara y Ohgushi (1987), Kuwabara y Takagi (1991), o bien sea de las trayectorias formánticas: Ingram *et al.* (1996).
d) Misma prioridad para la F_0 y la estructura formántica: La Riviere (1975).
e) Variables temporales: Pruzansky (1963), Wolf (1972), Doherty y Hollien (1978), Brown (1981), Johnson *et al.* (1984).
g) No es factible establecer una prioridad, la importancia de cada parámetro puede diferir de hablante a hablante y depende también de la naturaleza de las muestras: Gobl (1989) y Kuwabara y Sagisaka (1995).
h) Contorno tonal: Atal (1972), van Dommelen (1987).

Por lo que se refiere a qué formantes o qué distancias entre formantes son las más informativas, tampoco hay coincidencia en las posturas, pero la mayoritaria defiende el estudio de los tres primeros formantes. En este sentido, la falta de resultados coincidentes y concluyentes acerca de cuál es el parámetro más decisivo para el reconocimiento del hablante puede deberse a los enfoques metodológicos empleados, que son muy diferentes. En cualquier caso, parece difícil establecer una jerarquía absoluta entre los parámetros. Varios autores apuntan la interdependencia entre los índices, cuya prioridad relativa dependería, asimismo, del hablante. Por ejemplo, los oyentes pueden tomar como clave primaria para el reconocimiento de un hablante A un F_0 bajo, y sin embargo apoyarse en la estructura formántica para el reconocimiento de un hablante B (van Dommelen 1987). Esto es, todos los rasgos mencionados conllevan un cierto grado de información sobre las características del hablante y son potencialmente válidos para la tarea de reconocimiento.

3. SELECCIÓN DEL CORPUS ANALIZADO

El segundo paso fue la revisión de los recursos orales existentes en español, considerando la conveniencia de reutilizar en la medida de lo posible unas bases de datos de gran calidad, obtenidas gracias al trabajo de muchos especialistas, y con frecuencia gracias a financiación pública.

En primer lugar descartamos los que no estaban disponibles para fines investigadores, por ser de propiedad privada[2]. Tampoco se adecuaban a nuestros intereses los que fueron creados con fines específicos, como la variación dialectal o sociolectal[3], el análisis del discurso o la conversación[4], etc. Otros (como ROARS, MATE, el corpus oral de la UAM, o el del propio CREA) no nos resultaban accesibles en su totalidad.

Los corpus analizados fueron los siguientes: AHUMADA- GAUDI, Albayzín, SpeechDat, EUROM1 y su subconjunto MULTEXT. A continuación resumimos las posibilidades que ofrecen para el estudio de la variabilidad:

Según el tipo de variación

Variación intralocutor

El mayor número de locutores repitiéndose a sí mismos los proporciona GAUDI: 455 personas leyeron tres veces distintas idéntico texto y diez frases fónicamente equilibrados. En ALBAYZIN encontramos 20 locutores que reproducen dos veces (una de ellas bajo el efecto Lombard) un conjunto de 250 frases equilibradas, mientras que en EUROM1 es una serie de 82 logatomos la que es leída cinco veces por doce emisores.

Variación interlocutor

La repetición de los mismos estímulos por diferentes sujetos es el fin de todas las bases de datos: SpeechDat, por ejemplo, nos ofrece 999 frases fonéticamente ricas leídas por 1002 personas a través del teléfono. Los 455 locutores mencionados de GAUDI también proporcionan una cantidad considerable de estímulos para analizar la variación interlocutor. Por orden decreciente, ALBAYZIN permite comparar las voces de 154 lectores de entre 200 y 500 frases equilibradas. Incluso los 60 locutores de EUROM1 proporcionan un material oral muy abundante.

---

2 CEUDEX, SPATIS, VESTEL, VOCATEL, VOGATEL, TANGORA, IBM Voice Type, etc.

3 The Sala Project – SpeechDat Across Latin America; Macrocorpus de la norma lingüística culta de las principales ciudades del mundo hispánico. Caracas-77: Estudio sociolingüístico de Caracas, 1977. Caracas-87: Estudio sociolingüístico de Caracas, 1987. CEAP: Corpus de Encuestas en Asunción de Paraguay.CSMV: Corpus Sociolingüístico de Mérida-Venezuela. CSC: Corpus para el estudio del español hablado en Santiago de Compostela. COVJA: Corpus oral de la variedad juvenil universitaria del español hablado en Alicante. ALMECOR: Universidad de Granada. PRESEEA Corpus Sociolingüístico del español. Fonética Acústica Experimental del Español de Canarias- U. La Laguna. ILSE – Universidad de Almería. Variedades Urbanas Andaluzas – U. Granada y U. Málaga.

4 Análisis del Discurso Público Actual – Universidad de La Coruña. Análisis del Discurso Oral – Universidad de Granada. ACUAH – Universidad de Alcalá de Henares. Corpus de conversación coloquial- Universidad de Valencia.

Según el nivel de análisis

En el nivel segmental, los datos disponibles permiten analizar todos los sonidos del español en distintas condiciones de naturalidad, desde la lectura de logatomos (con o sin frase portadora), que sería el extremo más artificial, pero también el de mayor control de variables (EUROM1), hasta la lectura de frases equilibradas o fonéticamente ricas (en todas las bases de datos), la lectura de párrafos o textos (EUROM1 y GAUDÍ), o el habla espontánea (sólo en GAUDÍ).

Sin embargo, el estudio del nivel suprasegmental se encontraría limitado a los 20 párrafos con modalidades oracionales interrogativas y exclamativas de EUROM1.

El ritmo de habla, por último, puede ser analizado comparando las lecturas del texto en GAUDÍ, a tres velocidades distintas.

3.1 El Corpus Ahumada-Gaudí

Finalmente, nos decantamos por AHUMADA-GAUDÍ, cuya descripción se encuentra en Ortega *et al.* 2000. Por una parte, porque nos ofrecía abundante material para el estudio de la variación intralocutor, en condiciones de grabación muy controladas, tanto en el aspecto técnico como en el intervalo temporal entre unas sesiones y otras[5]. Por otra, por la existencia de unos minutos de habla espontánea para cada uno de los sujetos. Y por último, porque el material proporcionado también parecía sobradamente abundante para analizar la variación entre hablantes[6].

Dejando a un lado las tareas con dígitos, a todos los locutores se les pidió:

a) Lectura de 10 frases fónicamente equilibradas: 409 fonemas
b) Lectura de un texto de 179 palabras, también equilibrado (712 fonemas), a tres velocidades distintas (normal, rápida y lenta).
c) Lectura de un texto específico para cada locutor.
d) Habla espontánea durante más de un minuto, describiendo cuadros y dibujos que tenían preparados, o sus vacaciones, el lugar donde nacieron, etc.

Todas esas tareas se realizaron en seis sesiones de grabación: tres *in situ* con micrófono y tres por teléfono. Las sesiones microfónicas fueron la 1, la 3 y la 5, separadas entre sí por intervalos de entre 20 y 40 días.

4. ANÁLISIS ACÚSTICO

Para el estudio de la variación interlocutor e intralocutor del proyecto VILE se han seleccionado las grabaciones microfónicas del corpus Gaudí-Ahumada correspondientes a

---

5 Limitándonos a las grabaciones con micrófono, las 10 frases leídas tres veces por cada locutor suman 13.650 frases. El texto fue leído nueve veces, tres a velocidad normal, tres a mayor velocidad y tres más lentamente, en total, 4.095 repeticiones.

6 Si se suman las tres repeticiones por sujeto, cada frase ha sido leída 1365 veces, 432 por locutores masculinos y 393 por locutores femeninos. El texto, considerando sólo una velocidad de elocución, ha sido repetido las mismas veces.

Figura 1. Segmentación y extracción automática de valores de F_0 en el grupo fónico *pero yo me creía muy mayor* procedente del corpus de lectura.

muestras de habla de dos tipos: una, en tarea de lectura y contextos controlados (texto de 179 palabras, a velocidad normal); la otra, a tres sesiones de habla espontánea.

El análisis acústico del corpus, a partir de una selección de las grabaciones de 30 locutores masculinos, se realizó mediante el programa *Praat* (http://www.praat.org), desarrollado por Paul Boersma y David Weenink en el Instituto de Fonética de la Universidad de Ámsterdam.

Se eliminaron aquellos locutores cuya grabación de habla espontánea se obtenía mediante la descripción de una lámina, ya que en estos casos el locutor se preparaba con antelación la descripción y, por tanto, su elocución no presenta las características propias del habla espontánea no planificada.

Los parámetros acústicos considerados en el proyecto pueden clasificarse dependiendo del tipo de fuente, de la acción de los resonadores y de los parámetros temporales, tal como se describe a continuación.

4.1 Parámetros referidos a la fuente

El corpus de lectura o de habla espontánea fue segmentado en grupos fónicos (el segmento de habla comprendido entre dos silencios), y de cada uno de ellos se extrajo el valor medio de la F_0 y su fluctuación (la desviación estándar). En la figura 1 puede observarse un ejemplo de segmentación de un grupo fónico y el resultado de la extrac-

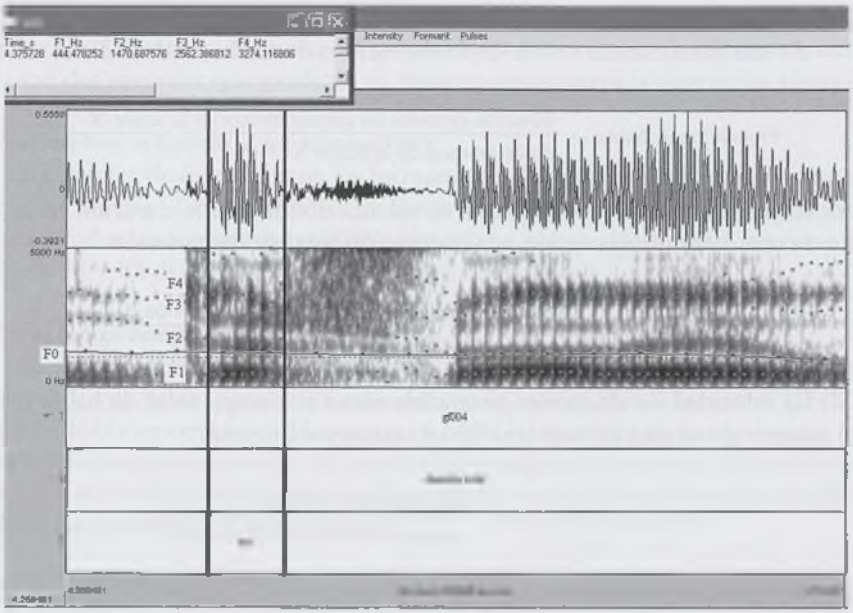


Figura 2. Segmentación y extracción automática de los valores de los cuatro primeros formantes de la vocal [o] del grupo fónico *ese viento suave vuela* procedente del corpus de lectura.

ción automática del valor medio y de la desviación estándar de la F_0 (ventana superior izquierda).

4.2 Parámetros referidos a los resonadores

En el corpus de lectura se segmentaron todas las vocales —tónicas y átonas— precedidas de cualquier oclusiva sorda (/p, t, k/) o de fricativa alveolar sorda (/s/) y, al mismo tiempo, seguidas de cualquier oclusiva sorda o de fricativa alveolar sorda, que apareciera en una estructura silábica del tipo CV $ C. Así mismo, se segmentaron todas las fricativas alveolares sordas precedidas y seguidas de las vocales más frecuentes del español, es decir de /a/ y /e/.

En los segmentos vocálicos se analizaron el valor de la F_0, las frecuencias centrales de los cuatro primeros formantes y sus anchos de banda (véase la figura 2). A partir de estos valores, se extrajo la distancia entre los valores de frecuencia de F_0, F1, F2, F3 y F4. Además, en los segmentos fricativos alveolares sordos se tuvo en cuenta el pico de mayor intensidad situado en la banda de frecuencia entre 0 y 8 kHz.

4.3 Parámetros temporales

En lo que respecta a los parámetros temporales, se consideraron:

a) El tiempo total de habla, es decir, la duración total de cada una de las sesiones analizadas.

b) La proporción de habla, definida como la duración de todos los grupos fónicos de una determinada sesión dividida entre la duración total de dicha sesión.

$$\text{Proporción de habla} = \frac{\text{duración de todos los grupos fónicos de la sesión X}}{\text{duración total de la sesión X}}$$

c) La proporción de silencios, que se calcula dividiendo la duración total de los silencios de una determinada sesión por la duración total de dicha sesión.

$$\text{Proporción de silencios} = \frac{\text{duración total de silencios de la sesión X}}{\text{duración total de la sesión X}}$$

d) La velocidad de elocución, entendida como el tiempo total de habla dividido entre el número de sílabas de cada una de las sesiones analizadas.

$$\text{Velocidad de elocución} = \frac{\text{tiempo total de habla}}{\text{número de sílabas de la sesión X}}$$

e) La velocidad de articulación, que es el resultado de la división entre el tiempo total de habla, exceptuando la duración de los silencios, y el número de sílabas de cada una de las sesiones analizadas.

$$\text{Velocidad de articulación} = \frac{\text{tiempo total de habla} - \text{silencios}}{\text{número de sílabas de la sesión X}}$$

El corpus de lectura se utilizó para el análisis de todos los parámetros descritos; para el estudio de los parámetros temporales y de los parámetros relacionados con la fuente se empleó también el corpus de habla espontánea.

5. RESULTADOS

Los resultados obtenidos mostraron variaciones significativas en los tres tipos de parámetros analizados, tal como se expone a continuación

5.1 Parámetros referidos a la fuente

Los valores medios de la F_0 presentan variaciones muy significativas en función del estilo de habla. Sin embargo, la tendencia no es la misma para los dos estilos observados: en el habla espontánea (excepto para siete locutores, 23,33% del total) se producen variaciones cuadráticas en el valor medio de la F_0 a medida que aumenta el número de sesiones; en lectura, en cambio, no se producen estas variaciones. En 8 locutores (26,66% del total) no se dan diferencias significativas en el valor medio de la frecuencia de F_0 en función del estilo de habla.

La desviación típica de la F_0 no presenta variaciones significativas debidas al estilo de habla. No obstante, la duración del grupo fónico y el número de segmentos del grupo fónico son dos variables que predicen de manera significativa la desviación típica de F_0.

5.2 Parámetros referidos a los resonadores

Los valores de frecuencia de los formantes de las vocales analizadas ayudan a identificar a los sujetos. Las resonancias en zonas de alta frecuencia y la F_0 permiten diferenciar en todos los segmentos vocálicos un mayor número de sujetos, mientras que los valores de F1 y F2 distinguen un número más reducido.

5.3 Parámetros temporales

La duración del grupo fónico también presenta diferencias significativas en función del estilo de habla, pero no en relación con la sesión. Sólo 5 locutores (16,66% del total) en el habla espontánea y 2 locutores (6,66% del total) muestran en lectura diferencias significativas en la duración del grupo fónico respecto a la sesión. En la velocidad de elocución y en la velocidad de articulación tampoco se producen diferencias significativas debidas a las distintas sesiones analizadas.

6. CONCLUSIONES

Los primeros resultados del análisis acústico hacen pensar que existen elementos que permiten la caracterización de los dos tipos de variación abordados en el proyecto. Por una parte, ponen de manifiesto que el estilo de habla es una variable que debe tomarse en consideración a la hora de comparar muestras de habla –en un contexto judicial, por ejemplo- y, por otra, que ciertos parámetros temporales y frecuenciales parecen mantener una cierta constancia en el mismo locutor a lo largo de las sesiones de grabación.

Además de refinar el tratamiento estadístico de los datos, en la segunda fase de VILE (*Estudio perceptivo de la variación inter e intralocutor en español*, HUM2005-06980/FILO) se plantea la validación perceptiva de los resultados, siguiendo la línea iniciada en el trabajo de Marrero *et al.* (2003), recurriendo tanto a la manipulación del habla natural como a la síntesis.

BIBLIOGRAFÍA

ATAL, B.S.: "Automatic Speaker recognition based on pitch contours", *J.A.S.A* 52, 1972, págs. 1687-1697.

BORDONE-SACERDOTE, C. y G.G. SACERDOTE: "Some spectral properties of individual voices", *Acustica* 21, 1969, págs. 199-210.

BROWN, R.: "An experimental study of the relative importance of acoustic parameters for auditory speaker recognition", *Language and Speech* 24, 4, 1981, págs. 295-310.

CARRELL, T.D.: "Contributions of fundamental frequency, formant spacing, and glottal waveform to talker identification", *Research on Speech Perception Technical Report* (Indiana University Speech Laboratory), 5, 1984.

COMPTON, A.J.: "Effects of filtering and vocal duration upon the identification of speakers aurally", *J.A.S.A.* 35, 1963, págs. 1748-1752.

DODDINGTON, G.R.: "Speaker recognition. Identifying people by their voices", *Proc. IEEE* 73, 1985, págs. 1651-1664.

DOHERTY, E. T.: "An evaluation of selected acoustic parameters for use in speaker identification", *Journal of Phonetics* 4, 1976, págs. 321-326.

DOHERTY, E. y H. HOLLIEN: "Multiple factor speaker identification of normal and distorted speech", *Journal of Phonetics* 6, 1978, págs. 1-8.

DOMMELEN, W. A. van: "The contribution of speech rhythm and pitch to speaker recognition", *Language and Speech* 30, 4, 1997, págs. 325-338.

ESKENAZI, D., G. CHILDERS y D.M. Hicks: "Acoustic correlates of vocal quality", *Journal of Speech and Hearing Research* 33, 1990, págs. 298-306.

FURUI, S.: "Research on individuality features in speech waves and automatic speaker recognition techniques", *Speech Communication* 5, 2, 1986, págs. 183-197.

GOBL, C.: "A preliminary study of acoustic voice quality correlates", *STL-Quarterly Progress Status Report* 4, 1989, págs. 9-22.

GLENN, J. W. y N. KLEINER: "Speaker identification based on nasal phonation", *J.A.S.A.* 43, 1968, págs. 368-372.

HOLLIEN, H.: *The Acoustics of Crime. The New Science of Forensic Phonetics,* Nueva York, Plenum, 1990,.

HOLLIEN, H.: "The profile approach to speaker identification", *Actes du XIIème Congrès International des Sciences Phonétiques* (Aix-en-Provence, 1991), Aix, Université de Provence, 1991, págs. 396-401.

HOLLIEN, H. y MAJEWSKI, W.: "Speaker identification by long-term spectra under normal and distorted speech", *J.A.S.A.* 62, 1977, págs. 975-980.

INGRAM, J.C.L., PRANDOLINI, R. y ONG, S.: "Formant trajectories as indices of phonetic variation for speaker identification", *Forensic Linguistics*, vol. 3-1, 1996, 129-145.

ITOH, K. y SAITO, S.: "Effects of acoustical feature parameters of speech on perceptual identification of speaker", *IECE Trans.* Vol. J65-A, 1982, págs. 101-108.

JOHNSON, C.C., HOLLIEN, H., y J.W. HICKS Jr.: "Speaker identification utilizing selected temporal speech features", *Journal of Phonetics* 12, 1984, págs. 319-327.

KARLSONN, I.: "Glottal waveform parameters for different speaker types", *Proc. Speech '88, 7th FASE Symposium,* vol. 1, 1988, págs. 225-231.

KUWABARA, H. y OHGUSHI, K.: "Contributions of vocal tract resonants frequencies and bandwidths to the personal perception of speech", *Acustica* 63, 1987, págs 121-128.

KUWABARA, H. y SAGISAKA, Y:. "Acoustic characteristics of speaker individuality: Control and conversion", *Speech Communication* 16, 1995, págs. 165-173.

KUWABARA, H. y TAKAGI, T.: "Acoustic parameters of voice individuality and voice-quality control by analysis-synthesis method", *Speech Communication* 10, 1991, págs. 491-495.

LARIVIERE, C.: "Contribution of fundamental frequency and formant frequencies to speaker identification", *Phonetica* 31, 1975, págs. 185-197.

MARRERO, V., J. GIL y E. BATTANER, E.: "Inter-Speaker Variation in Spanish. An Experimental and Acoustic Preliminary Approach", *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona, 2003, págs. 703-706. http://liceu.uab.es/~joaquim/phonetics/VILE/VILE_ICPhS03.pdf

MATSUMOTO, H. *et al.:* "Multidimensional representation of personal quality of vowels and its acoustical correlates", *IEEE Trans.* Vol. AU, 21, 1973, págs. 428-436.

MILLER, J.E.: "Decapitation and recapitation: a study of voice quality", *J.A.S.A.* 36, 1964.

ORTEGA, J., J. GONZÁLEZ y V. MARRERO: "AHUMADA: A large corpus in Spanish for speaker characterization and identification", *Speech Communication* 31 (2-3), 2000, págs. 255-264

PITTAM, J.: "The long-term spectral measurement of voice quality as a social and personality marker: a review", *Language and Speech* 30, 1987, págs. 1-13.

PRUZANSKY, S.: "Pattern matching procedure for automatic for automatic talker recognition", *J.A.S.A.* 35, 1963, págs. 354-358.

SHEARME, J.N. y J.N. HOLMES: "An experiment concerning the recognition of voices", *Language and Speech* 2, 1959, págs. 123-131.

STEVENS, K.: "Sources of inter- and intra-speaker variability in the acoustic properties of speech sounds", *Proceedings of the 7th International Congress of Phonetic Sciences.*, Montreal, 1971, págs. 206-227.

STEVENS, K. *et al.:* "Speaker identification and authentication: a comparison of spectrographic and auditory presentation of speech materials", *J.A.S.A.* 44, 1968, págs. 1596-1607.

SU, L. S, LI, K. P. y FU, K. S.: "Identification of speakers by use of nasal coarticulation", *J.A.S.A.* 56, 1974, págs. 1867-1882.

WOLF, J. J.: "Efficient acoustic parameters for speaker recognition", *J.A.S.A.* 51, 1972, págs. 2044-2056.

TRANSCRIBING PROSODY USING SYNTAX AND PRAGMATICS

DANIELA BRAGA
Faculty of Philology of University of Coruña dbraga@udc.es
LUÍS COELHO
Polytechnic Institute of Porto luiscoelho@eseig.ipp.pt
DIAMANTINO FREITAS
Faculty of Engineering of University of Porto dfreitas@fe.up.pt

INTRODUCTION

In recent developments on speech processing, the expressions "attitudinal prosody", "emotional speech" and "affective computing" became frequent. This seems to indicate that, once the segmental quality of TTS systems is controlled, the next step will be the optimized manipulation of suprasegmental or prosodic levels, which is determined not only by linguistic factors (lexical, syntactic and semantic), but also by pragmatic and communicative factors. Therefore, prosody, known as one of the main factors involved in the so much wanted synthetic voice naturalness, can also give information about the speaker's pragmatic and communicative objective, about the speaker's certainty or uncertainty about the propositional content of his utterance and even about his formal or informal behaviour towards the listener.

However, it seems that there aren't any available resources for European Portuguese that is specifically orientated for the interfaces between syntax, prosody and pragmatics.

Another problem has to do with corpora transcription. The task of transcribing natural speech and dialogue involves important decisions concerning the segmental features to be chosen. But the most difficult task is to tackle the suprasegmental level in speech, as well as to choose the prosodic categories whose comprehension is so necessary, for example, to TTS naturalness and discourse analysis.

Two basic approaches have been followed regarding this subject: the transcription parameters that were brought up by linguistic research on conversation analysis, on one hand, and the transcription systems that were developed by speech science and technology research, on the other hand.

Linguistic studies on conversation and discourse analysis have paid attention to natural speech paralinguistic phenomena since the late seventies and early eighties. Some of the conventions that became common in conversational analysis were proposed by Gail Jefferson [1], Maxwell Atkinson, Paul Drew [2]. The labels used in database construction were still very simple and the prosodic information provided was very close to the one given by punctuation marks, which means that it was not accounting the syllable accent fall/ rises nor the durations and intra-phrasal boundaries.

In what speech science is concerned, we would like to mention at least four prosodic models that have been largely used for multi-language prosody transcription and/ or prediction.

ToBI (Tones and Break Indices) [3] is a spoken language transcription system composed by prosodic labels that represent F0's increases and decreases concerning both syllable accents and phrase accents. Besides this, in the ToBI model, initial and terminal phrase boundaries are also marked as well as the duration of the boundary.

The TILT model [4] represents intonation in the shape of a linear sequence of events, which may be f_0 accents or boundary tones. Each event is characterized according to continuous parameters representing amplitude, duration and tilt (measure of the shape of the event).

INTSINT (INternational Transcription System for INTonation) [5] is an intonation transcription system which codifies f_0 patterns using a set of abstract tone symbols, such as Absolute symbols (*Top, Mid, Bottom*), for a speaker f_0 variance range, and relative symbols (Higher, Same, Lower, Up stepped, Down stepped) relative to the previous target-point. Each symbol characterizes a target-point in the phonetic transcription and these points are later expanded by the MOMEL algorithm (MOdélisation de MELodie), which allows for automatic modelling of the macroprosodic component of the f_0 contour with a sequence of points which define a quadratic spline function.

Finally, the Fujisaki model [6] is a physiologically inspired and mathematically-based model that uses accent patterns (derived from accent commands) and phrase patterns (derived from phrase commands) in order to model and predicts f_0 contours. It employs two functions with parameters which produce logarithmic quantities that can be added to produce the desired logarithmic F_0 patterns. The parameters must be found from experimental studies of the real sounds. The base value of f_0 and the timing commands (inputs to the two Fujisaki functions) are required, with precise time and amplitude values. The Fujisaki model has been applied to several other languages with success and was implemented for EP by J. P. Teixeira [7]. For several reasons, we have been using this model for prosody prediction, but we believe that syntactic and pragmatic information can improve the model efficiency.

In what pragmatics is concerned, we have been working within the discourse analysis framework, specifically within the French interactionism [8]. The main pragmatic topics we are considering are speech acts, style, attitude, context and the expression of emotion.

The syntactic framework we are using is a hierarchical dependency model.

CORPUS DESCRIPTION AND METHODOLOGY

Spontaneous speech data is a type of resource that provides valuable information on dialogue interaction and structure, on speaker's communicative intention, on phonetic and phonologic language system and on prosodic behaviour. Moreover, it becomes possible to understand the correlation between individual prosodic behaviour, syntactic motivations and pragmatic strategies in oral discourse. This can be done by defining syntactic and pragmatic units and putting them in correlation with the prosody levels. A database of this kind, both with linguistic, pragmatic and prosodic labels, is also important for development of Text-to-Speech (TTS) intelligibility and naturalness, since it is possible to extract f_0 patterns and duration models either by using a linguistic rule-based approach (deriving and generating f_0 from syntactic boundaries, called phrase boundaries) or by using neural networks, Hidden Markov Models or other stochastic methods of pattern recognition and classification.

Nevertheless, natural speech corpora for EP are scarce and when they exist three possible situations can happen: they are hardly accessible; they are not prosodically labelled, which limits a large scope of studies concerning pragmatics and prosody; and if they are prosodically labelled, they often have a rudimentary system based on punctuation marks. Another problem often found in this sort of databases is the difficult access to the original sound file, which only allows us to work with the transcribed text file.

To pursuit our objectives of studying the correlation between syntax, prosody and pragmatics, we have been building a spontaneous speech corpus in EP conventionally named ProGmatica.

The used criteria for structuring our database, which is still ongoing, are based on the well known John Searle's speech acts typology [9]. For our database we also considered argumentative speech acts, based on C. Kerbrat-Orecchioni's research [8] on verbal interactions. We have then expanded the former Searle's typology by including some more speech acts, such as agreement, disagreement, criticism and mockery, as shown in figure 1. Some speech acts were subdivided into categories, such as questions, whenever their syntactic and prosodic structure justified the distinction.

At this stage, no declarative speech acts were considered because of their inferior importance for TTS systems.

ProGmatica consists of broadcasted television materials in which natural spontaneous speech occurs, such as interviews, talk shows and political debates. Soap operas have also been included in this database. Although they can not be considered spontaneous speech, since they follow a script, they are performed by professional actors. This way, we think that soap operas may offer more varied speech styles and that it may be a privileged environment to study emotional speech, even though it is acted. From the whole recorded material, we selected, extracted and saved prototypical speech acts with the help of Praat tool [10].

Comunicative goal (Speech Act)				
To obtain listener's reaction	To fulfil social functions	To compromise	To persuade	To relate speaker with truth
To order	To thank	To promise	To criticize	To aprove
To request	To apologize	To swear	To agree	To disaprove
To suggest	To flater		To disagree	To inform
To advise	To say hello		To mock	To afirm
To question	To say goodbye		To refute	
To request info.	To feel sorry			

Fig. 1: Pragmatic structure of ProGmatica.

The corpus was recorded in a VHS format and then converted to a digital format, with a sampling rate of 11025 Hz. After this, the speech files were labelled according to the categories presented in the next chapter.

Therefore, we have a multi-speaker corpus, where linguistic, paralinguistic and extra linguistic information are related to each other. So far, this task is being performed by a professional linguist but perceptive tests must be done afterwards in order to validate these selections and classifications.

TRANSCRIPTION LEVELS

In the transcription of the sound material of ProGmatica, the following tiers were considered: phonetic, orthographic, morpho-syntactic, phrasing, focus, emotion type (among seven: neutral, joy, sadness, anger, fear, indignation, surprise), style (polite/impolite; formal/informal) and context information.

In the phonetic tier, a phone transcription is performed with the help of an automatic annotation tool we have developed named EasyNote [11]. This labelling tool considers some special features for the acoustic and articulatory dimension of EP and uses language models and phonological rules for error detection and correction, leading this way to results close to those obtained by human annotators. Tonic syllables are marked in this tier, although it is accent information and therefore prosodic information. This is an example of information levels' correlation. In the phonetic level, paralinguistic signals, whenever produced by the participant who is speaking, are manually marked, as well as silences and breaks.

Acoustic analysis is also obtained with the help of Praat tool. Several parameters were considered, such as intensity, formants, f_0 average, maximum and minimum pitch.

All utterances were orthographically transcribed so as to guide users along the database. We have also considered a morphological and syntactic level in which words are classified according to their grammatical or functional class.

The prosodic annotation in this level directly derives from syntax. In this level, phrasing and focus are marked.

Higher tiers concern to emotion type, even though we should talk in terms of emotion degree and not category, style and context. In this last tier, details about the program, date of broadcasting and speaker can be found.

In figure 2, 3, 4 and 5 below, examples of transcribed segments of this corpus are displayed.

In the first tier, right after the wave form and the spectrogram analysis, the corresponding phonetic transcription is displayed together with the tonic syllable markup. In the second tier, an orthographic transcription is done. Subsequently, there is the morpho-syntactic tier. Anchored in the last one, there is the phrasing and the focus tier, where syntax and prosody are in deeply connection. Immediately after, the pragmatic tiers of emotion, style and context are displayed.

This transcription system allows both linguists and speech scientists to extract patterns and correlations between the morpho-syntactic and phonological levels and the prosodic level. In other words, the f_0 accent commands are anchored in the syllabic labels, on one hand, and the phrase commands derive from the phrase boundaries, on the other hand. This approach is made in the scope of prosodic modelling (stated in [7]) to extract f_0 patterns from natural speech in order to generate artificial f_0 in TTS systems. In this model, we combine in varying degrees morpho-syntactic information (rules) and statistical information (artificial neural networks) to derive prosodic patterns.

Moreover, the pragmatic levels regarded together with the prosodic and linguistic levels can be more enlightened especially when analyzing the importance of word/syllable duration, f_0 and focus manipulation. Besides, both syntactic/prosodic correlation [12] and pragmatics/prosody interface [13] are topics we have been working on for some time.

SOME RESULTS AND DISCUSSION

The analysis of the transcribed data has been giving us many interesting results from which we can only give a few examples here, due to printing constraints.

In figure 2, there is an example of an utterance which was categorized as a promise, in terms of speech acts. The pragmatic reasons stated to justify this classification are the context in which it was said (a political debate on the eve of a government election) and the political profile of the speaker, as he was one of the proposed candidates. In this particular utterance, no emotion could be noticed. However, we can say that in promises there is always a certain degree of assertiveness that is derived from the speech rate and focus. In other words, a higher speech rate conveys more assertiveness and certainty to the promise that is being uttered. In what morpho-syntax is concerned, we can find future tense verbal constructions, either verbal periphrasis (e.g. *"vai haver"*, *"vai ser"* it will be) or simple future constructions (*"faremos"* we will do). Topicalized constructions can also be found (e.g. *"E o que faremos é tirá-los da pobreza"* And what we will do is take them out of poverty).

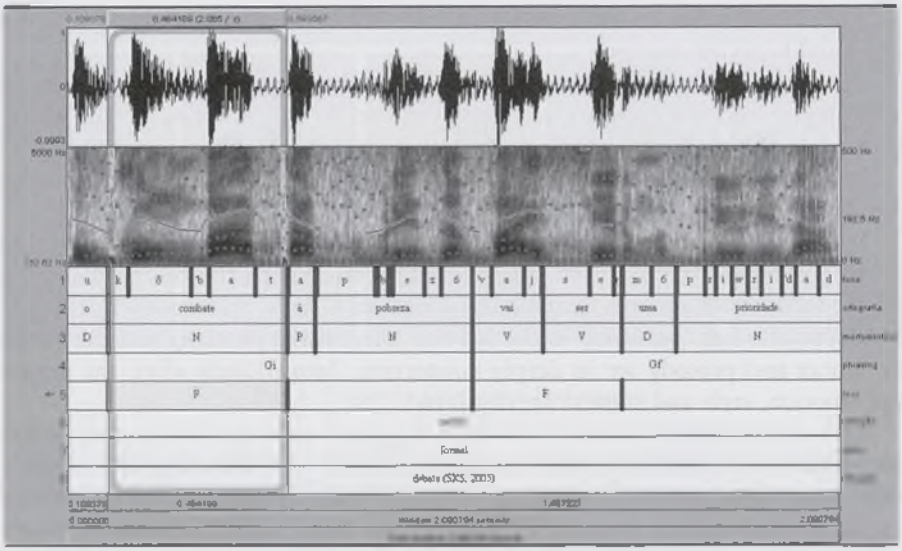


Fig. 2. Example of a Promise: "O combate à pobreza vai ser uma prioridade" (The combat to poverty will be a priority).

In figure 2, we can see an example of a promise syntactically uttered trough a declarative sentence composed by simple subject and predicate. Nouns are more prone to be the syntactic and prosodic focus. As we can see in figure 2, the subject word "combate" (combat) has an f_0 increase, which is higher over the tonic syllable, when comparing it with the other nouns. In the predicate, there is also an f_0 increase in the verbal expression "vai ser", marking a secondary focus and a syntactic boundary as well, very common between subject and predicate. Prosodically, this boundary is marked with a decrease followed by an increase of f_0 on the tonic syllables. In other observed cases, we could see that phrasing can also be marked by short pauses or silences. Phrasing is, therefore, very connected with focus, since it is also a strategy to enhance the focused word, together with f_0 manipulation and tonic syllable intensification.

In figure 3, a sample of a refutation speech act is shown. Pragmatically, the main purpose of this speech act is to deny the other's statement or disagree with the other's opinion. It can be included in a type of assertive modality, since the final aim is to convey certainty and accuracy to what is being denied. Syntactically speaking, it is characterized by the multiple repetitions of negative adverbs, especially the "*não*" (no) adverb, but the negation and repetition of parts of the sentence can also be found. Emphatic expressions are also common, particularly with the expression "*é que*", as shown in figure 3. Prosodically, focus and phrasing are more prone to happen in the negative adverb. In other words, negative adverbs are often intensified through a pitch range increase, generating

a certain broad focus voice segment. Most of the times there is a serious increase of f_0 and intensity in the negative adverb. Intensive phrasing is also common.

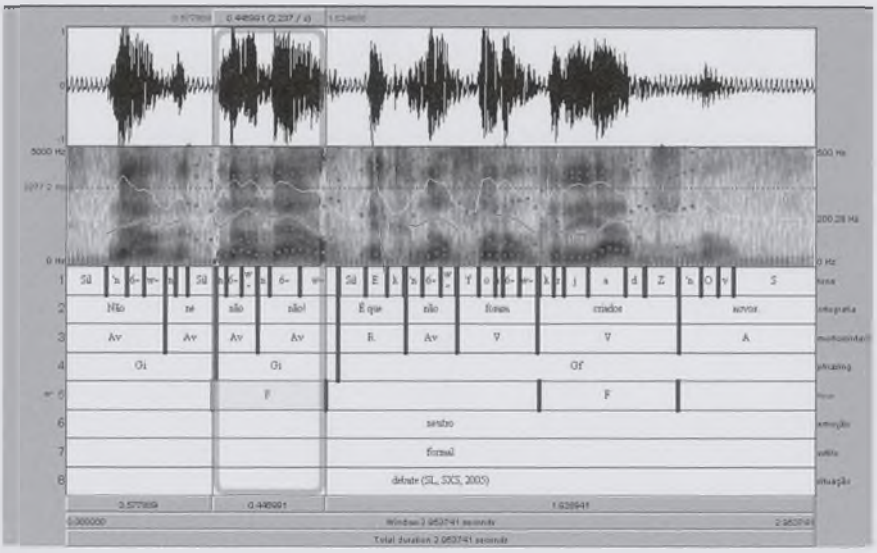


Fig. 3. Example of a Refutation: "Não, não não! É que não foram criados novos" (No, no no! The point is they didn't create new ones).

In figures 4 and 5 we have an example of how prosody and pragmatics can be decisive in the understanding of syntactical ambiguous sentences.

In figure 4, an example of a typical WH- question in EP is shown. In terms of syntactic characterization, it is an interrogative sentence that begins with an interrogative word ("*o que*" what), either adjective or pronoun. In terms of pragmatic objective, it is a type of speech act in which the speaker wants to obtain a listener's reaction, which is, in this particular case, the listener's point of view (the question is asked by an interviewer to a politician in the context of a political debate). Prosodically, we can see a general f_0 decrease till the end. Besides, this event is widely described for EP.

In figure 5, a similar syntactic structure can be seen: an interrogative sentence starting with a WH- word ("*Por que*" why). In questions, vocatives are very common, introducing phrase boundaries.

When we contrast figure 4 with figure 5, an interesting difference can be often seen: instead of the expected pitch range final falling, we hear a pitch range final increase. In other words, in figure 5 the announced syntactic structure is different from the pragmatic intention. The speaker uses a WH- question, but despite this fact, he is not expecting to be answered, since in a pragmatic point of view he is asking a rhetorical

question. That is probably why the final f_0 pattern doesn't fall down, as it usually happens in this kind of phrase, but in spite of that it increases in the last voiced syllable.

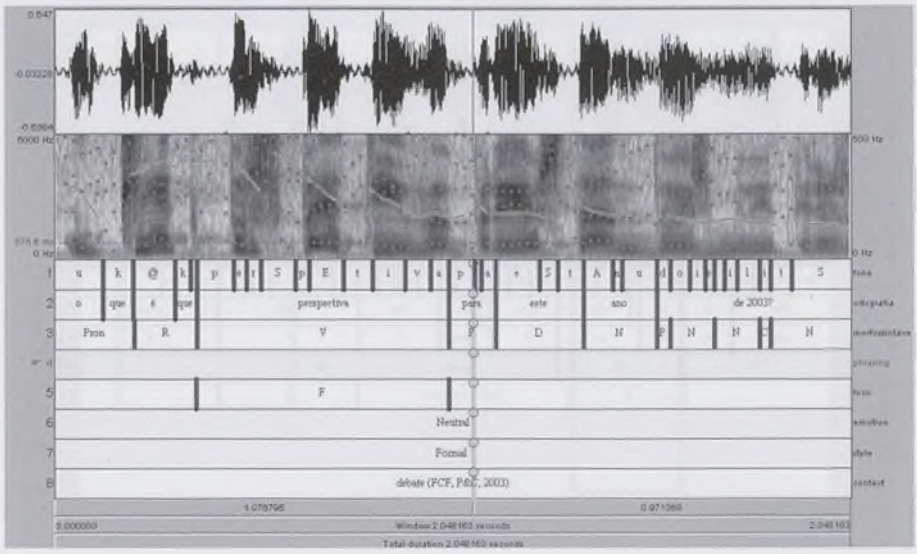


Fig. 4. Example of a WH- Question: *"O que é que perspectiva para este ano de 2003?"* (What do you foresee for this year of 2003?)

Finally, we would like to mention that phrase boundaries only perceived by pitch breaks (what is called phrase commands, within the Fujisaki model) are more frequent in oral speech style than in reading speech style. This fact can be related to the argumentative force and the illocutionary objective of the speaker. In other words, by making more breaks in their speech flow, the speaker may convey his idea in smaller sequences, allowing the listener to have a closer understanding of his argument, especially when analysing the political debate type of interaction. Thus, the participants state their reasons in a more effective way by using prosody.

However, this task of description and overlapping all linguistic, paralinguistic and non linguistic information has been showing that there are a huge amount of ambiguous cases when identifying speech acts and emotions. In other words, a promise can be mistaken with a simple assertion or affirmation for a computer program, which means that probably semantics could help in this difficult task. Work on this subject is presently on progress.

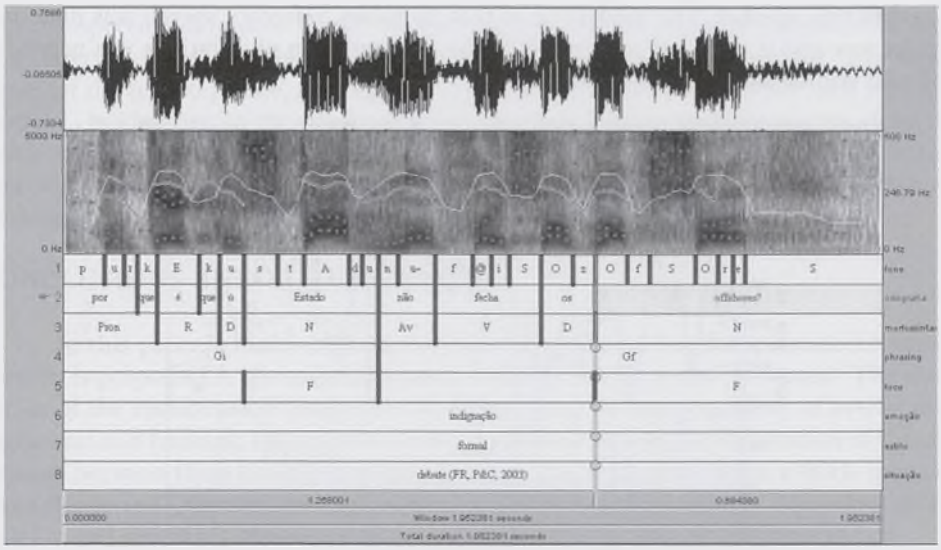


Fig. 5. Example of a Rethorical Question: "Por *que é que o Estado não fecha os offshores?*" (Why doesn't the government close the offshores?)

APPLICATIONS

The main type of applications of the present study and labelling system is to prosody research for spontaneous speech. A second type of research that is targeted is discourse modelling.

As an illustration of some possibilities, we present in the following the Fujisaki model application to an assertion only in terms of f_0. Duration studies are also possible, and were already performed for EP reading speech style by Teixeira [7], but we still didn't apply it to spontaneous speech, although this action is foreseen. With this aim, we selected a specific segment of speech, labelled with this type of speech act, and we superimposed an f_0 generated by the Fujisaki model by use of the developed tool [7]. This leads us to a very good match to the original f_0, as can be seen in the screen-shot of Figure 6 in the middle graph where the "+" signs mark the original f_0 values, the lower thin line marks the phrase pattern and the upper thin line marks the full pattern. The outcome is a set of model parameters that can be identified in the lower graphs of figures 6 and 7.

In order to understand which are the specific f_0 movements related to the assertion, in a first preliminary study, manipulation of the pattern was done in order to reach a sentence that sounded like read speech. From the differences we extracted a few rules about the prosodic characteristics of the assertion modality. It must be said that this manipulation was, at this phase of the work, guided by example from read speech recorded

by a different speaker. The modelling of this, so to say, reference speech was done in a similar way and is depicted in figure 6. Some manipulation of durations was needed as well and will be addressed in a future paper.

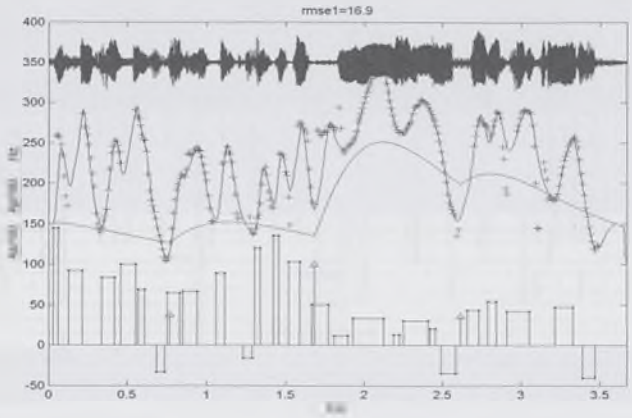


Fig. 6. Example of f0 modelling of the assertion "*e sobre isso, o Governo não tem política nenhuma. pelo contrario*", upper graph – speech waveform, middle – f0 contours (+ indicates original values), lower – accent and phrase Fujisaki commands.

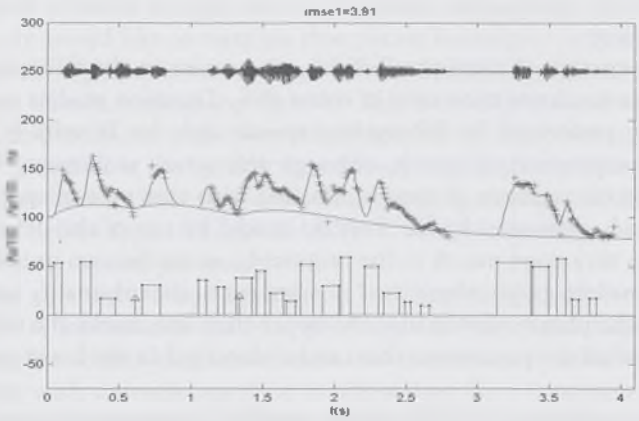


Fig. 7. Example of f0 modelling of the same utterance of figure 6, but in reading style by a different male speaker. Major differences can be observed especially in words "*nenhuma*" and "*contrario*" and in general declination. Other utterances showed different average f0 values as well.

Inspection of figures' 6 and 7 middle graphs and lower graphs suggests the main f_0 differences introduced by assertion: a strong accent in the focus word "nenhuma" (none) of the second phrase, a strong accent in the word "contrario" of the final phrase, a deeper f_0 modulation in the whole utterance producing some negative accents along the whole sentence and a smaller global declination. F_0 values are also higher that in read speech in average (preliminary calculations indicate approximately 0, 35 octaves increase in average f_0).

CONCLUSIONS AND FUTURE WORD

In this paper, a multi-informational transcription system for spontaneous speech corpora is proposed. A spontaneous speech database for EP is also presented. The innovation of the transcription categories considered is set on the possibility of overlapping segmental and prosodic information with pragmatic features. An illustration of the correlation between these three levels was proved in the contrast between a WH- question and a Rhetorical Question, whose success and interpretation is directly dependent on the syllabic and general f_0 behaviour that was in this case opposite to the expected one.

Although in a preliminary stage of results, it was also observed that, in the assertion, the prosodic groups are shorter, deriving from smaller syntactic units, limited with long pauses, this way conveying more effective information to the audience, to whom is given enough time to understand what is being said.

The analysis of the assertion as a speech act whose effectiveness is mainly supported by an f_0 enhancement is quite interesting for pragmatic studies. However the possibility of extracting a prosodic pattern of the assertion or other pragmatic speech act is of extreme importance to speech synthesis naturalness and variety.

Besides, our corpus can provide a lot of information, concerning durations, speech rate, micro-prosody, stylistic prosodic diversity, emotion and style, in order to extract rules and trends that can be submitted to neural networks, HMMs or other prosodic modulation solutions.

This spontaneous speech corpus is intended to be a work platform to lexicometric analysis for pragmatic purposes as well as f_0 and durations Fujisaki models (platform under development for TTS applications) applied to EP, in a more extensive application. Recognition could also benefit from this linguistic resource: once it is labelled, it is possible to extract the words' acoustical characteristics and then use them in a word based neural network recognition model.

In the near future, a more detailed description of the pragmatic features associated with the prosodic patterns and syntactic structures is foreseen, allowing to synthesize an even more natural intonation pattern and to contribute to a better understanding of argumentation in conversation. This ongoing study is meant to observe and describe real speech, enabling us to prepare texts to be spoken with a better performance.

Regarding to speech emotions, we are planning to use semantic knowledge in order to tackle ambiguity in emotion identification.

BIBLIOGRAPHY

1. JEFFERSON, Gail.: "Sequential Aspects of Story-telling in Conversation", in SCHENKEIN (ed) *Studies in Organization of Conversational Interaction*, New York, Academic Press, 1978.
2. ATKINSON, Maxwell; DREW, Paul: *Order in Court*, Atlantic Highlands, New Jersey, 1979.
3. SILVERMAN, et al: "ToBI: a standard for labelling English prosody", in MANNELL, R.; ROBERT-RIBES, J (eds), *Proceedings of the 1998 International Conference on Spoken Language Processing*, vol. 2 Sydney, ASSTA, 1992.
4. TAYLOR, et al.: "The rise/ fall/ connection model of intonation", in *Speech Communication 15*, 1994.
5. DI CRISTO, A. et al.: "A metrical model of rhythm and intonation for French text-to-speech synthesis", in Botinis, A (eds), *Intonation: theory, models and applications. Proceedings of an ESCA workshop*, Athens: European Speech Communication Association, 1997.
6. FUJISAKI, Y. "Analysis of voice fundamental frequency contours for declarative sentences of Japanese", *Journal of Acoustical Society of Japan* (E), 1984.
7. TEIXEIRA, J.: *A Prosody Model to TTS Systems*, Faculty of Engineering of University of Porto, PhD Thesis, 2004.
8. KERBRAT-ORECCHIONI, C.: *La conversation*. Paris: Seuil, 1996.
9. SEARLE, J.: *Speech Acts. An essay in the philosophy of language*, Cambridge: Cambridge University Press, 1969.
10. PRAAT. Available at: http://www.fon.hum.uva.nl/praat/
11. COELHO, L.: *Etiquetagem Automática de Sinais de Fala*. Faculty of Engineering of University of Porto, MSc Thesis, 2005.
12. FREITAS, D. and BRAGA, D.: "Correlation between phonetic factors and linguistic events regarding a prosodic pattern of European Portuguese: a practical proposal", *Proceedings of ICSP2001*, Korea, 2001.
13. BRAGA, D.; MARQUES, A.: "The pragmatics of prosodic features in the political debate", *Proceedings of Speech Prosody 2004*, Japan, 2004.

PATRONES MELÓDICOS DEL ESPAÑOL EN HABLA ESPONTÁNEA

FRANCISCO JOSÉ CANTERO SERENA
Laboratorio de Fonética Aplicada – Universitat de Barcelona

1. INTRODUCCIÓN

Este trabajo[1] constituye la culminación de los trabajos presentados en Cantero et al. (2002) sobre la entonación interrogativa y en Cantero et al. (2005) sobre la entonación enfática. En él se presentan los resultados de nuestra investigación sobre la entonación del español en habla espontánea, realizada en nuestro centro a lo largo de los últimos seis años[2].

El corpus de habla espontánea y el método de análisis

Nuestro centro de investigación, el Laboratorio de Fonética Aplicada de la Univ. de Barcelona, está ubicado en un contexto académico (la Facultad de Formación de Profesores) que orienta toda toda su actividad científica en torno a la formación de profesores, maestros y logopedas. En este contexto, nos hemos planteado el reto de ofrecer modelos de análisis y patrones de actuación didáctica directamente relacionados con la realidad lingüística y comunicativa, más allá de los tradicionales modelos abstractos que suele ofrecer la fonética teórica y descriptiva. Así, todas nuestras actividades de investigación están relacionadas con la descripción del habla espontánea, y muy especialmente con el análisis de la entonación.

---

1 Esta investigación se inscribe en el proyecto *Fonética Aplicada a la Educación* (ref. BSO2002-03479), financiado por el Ministerio de Ciencia y Tecnología.

2 Paralelamente, se han realizado los trabajos de Font et al. (2002) y Font (2005) sobre la entonación del catalán, así como los de Liu (2003; Liu & Cantero, 2002) y Cortés (2000, 2004, 2005) centrados en la adquisición de la entonación española por parte de hablantes de una lengua tonal como es el chino, y el de Torregrosa (1999) sobre correlación kinésico-entonativa (cuya finalización está prevista para 2006), además de numerosos trabajos sobre análisis contrastivo y didáctica de la pronunciación: v. novedades en http:/www.ub.es/lfa).

Aunque en nuestros días contamos con sofisticados medios técnicos y metodológicos, los fonetistas aún solemos encontrarnos con trabajos que describen la entonación únicamente del habla de laboratorio (es decir, leída o inducida por el investigador), con la peregrina excusa de la calidad acústica de las grabaciones. En nuestros trabajos, hemos demostrado que podemos adquirir habla genuinamente espontánea (por ejemplo, en registros televisivos) con una sobrada calidad de grabación.

En realidad, si no se afronta el habla espontánea es por razones exclusivamente metodológicas: no todos los métodos de análisis de la entonación permiten afrontar la variabilidad del habla espontánea. O, lo que es lo mismo, algunos métodos de análisis sólo permiten comprobar lo que el investigador ya sabe (puesto que en el laboratorio sólo puede describirse lo que ya se había previsto, lo que se había preparado de antemano).

Otra característica del habla de laboratorio es su "singularidad": a menudo, se describe el habla (leída, o inducida) de un único informante (en ocasiones, ¡el mismo investigador!) o de un reducido número de informantes, advertidos de su participación en una investigación.

En nuestra investigación, por el contrario, hemos contado con un total de 57 informantes distintos (25 hombres y 32 mujeres), todos ellos anónimos, hablantes nativos de español, de distinta procedencia geográfica y nivel cultural, que componen una muestra ponderada y diversa de las hablas peninsulares. Ninguno de ellos, desde luego, tenía la más mínima idea de que podía ser informante en nuestras investigaciones.

El corpus se elaboró a partir de más de 6 horas de grabación tomada de diversos programas de televisión, durante el mes de agosto de 1999. Consta de 136 enunciados[3] emitidos en situación de habla espontánea genuina: no inducida, ni manipulada en ningún sentido. De ellos, 46 enunciados constituían entonaciones inequívocamente interrogativas en su contexto; 40 enunciados, entonaciones inequívocamente enfáticas; 50 enunciados, entonaciones neutras.

El método de análisis melódico empleado en la investigación se ha expuesto en Cantero (1999) y Cantero (2002), así como en Font (2005b, en este mismo volumen). Constituye un método robusto, que permite el análisis de habla espontánea, ya que las unidades que maneja no son apriorísticas sino genuinamente descriptivas: el segmento tonal (típicamente, el valor tonal de una vocal, o de una mora) y las relaciones tonales (expresadas en porcentajes de ascenso o descenso, con respecto al segmento tonal anterior).

El protocolo completo de nuestro método de análisis melódico se completa con una fase perceptiva, a partir de la síntesis de las curvas estandarizadas. Todos los patrones melódicos expuestos en este trabajo han sido validados perceptivamente, con índices de aprobación de más de un 80%.

De hecho, todos ellos son fácilmente sintetizables, a partir de cualquier enunciado previo: una vez identificados los segmentos tonales del enunciado, no hay más que modificar sus valores relativos en un programa de síntesis como *Praat* [4].

---

3 La lista de enunciados aparece en los trabajos ya reseñados Cantero et al. (2002) y Cantero et al. (2005).

4 Mediante sus rutinas *Manipulation* y *Resynthesis (PSOLA)*. Cfr: Boersma & Weenik (1992-2005).

Qué entendemos por patrones melódicos

En nuestro marco teórico[5], la entonación constituye un fenómeno comunicativo diverso: un fenómeno discursivo (o *prelingüístico*, en nuestra terminología), un fenómeno lingüístico (de tipo fonológico) y un fenómeno *paralingüístico* (o *expresivo*).

La entonación prelingüística (que reúne los fenómenos del acento y la entonación, en el complejo que llamamos *jerarquía fónica*) constituye el elemento integrador del habla, el principio unificador del discurso, que permite su inteligibilidad[6].

La entonación lingüística se refiere al nivel de análisis mediante el cual podemos designar y estudiar las unidades fonológicas de carácter suprasegmental capaces de distinguir unidades de discurso significativas (como las frases). Dichas unidades fonológicas, llamadas *tonemas*, se caracterizan mediante los rasgos fonológicos: /± interrogativa/ /± enfática/ /± suspendida/. En Cantero (2002) hemos distinguido un total de 8 tonemas en español:

1. /+interrog. +enf. +susp./
2. /+interrog. +enf. –susp./
3. /+interrog. –enf. +susp./
4. /+interrog. –enf. –susp./
5. /–interrog. +enf. +susp./
6. /–interrog. +enf. –susp./
7. /–interrog. –enf. +susp./
8. /–interrog. –enf. –susp./

La entonación paralingüística, por su parte, se refiere a los amplios márgenes de dispersión de los tonemas, en los que caben las (muy) variadas realizaciones expresivas, emotivas e idiosincrásicas del hablante. La complejidad de la entonación expresiva, de hecho, constituye los márgenes de dispersión de los tonemas /+ enfáticos/.

En este marco teórico, entendemos como *patrones melódicos* las variantes tipo de los tonemas: es decir, las melodías típicas de cada tonema, aquellas que hemos encontrado con asiduidad en nuestro corpus, que han permitido su síntesis melódica y que en las pruebas perceptivas han aparecido como entonaciones claramente reconocibles.

Los patrones melódicos, por tanto, y como *variantes tipo* de los tonemas, constituyen un nivel de abstracción intermedio entre el nivel fonológico (cuya unidad es el tonema) y el nivel fonético (cuya unidad es el rasgo melódico): la unidad de aplicación idónea en aplicaciones tecnológicas (síntesis de voz y RAH) y, sobre todo, en aplicaciones educativas. Los patrones melódicos, por tanto, pueden emplearse como modelos entonativos del habla realistas y fiables: por ejemplo, como modelos didácticos de normalidad y de excelencia.

2. EL PATRÓN MELÓDICO DE LA ENTONACIÓN NEUTRA

El primer patrón melódico, caracterizado fonológicamente como una variante de tonema /–interrog. –enf. –susp./, es el patrón melódico de la entonación neutra, llamada así por no estar marcada por ningún rasgo fonológico positivo.

---

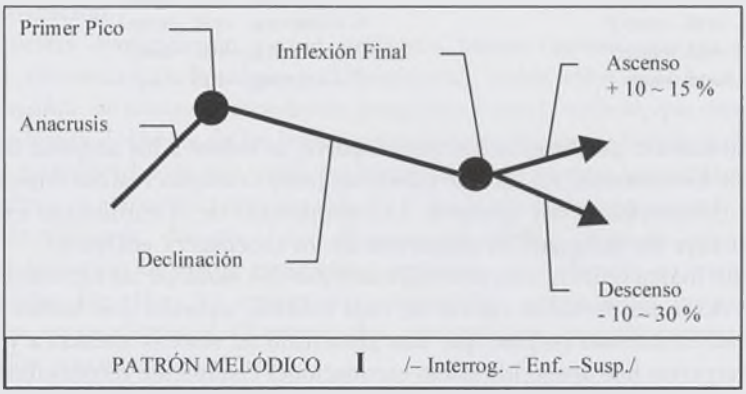
5 Expuesto detalladamente en Cantero (2002), reinterprerado y relacionado con otros modelos teóricos en Fonc (2005), y empleado en numerosas investigaciones paralelas (v. nota 2).

6 Cuya organización caracteriza los distintos dialectos (el "acento dialectal"), y cuya desorganización afecta al habla de los extranjeros (el "acento extranjero").

En principio, la entonación neutra debería ser la más frecuente, por tratarse de la menos marcada; sorprendentemente, sin embargo, en habla espontánea la encontramos con relativamente poca frecuencia, en comparación, sobre todo, con la entonación enfática (con mucho, la más común): la entonación neutra, sin ninguna marca positiva, tal vez sólo sea frecuente en la lectura en voz alta (especialmente, de textos expositivos).

En la figura del *Patrón Melódico I* aparecen los distintos elementos constituyentes del contorno entonativo: *anacrusis* (segmentos tonales previos al primer pico); *primer pico* (en principio, la primera vocal tónica; a menudo, la siguiente vocal átona); la *declinación* (sucesión de segmentos tonales entre el primer pico y la inflexión final, que típicamente constituye una leve declinación); y la *inflexión final* (la parte más relevante, y a veces la única parte del contorno: los segmentos tonales a partir de la última vocal tónica).

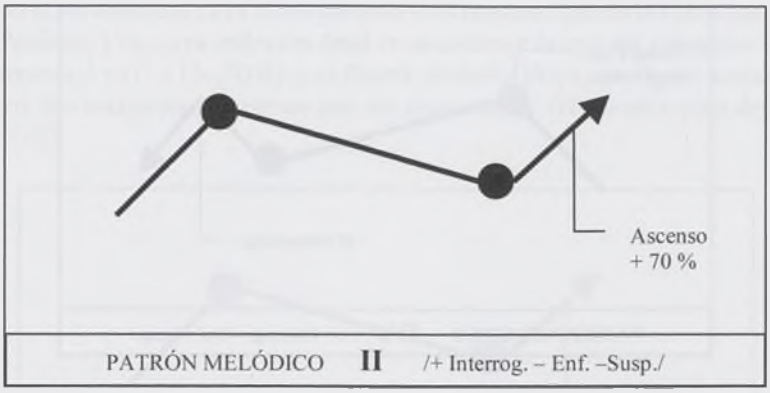
En el *Patrón Melódico I*, la inflexión final se caracteriza por un descenso moderado (entre un -10~30%), o bien por un leve ascenso final, no mayor a un +10~15%:



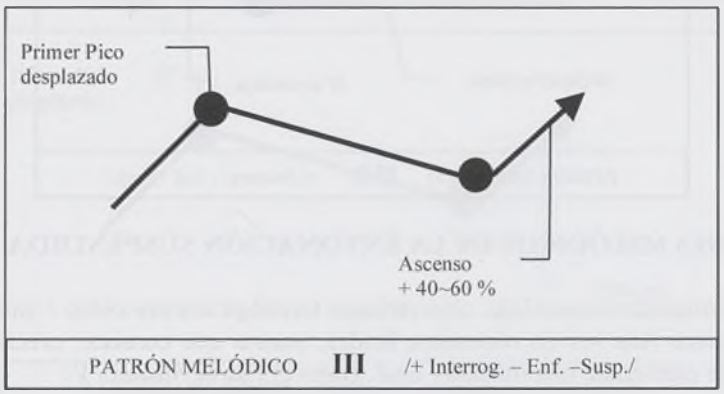
3. PATRONES MELÓDICOS DE LA ENTONACIÓN INTERROGATIVA

Los patrones melódicos de la entonación interrogativa constituyen variantes del tonema /+interrog. –enf. –susp./. En Cantero et al. (2002) se definieron estos patrones, que aquí hemos reagrupado y renombrado.

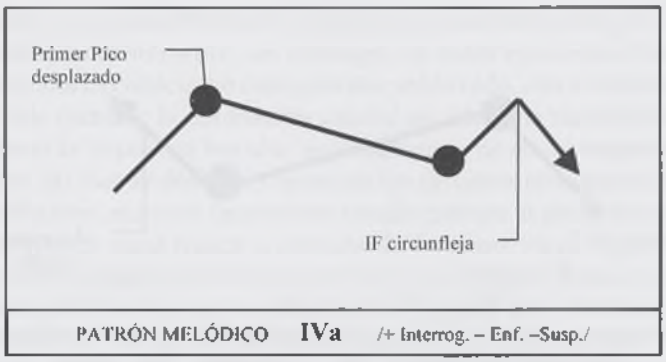
El *Patrón Melódico II* constituye el patrón típico de la entonación interrogativa, y se define en función de la inflexión final, cuyo ascenso debe ser superior a un +70~80% (y que, normalmente, es superior a un +100%):

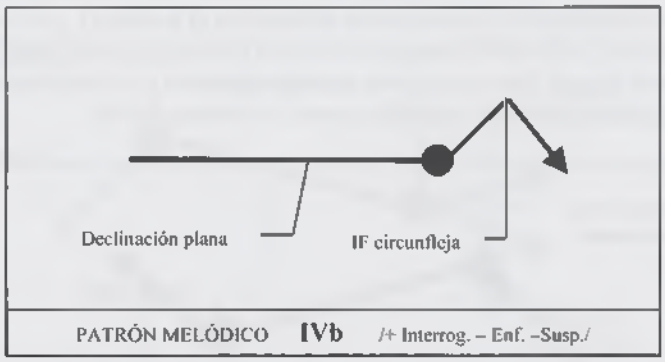


El *Patrón Melódico III* constituye una alternativa al patrón *II*, con un ascenso final menos pronunciado (+40~60%) compensado con el primer pico desplazado (desplazado a la siguiente vocal átona). Esta conjunción de rasgos (primer pico desplazado más ascenso final moderado) define el patrón melódico como /+ interrogativo/:



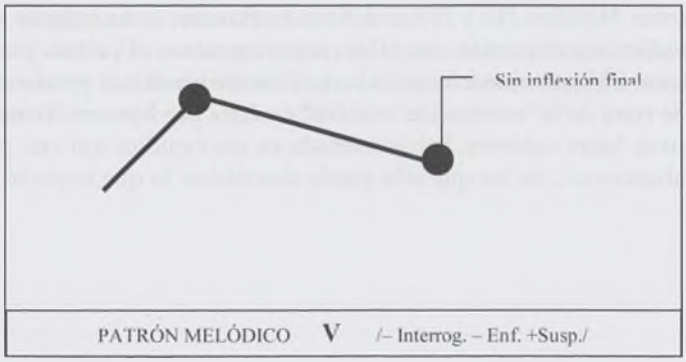
Los *Patrones Melódicos IVa* y *IVb* se definen en función de su *inflexión final circunfleja ascendente-descendente*: en el primer caso (*IVa*), conjuntamente al primer pico desplazado; en el segundo caso (*IVb*), conjuntamente a la declinación plana (sin pendiente, y sin picos ni anacrusis). Se trata de la "entonación relativa" descrita por Navarro Tomás (1944), que ningún otro autor, hasta nosotros, había reseñado en sus estudios (tal vez, por tratarse de estudios "de laboratorio", en los que sólo puede describirse lo que se prevé):



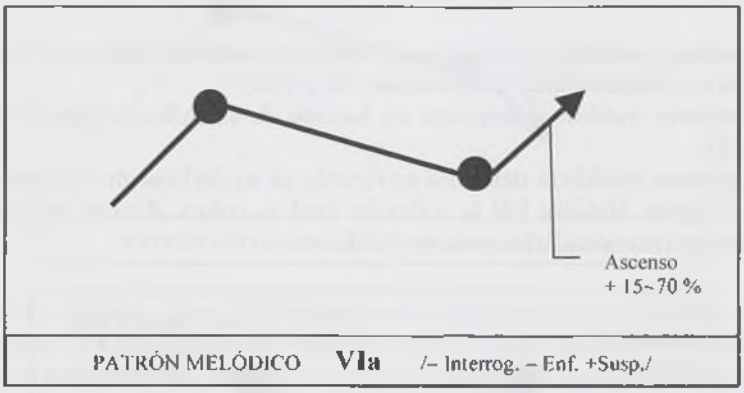


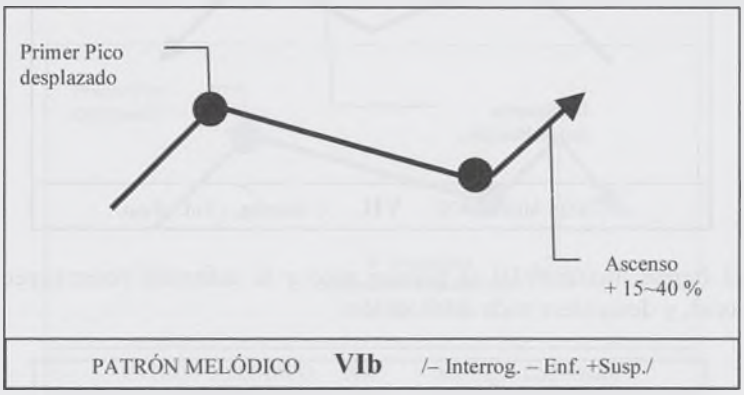
4. PATRONES MELÓDICOS DE LA ENTONACIÓN SUSPENDIDA

La entonación suspendida, caracterizada fonológicamente como /–interrog. –enf. +susp./, aparece rara vez en contornos finales, puesto que consiste, precisamente, en melodías que carecen de una inflexión final, como el *Patrón Melódico V*:



O bien, en melodías cuya inflexión final está marcada por no ser terminativa, como el *Patrón Melódico VIa*, cuya inflexión final es ascendente (entre los márgenes impuestos por los patrones *I* y *II*: +15~70%) y el *Patrón Melódico VIb*, cuyo ligero ascenso final se mueve entre los márgenes impuestos por los patrones *I* y *III* (primer pico desplazado y un +15~40%):





Lo más común es que los contornos suspendidos aparezcan como *contornos interiores* del discurso. En el análisis de la conversación, los contornos interiores /+susp./ permiten al hablante mantener su turno de habla, marcar que aún no ha terminado, por lo que constituyen uno de los recursos más relevantes en los procesos de interacción oral.

5. PATRONES MELÓDICOS DE LA ENTONACIÓN ENFÁTICA

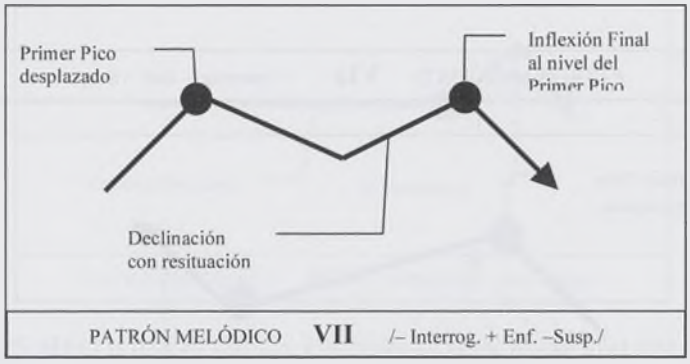
La entonación enfática se caracteriza fonológicamente como /–interrog. +enf. –susp./. Entendemos por énfasis el fenómeno melódico contrastante que permite individualizar cada enunciado, dentro de los amplios márgenes de dispersión del tonema

/+ enfático/. Estos márgenes de dispersión tan amplios son los que constituyen lo esencial de la entonación paralingüística, de la expresividad personal. En Cantero et al. (2005) hemos presentado un análisis pormenorizado de los rasgos melódicos de énfasis hallados en nuestro corpus, así como diversas muestras audio de los enunciados analizados (gracias a las ventajas multimedia de la publicación digital).

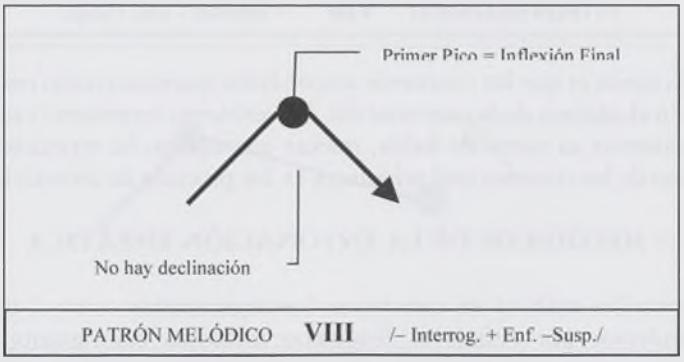
Los patrones melódicos de énfasis pueden dividirse en tres grupos:

- patrones melódicos en los que el énfasis se establece mediante el juego primer pico – inflexión final (los patrones *VII* y *VIII*)
- patrones melódicos definidos en función de su inflexión final (*IX*, *Xa-Xb* y *XI*)
- patrones melódicos definidos en función de su declinación (*XIIa-XIIb-XIIc*)

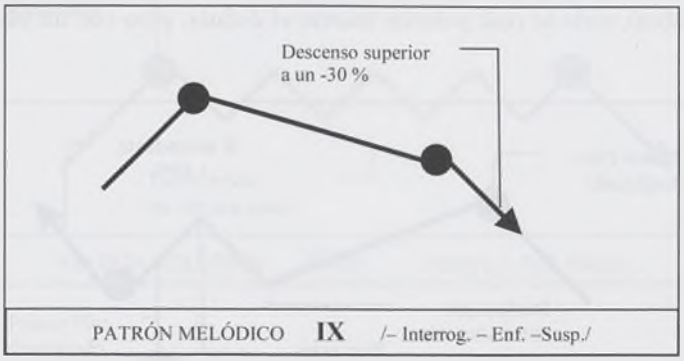
En el *Patrón Melódico VII* la inflexión final se coloca al nivel del primer pico, mediante una resituación al alza en la declinación:



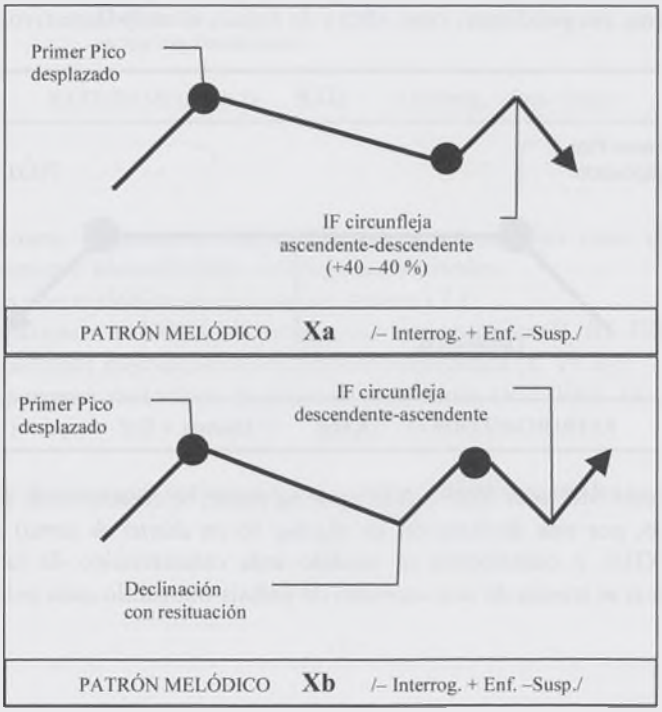
En el *Patrón MelódicoVIII* el primer pico y la inflexión constituyen el mismo segmento tonal, y desaparece toda declinación:



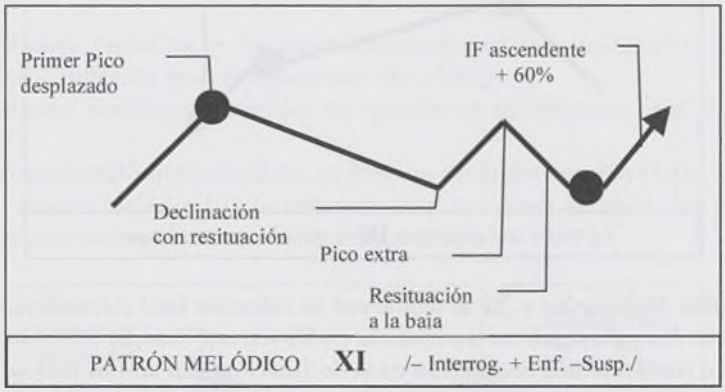
El *Patrón Melódico IX* es similar al patrón *I*, pero con un gran descenso en la inflexión final, superior al 30%:



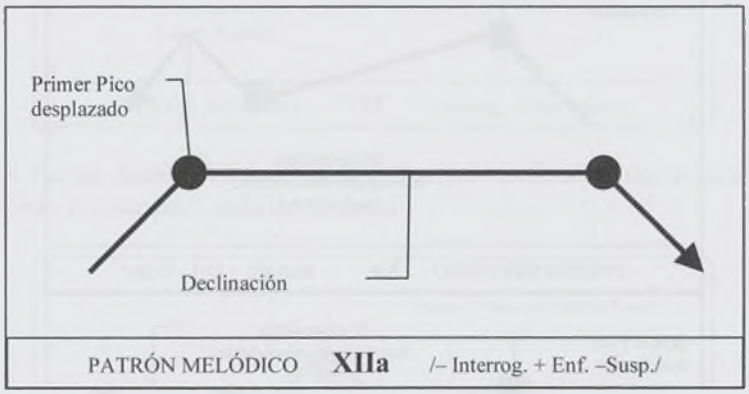
El *Patrón Melódico Xa* y *Xb* se define por su inflexión final circunfleja, ascendente-descendente en *Xa* y descendente-ascendente en *Xb* (en este caso, la inflexión comienza al alza, de nuevo mediante una resituación en la declinación). En ambos patrones, el primer pico desplazado es un rasgo conjunto a la inflexión final circunfleja, como vimos en *IVa*:



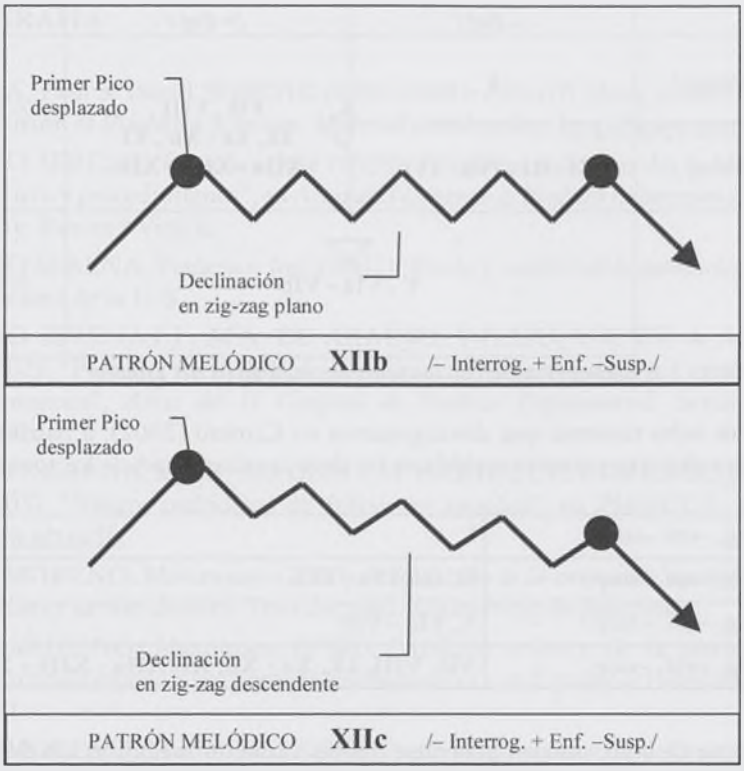
El *Patrón Melódico XI* es similar a los patrones *II* o *III*, con el primer pico desplazado a la vocal átona, pero con un gran ascenso, superior a un +60%, precedido por una declinación quebrada en la que aparece un pico interior extra (como si se tratase de un énfasis de palabra), todo lo cual permite marcar el énfasis, pero con un valor fonológico /– interrog./:



El *Patrón Melódico XIIa* se define por un primer pico desplazado y por una declinación plana, sin pendiente, cuyo efecto de énfasis es muy llamativo, por contraste:



Los *Patrones Melódicos XIIb* y *XIIc*, por su parte, se caracterizan, junto al primer pico desplazado, por una declinación en *zig-zag* (o en *dientes de sierra*) plana (*XIIb*) o descendente (*XIIc*), y constituyen el modelo más característico de las entonaciones enfáticas, como si se tratara de una sucesión de énfasis marcando cada palabra:



6. CONCLUSIÓN

En resumen, en nuestro corpus hemos encontrado un total de 12 patrones melódicos claramente identificables, con diversas variantes:

- un patrón melódico de entonación neutra (*I*)
- tres patrones melódicos de entonación interrogativa (*II, III, IVa-b*)
- dos patrones melódicos de entonación suspendida (*V, VI a-b*)
- seis patrones melódicos de entonación enfática (*VII, VIII, IX, Xa-b, XI, XII a-b-c*)

En el siguiente cuadro se resumen los patrones melódicos, con los rasgos fonológicos de cada uno:

<table>
<tr><td></td><td>/– Enf./</td><td>/+ Enf./</td><td></td></tr>
<tr><td>/– Interrog./</td><td>I</td><td rowspan="2">VII , VIII , IX, Xa - Xb , XI XIIa - XIIb - XIIc</td><td rowspan="2">/– Susp./</td></tr>
<tr><td>/+ Interrog./</td><td>II , III , IVa - IVb</td></tr>
<tr><td></td><td colspan="2">V , VIa - VIb</td><td>/+ Susp./</td></tr>
</table>

Cuadro–resumen de los patrones melódicos del español

De los ocho tonemas que distinguíamos en Cantero (2002), a partir de nuestro corpus hemos descrito patrones melódicos (es decir, *variantes tipo*) de los tonemas:

8. /–interrog. –enf. –susp./	**I**
4. /+interrog. –enf. –susp./	**II, III, IVa - IVb**
7. /–interrog. –enf. +susp./	**V, VIa - VIb**
6. /–interrog. +enf. –susp./	**VII, VIII, IX, Xa - Xb, XI, XIIa - XIIb - XIIc**

A partir de aquí, pueden generarse nuevas variantes melódicas añadiendo énfasis (para pasar de /–enf./ a /+enf./) o eliminando su inflexión final (para pasar de /–susp./ a /+susp./), hasta generar variantes de todos los tonemas del español:

/–enf./ ! /+enf./	
8. /–interrog. –enf. –susp./	6. /–interrog. +enf. –susp./
4. /+interrog. –enf. –susp./	2. /+interrog. +enf. –susp./
7. /–interrog. –enf. +susp./	5. /–interrog. +enf. +susp./
/–susp./ ! /+susp./	
8. /–interrog. –enf. –susp./	7. /–interrog. –enf. +susp./
4. /+interrog. –enf. –susp./	3. /+interrog. –enf. +susp./
6. /–interrog. +enf. –susp./	5. /–interrog. +enf. +susp./
2. /+interrog. +enf. –susp./	1. /+interrog. +enf. +susp./

BIBLIOGRAFÍA

BOERSMA, Paul & David WEENIK (1992-2005): *PRAAT. Doing phonetics by computer*. Institute of Phonetic Sciences, Univ. of Amsterdam. http://www.praat.org

CANTERO SERENA, Francisco José (1999): "Análisis melódico del habla: principios teóricos y procedimiento", en *Actas del I Congreso de Fonética Experimental*. Tarragona: Univ. Rovira i Virgili.

CANTERO SERENA, Francisco José (2002): *Teoría y análisis de la entonación*. Barcelona: Edicions de la U.B.

CANTERO SERENA,F.J.; Mª A. DE ARAÚJO; Y-H.LIU; Y-K.WU & A.ZANATTA (2002): "Patrones melódicos de la entonación interrogativa del español en habla espontánea", *Actas del II Congreso de Fonética Experimental*. Sevilla: Univ. de Sevilla.

CANTERO SERENA, F.J.; R. ALFONSO; M. BARTOLÍ; A. CORRALES & M. VIDAL (2005): "Rasgos melódicos de énfasis en español", en *PHONICA, vol. 1* http:/ www.ub.es/lfa

CORTÉS MORENO, Maximiano (2000): *Adquisición de la entonación española por parte de hablantes nativos de chino*. Tesis doctoral. Universitat de Barcelona.

CORTÉS MORENO, Maximiano (2004): "Análisis acústico de la producción de la entonación española por parte de sinohablantes", en *Estudios de Fonética Experimental. XIII.*

CORTÉS MORENO, Maximiano (2005): "Análisis experimental del aprendizaje de la acentuación y la entonación españolas por parte de hablantes nativos de chino", en *PHONICA. vol. 1* http:/www.ub.es/lfa

FONT ROTCHÉS, Dolors (2005): *L'entonació del català*. Tesis doctoral. Laboratori de Fonètica Aplicada. Universitat de Barcelona.

FONT ROTCHÉS, Dolors (2005b): "Patrones entonativos del catalán en habla espontánea", en este mismo volumen.

FONT ROTCHÉS, D.; A. CANALS; G. ESTER; A. HERMOSO & F.J. CANTERO (2002): "Patrones melódicos de la entonación interrogativa del catalán en habla espontánea", en *Actas del II Congreso de Fonética Experimental*. Sevilla: Univ. de Sevilla.

LIU, Yen-Hui (2003): *La entonación del español hablado por taiwaneses*. Tesis doctoral. Versión digital (2005) en la *Biblioteca Phonica* http:/www.ub.es/lfa

LIU, Yen-Hui & Francisco José CANTERO (2003): "La entonación prelingüística del español hablado por taiwaneses: establecimiento de un corpus", en *Actas del II Congreso de Fonética Experimental*. Sevilla: Univ. de Sevilla.

NAVARRO TOMÁS, T. (1944): *Manual de entonación española*. New York: Hispanic Society. (1974[4]) Madrid: Guadarrama.

TORREGROSA, José (1999): "Correlación de patrones entonativos y kinésicos: análisis de un debate televisado", en *Actas del I Congreso de Fonética Experimental*. Tarragona: Univ. Rovira i Virgili.

RELAÇÃO ACENTO PROSÓDIA NO CORPUS AMPER-POR

Rosa Lídia Coimbra, António Teixeira, Mário Martins Pereira, Lurdes De Castro Moutinho
Universidade de Aveiro

1. INTRODUÇÃO

A presente pesquisa, baseada numa pequena amostragem do corpus AMPER-POR, enquadra-se no projecto AMPER (Atlas Multimédia Prosodique des Langues Romanes), seguindo a metodologia comum às outras equipas que nele participam e de acordo com as directrizes definidas pela coordenação geral, sediada na Universidade de Grenoble (Contini et al., 2003).

O objectivo deste estudo é o de avaliar, na prosódia, o efeito das diferentes estruturas do acento lexical (oxítono, paroxítono e proparoxítono) ocorrendo em diferentes posições frásicas, para dois tipos de frase: declarativa e interrogativa global. Nesse sentido, seleccionámos, do nosso corpus, seis informantes masculinos, sendo cada um deles de um ponto geográfico diferente de Portugal continental.

2. O CORPUS E METODOLOGIA DE ANÁLISE

O corpus analisado é constituído por um total de 144 enunciados, incluindo, para cada informante, 3 repetições de cada uma das seguintes frases:

Kwpa – O capataz toca no pássaro.	Pwka – O pássaro toca no capataz.
Kwpi – O capataz toca no pássaro?	Pwki – O pássaro toca no capataz?
Pwta – O pássaro toca no Toneca.	Twpa – O Toneca toca no pássaro.
Pwti – O pássaro toca no Toneca?	Twpi – O Toneca toca no pássaro?

Os informantes seleccionados são todos do sexo masculino, com a escolaridade básica, cobrindo, uma faixa etária entre os 45 e os 65 anos, e oriundos das seguintes proveniências geográficas: Beira Litoral, Alentejo, Algarve, Beira Interior, Minho, Trás-os-Montes.

O sinal acústico proveniente das gravações efectuadas foi digitalizado, segmentado no programa CoolEdit e tratado em MatLab de acordo com a metodologia e programação informática desenvolvidas para o efeito pela coordenação geral do projecto.

3. ALGUNS RESULTADOS

Feita a análise de sinal e elaborados os gráficos de Fo no programa Excell, procedemos à análise dos resultados obtidos. Com essa finalidade, agrupáram-se as representações gráficas dos primeiros e dos últimos grupos tonais nas duas modalidades – declarativa e interrogativa – e acentuações lexicais possíveis para o Português.

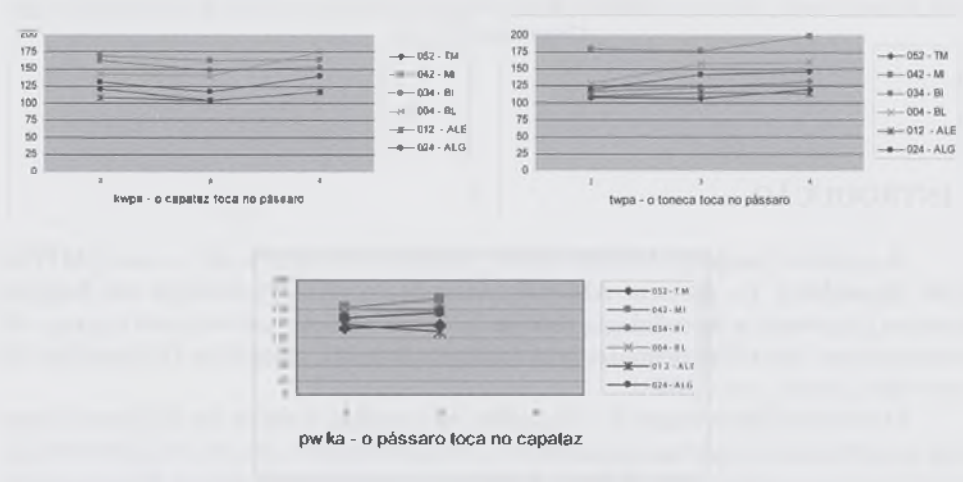


Figura 1 – O primeiro grupo tonal das frases declarativas nas três acentuações lexicais

Na figura 1 acima, podemos observar o contorno melódico do primeiro grupo tonal nas frases declarativas selecionadas, ou seja, o grupo nominal cujo núcleo é ocupado, respectivamente por: "capataz" (oxítona), "Toneca" (paroxítona) e "pássaro" (proparoxítona). Devemos, no entanto, salientar que, no último caso, a última vogal não obteve, em nenhum dos informantes, uma realização fonética. A elisão das vogais átonas em final de frases é, de facto, uma característica comum e prevalecente no Português Europeu, o que, por vezes, dificulta uma observação clara do que acontece no movimento de Fo, sobretudo quando tal ocorre em final de frase.

Genericamente, podemos afirmar que estas frases declarativas apresentam contornos prosódicos semelhantes para as três acentuações, com uma tendência para uma configuração côncava ligeiramente ascendente. Esta semelhança de contornos permite-nos concluir que os diferentes acentos lexicais não parecem condicionar a configuração da curva melódica, quando tal ocorre no primeiro grupo tonal desta modalidade.

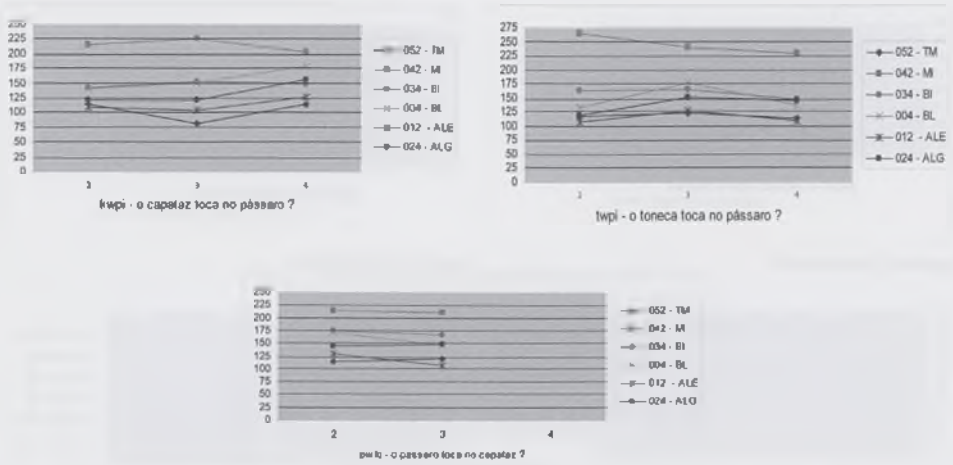
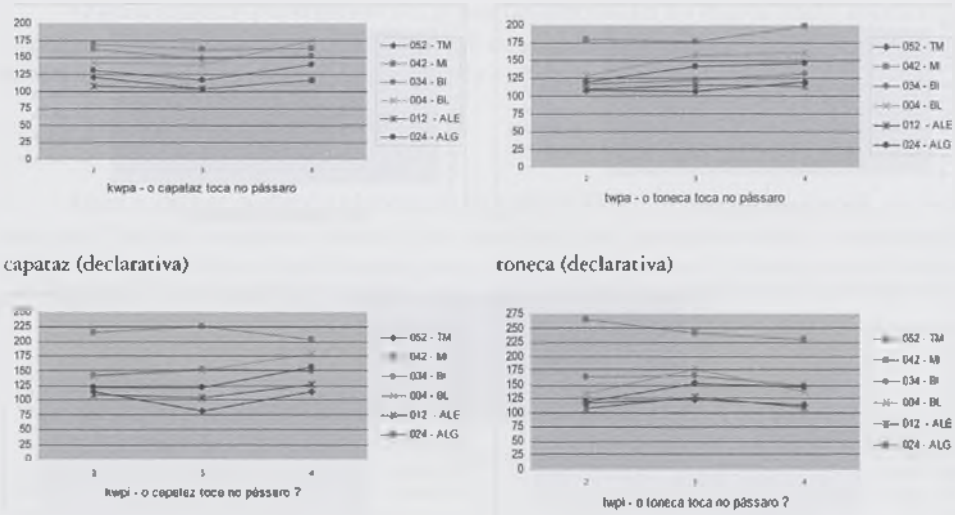


Figura 2 - O primeiro grupo tonal das frases interrogativas nas três acentuações lexicais

Na figura 2, podemos observar que, ao contrário do que acontece com as declarativas, nas frases interrogativas a mudança do acento lexical provoca uma ligeira alteração no movimento da curva melódica. Mantém-se a elisão da última vogal na acentuação de "pássaro", tal como na declarativa, mas no caso de "toneca" e "capataz" vamos distinguir um movimento ligeiramente convexo no primeiro caso e côncavo no segundo. No entanto, todas estas configurações se apresentam como tendencialmente muito suaves, sem grandes picos e oscilações na curva de Fo.

Para além destas observações com base nas acentuações lexicais, interessou-nos também agrupar os gráficos segundo as modalidades. Apenas apresentamos gráficos para as acentuações oxítona e paroxítona, por serem apenas estas que mantêm todas as vogais, como se apresenta na figura 3.



capataz (declarativa) toneca (declarativa)

capataz (interrogativa) toneca (interrogativa)

Figura 3 – Primeiro grupo tonal: comparação dos dois tipos de frase

Observando a figura 3, podemos concluir que a informação do primeiro grupo tonal não permite estabelecer uma distinção clara entre os dois tipos de frase, principalmente no caso da acentuação oxítona, em que os movimentos de Fo são muito semelhantes e situados dentro do mesmo intervalo de frequência.

Quanto ao segundo grupo tonal, analisámos, igualmente, as realizações do lexema distintivo das três acentuações possíveis, mas, neste caso, contemplamos, nos gráficos, as três medições que, segundo a metodologia AMPER, são realizadas em cada vogal. Esta opção prende-se com o facto de que neste grupo tonal a elisão da última vogal torna difícil, sem esta informação mais detalhada, a observação do contorno final dos enunciados, fundamental, como veremos, para a distinção entre os dois tipos de frase analisados.

Observando a figura 4, constituída pelos gráficos relativos aos enunciados de tipo declarativo nas três acentuações, podemos constatar um movimento global descendente em todos os casos e uma tendência crescente para a elisão das vogais finais. Quer dizer: enquanto que nenhuma das vogais é elidida no caso da oxítona, já que termina em vogal acentuada, verificamos bastantes elisões na última vogal da paroxítona e uma tendência geral para esta elisão no caso da proparoxítona. Apesar destas diferenças, o movimento descendente é visível em todos os casos, tal como acontecia com o primeiro grupo tonal.

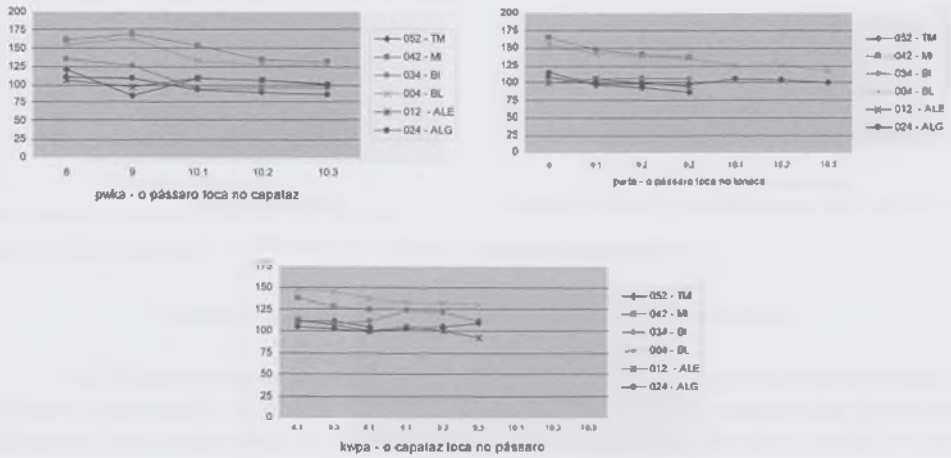


Figura 4 - O segundo grupo tonal das frases declarativas nas três acentuações lexicais

Esta tendência, verificada em todas as variedades regionais para o contorno final descendente, é típica e comum às declarativas do Português por nós já estudadas (cf. por exemplo, Moutinho et al., 2005) e constitui a principal característica prosódica que permite distingui-las das frases interrogativas globais correspondentes.

Na figura 5, construímos o mesmo tipo de comparação para os enunciados interrogativos. Nestes casos, e comparando com a figura 4, é nítida uma muito menor tendência para a elisão das vogais finais, possivelmente devido à intenção do falante em marcar prosodicamente um movimento final tendencialmente ascendente, característico do movimento prosódico dos enunciados interrogativos globais em Português.

Em todas as regiões e para as três acentuações, notamos, no caso dos enunciados interrogativos globais, uma configuração ascendente em final de frase, embora com contornos melódicos distintos, de acordo com as acentuações lexicais. Este movimento não é tão nítido nos casos em que existem vogais átonas em posição final, caso em que Fo sofre uma ligeira queda. No entanto, as vogais pré-tónica e tónica apresentam, em todos os casos, o contorno ascendente que torna prosodicamente perceptível a realização deste tipo de enunciados.

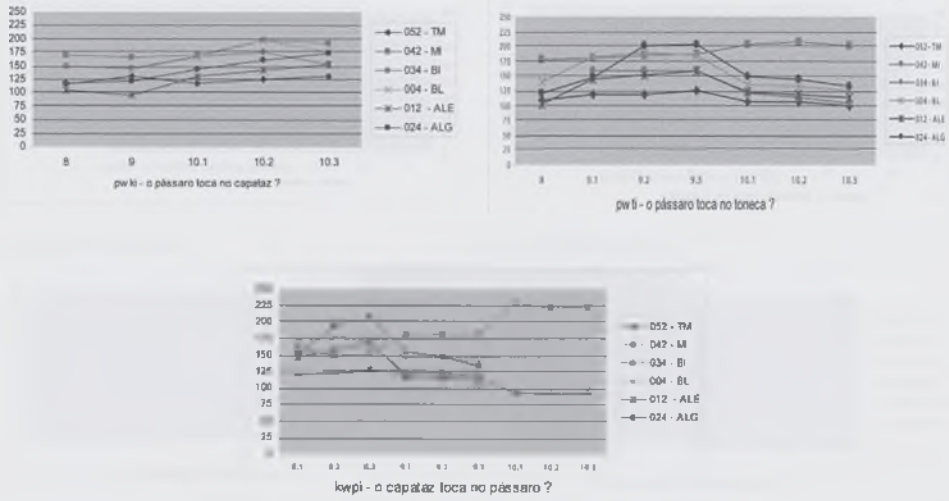
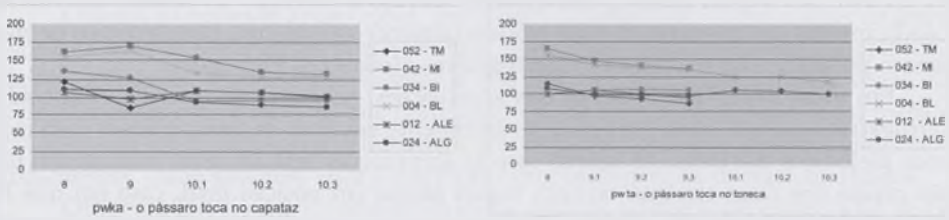
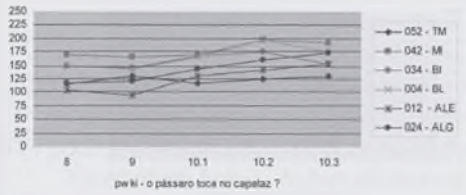


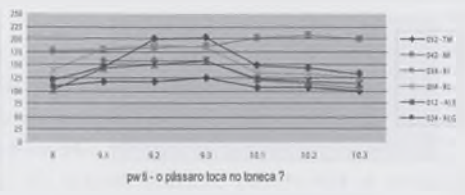
Figura 5 - O segundo grupo tonal das frases interrogativas nas três acentuações lexicais

À semelhança do que fizemos para as frases declarativas, interessou-nos também aqui uma comparação entre as realizações do último grupo tonal nos dois tipos de frase, obtendo a figura 6. Por uma questão de coerência, mantemos as acentuações lexicais escolhidas anteriomente: oxítona e paroxítona.



capataz (declarativa) toneca (declarativa)





capataz (interrogativa) toneca (interrogativa)

Figura 6 - Segundo grupo tonal: comparação dos dois tipos de frase

Esta figura leva-nos à conclusão de que, ao contrário do que observámos para o primeiro grupo tonal, no último tonal as realizações prosódicas dos enunciados permitem dsitinguir claramente entre os tipos declarativo e interrogativo. De facto, tanto no caso da oxítona, como no da paroxítona, verificamos um movimento cruzado entre as curvas correspondentes, sendo nítida a descida na declarativa, por oposição a uma subida na interrogativa e uma finalização com valores de Fo mais elevados neste último caso em relação aos valores finais das declarativas correspondentes. Será, portanto, este grupo tonal o responsável pela distinção prosódica entre os dois tipos de frase.

Verificando-se esta situação, quer para uma, quer para outra das acentuações comparadas, podemos concluir que a mudança de lugar dos diferentes acentos lexicais na frase parece não interferir grandemente no acento frásico da mesma.

4. CONTORNO GLOBAL DE F0

Para além da análise acima apresentada, objecto principal do presente estudo, decidimos também incluir os movimentos globais de Fo, apenas para a oxítona em final de frase, por ser esta a frase em que, embora se constate a elisão de vogais, como é habitual no PE, essa elisão coincide, para todos os falantes, nas mesmas vogais, tornando-se, por isso, mais adequada a realização de uma média global para todas as regiões analisadas.

Assim, apresentamos na figura 7 os resultados que nos permitem comparar as modalidades em análise (declarativa e interrogativa) da frase -"O pássaro toca no capataz" e na figura 8 os dados relativos à fras- "O pássaro toca no Toneca", onde se pode constatar a existência de variação regional, mais acentuada nuns casos do que noutros. Apesar disso, a média obtida para cada uma das modalidades para todas as regiões analisadas é consentâneo com o tradicionalmente descrito para o Português Europeu.

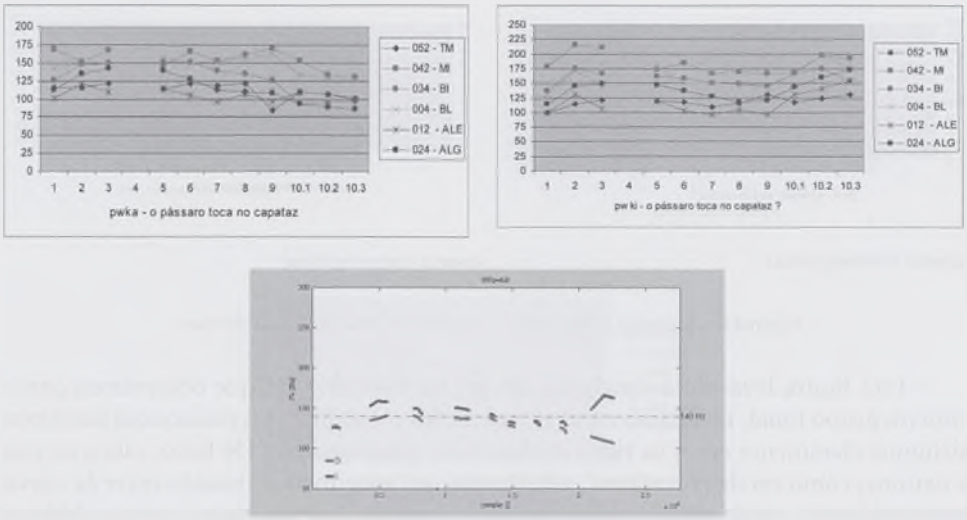


Figura 7 - O movimento global de Fo, para as seis regiões analisadas, nas duas modalidades, acnetuação final oxítona

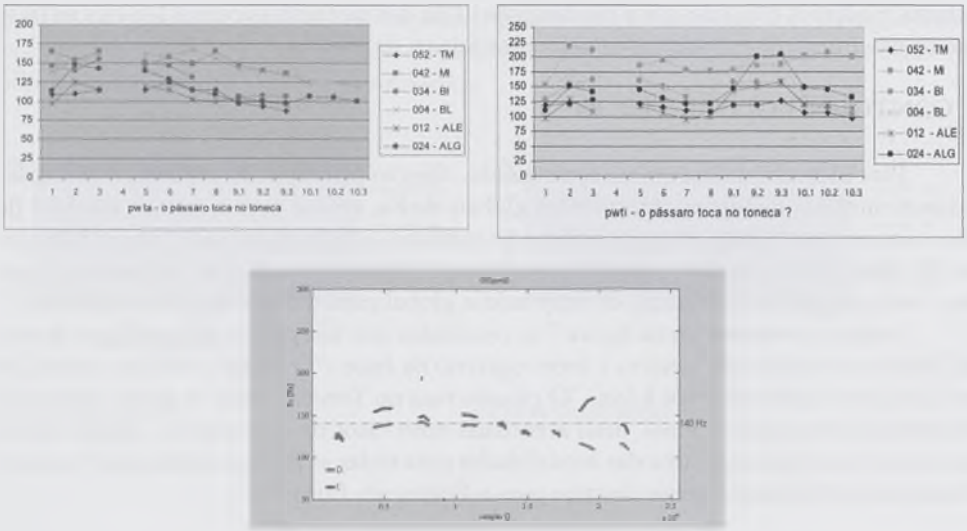


Figura 8 – O movimento global de Fo, para as seis regiões analisadas, nas duas modalidades, acentuação final paroxítona

5. CONCLUSÃO

De acordo com a pesquisa realizada, verificamos que a mudança de lugar dos diferentes acentos lexicais na frase parece não interferir grandemente no acento frásico.

No entanto, verifica-se que a mudança de acento lexical no segundo grupo tonal (SV) pode eventualmente alterar o movimento de Fo.

Este grupo tonal, independentemente do acento frásico, parece ser o principal responsável pela distinção entre as modalidades declarativa e interrogativa, com movimentos finais de Fo descendentes no primeiro caso e ascendentes no segundo.

Apesar desta tendência geral, consentânea com o já referenciado noutros estudos para o PE (Pereira et al., 1992; Cruz-Ferreira 1998), verificamos que, em todos os casos estudados, existem diferenças mais ou menos notórias entre os informantes, o que pode ser indiciador de uma variação regional em termos prosódicos.

BIBLIOGRAFIA

CONTINI M., *et al*, 2002, "Un Projet d'Atlas Multimédia Prosodique de l'Espace Roman", *Proc. of the Int. Conf. Speech Prosody 2002*, Aix-en-Provence, 227-230.

CONTINI M., *et al*, 2003, « Vers un Atlas Prosodique Parlant des Variétés Romanes ». In. J.C. Bouvier et al. (eds) Mélanges offerts à X. Ravier, CNRS. Univ. Toulouse Le Mirail, 73-84.

CRUZ-FERREIRA, M., 1998, « Intonation in European Portuguese », in Hirst D. & Di Cristo A. (eds), *Intonation Systems. A Survey of Twenty Languages*, Cambridge University Press, 167-178.

MOUTINHO L.C., *et al*, 2001, "Project d'Atlas Prosodique Multimédia des Variétés Romanes", in *Travaux de L'Institut de Phonetique de Strasbourg (TIPS)*, 31, 61-70.

MOUTINHO L.C., *et al*, 2005, "Variação Entoacional em Três Áreas Dialectais de Portugal Continental", *Geolinguistique*, Hors série, 3: 19-38.

MOUTINHO L.C., *et al*, (no prelo), "Variação entoacional inter e intra regional no Português europeu", in Actas do II Congresso Int. de Fonética e Fonologia, UFM.

PEREIRA, I. et al., 1992, *Estudos em Prosódia*, Lisboa: Colibri, 33-73.

PRIMERAS APORTACIONES A LA DESCRIPCIÓN PROSÓDICA DEL EXTREMEÑO: INTERROGATIVAS ABSOLUTAS SIN EXPANSIÓN EN EL HABLA DE DON BENITO (BADAJOZ)

YOLANDA CONGOSTO MARTÍN
(Universidad de Sevilla)
Responsable del proyecto AMPER en Andalucía Occidental y Extremadura

1. INTRODUCCIÓN

> *Si bien es verdad que los estudios descriptivos sobre las particularidades entonativas del español en su complejidad panhispánica han adquirido en las últimas décadas un fuerte impulso (son especialmente significativas a este respecto las aportaciones realizadas por AMPER) también es cierto que aún quedan zonas prácticamente sin investigar desde el punto de vista experimental; es el caso de las hablas extremeñas. Si consultamos los distintos repertorios bibliográficos existentes a propósito de la caracterización dialectal de este ámbito lingüístico comprobaremos que poco más hay sobre sus peculiaridades entonativas -de especial riqueza e interés- que el tradicional estudio realizado por María Josefa Canellada en 1941*[1].

Como responsable y coordinadora del proyecto AMPER en Extremadura es mi intención realizar un estudio sistemático, minucioso y detallado de la prosodia del extremeño en su variedad dialectal[2]; los primeros pasos de este apasionante camino comienzan, así pues, hoy en Don Benito (Badajoz), en un intento de definir en sí misma y por comparación la estructura y caracterización fonética del contorno entonativo de los enunciados interrogativos no pronominales dombenitenses a través, por un lado, del análisis de sus rasgos melódicos, y, por otro, de la comparación de los mismos con los propios de

1 «Notas de entonación extremeña», *RFE*, XXV, 1941, 79-91.

2 Para un conocimiento más detallado sobre el subproyecto AMPER en Extremadura, véase el artículo Y. Congosto «Proyecto de un Atlas prosódico de Extremadura» *Anuario de Estudios Filológicos*, 29, 2006 (en prensa).

otro ámbito dialectal, en concreto, con los de Sevilla capital. Así mismo, analizaremos el comportamiento de los otros dos parámetros fundamentales que definen la entonación, el temporal o duración de las vocales y su intensidad, completando así, de este modo, el estudio entonativo que nos hemos propuesto.

Para este primer acercamiento hemos seleccionado, de acuerdo con los objetivos marcados por AMPER–España, un informante femenino representativo del punto encuestado, Ana de Paz Coronado, 25 años, natural de Don Benito.

El corpus analizado está compuesto por 9 (3 x 3) frases interrogativas absolutas sin expansión del tipo *SVO*, con 11 sílabas cada una, y que presentan una estructura acentual oxítona en el SN1, paroxítona en el SV, y las tres posibilidades acentuales: oxítona, paroxítona y proparoxítona en el SN2. Las frases seleccionadas en cuestión son: *el saxofón se toca con obsesión*, *el saxofón se toca con paciencia* y *el saxofón se toca con pánico*.

En cuanto a la metodología seguida para el análisis acústico, ésta responde a la establecida por AMPER.

2. ANÁLISIS ACÚSTICO

2.1. Curvas melódicas de los enunciados interrogativos no pronominales dombenitenses

2.1.1. Estructura y caracterización fonética del contorno. Descripción general.

El análisis acústico de los principales rasgos melódicos (rasgos fonéticos de la entonación: altura relativa del primer pico, declinación e inflexión final) de las curvas medias resultantes de los tres enunciados analizados presenta los siguientes resultados:

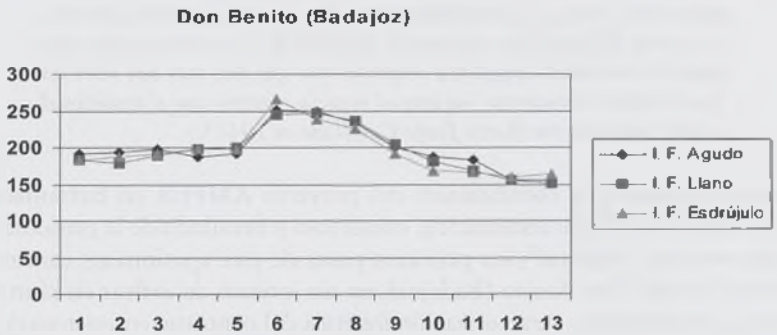


Figura 1. Medias de F0 (en Hz)[3]

3 Recordamos que las curvas han sido obtenidas con los valores centrales de cada vocal (del 2 al 12). El valor 1 y 13 corresponden al valor inicial de la primera vocal y al valor final de la última vocal, respectivamente; y lo mismo cabe decir para las tablas.

Altura relativa del primer pico

- Enunciados con final agudo: 251 Hz (pretónica del segundo acento tonal).
- Enunciados con final llano: 244 Hz (pretónica del segundo acento tonal).
- Enunciados con final esdrújulo: 266 Hz (pretónica del segundo acento tonal).

Declinación

Se inicia en el primer pico, con excepción de la estructura con final llano que la inicia en la siguiente vocal, la tónica del segundo acento tonal, produciéndose a partir de él un descenso progresivo del tono que oscila entre los 95 y los 103Hz.

Inflexión final

Descendente (tonema final en cadencia).

Campo tonal

	1	2	3	4	5	6	7	8	9	10	11	12	13
F.A	192	193	197	188	192	251	249	233	200	188	184	155	156
F.LL	183	178	189	197	201	244	247	236	204	181	169	156	151
F.E	184	188	192	198	197	266	238	225	191	169	166	160	163

Tabla 1. Valores medios de las vocales [en Hz]

Estructura del contorno entonativo

Anacrusis:	192Hz- 193Hz- 197Hz- 188Hz- 192Hz (F.A.)
	183Hz- 178Hz- 189Hz- 197Hz- 201Hz (F.LL.)
	184Hz- 188Hz- 192Hz- 198Hz- 197Hz (F.E.)
Primer Pico:	251Hz (F.A.)
	244Hz (F.LL.)
	266Hz (F.E.)
Cuerpo:	249Hz- 233Hz - 200Hz -188 Hz -184 Hz (F.A.)
	247Hz- 236Hz - 204Hz -181 Hz (F.LL.)
	238Hz- 225Hz -191Hz (F.E.)
Núcleo o Infl. final[4]:	155Hz- 156Hz (F.A.)
	169Hz- 156Hz -151Hz (F.LL.)
	169Hz- 166Hz -160Hz -163Hz (F.E.)

Contorno melódico terminal

- Enunciados con final agudo: L* L%
- Enunciados con final llano: L* L%
- Enunciados con final esdrújulo: L* L%

4 En los tres finales ofrecemos el valor medio y final de la última vocal.

Los valores expuestos nos muestran cómo los tres contornos comienzan con un tono de juntura inicial bajo, lo normal en español, para emprender con posterioridad una subida muy suave hasta la primera vocal tónica, produciéndose a partir de ella una inflexión ascendente[5] cuyo segundo segmento tonal, una vocal átona (pretónica del segundo acento tonal), constituye el primer pico tonal de la melodía, iniciándose a continuación un declive más o menos empinado (en función de la composición del tercer acento tonal)[6]; es decir, después del primer pico se produce un escalonamiento descendente del tono con una inflexión final en cadencia con valores muy por debajo de los iniciales.

Todo lo hasta aquí revelado nos permite definir, en principio y hasta nuevas investigaciones, el patrón entonativo de los enunciados interrogativos no pronominales dombenitenses y, por extensión, algunas de las características melódicas del habla de Badajoz, que quedaría representada por el siguiente contorno melódico:

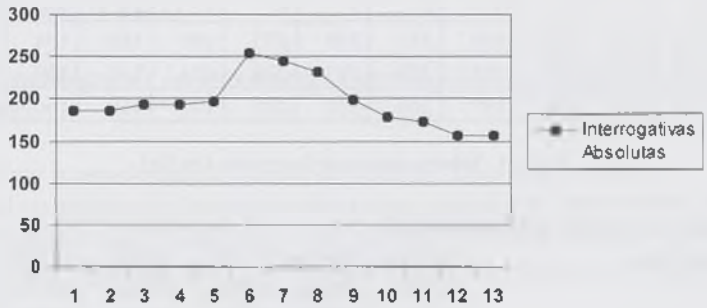


Figura 2. Medias generales de F0 (en Hz).

- *Altura del primer pico*: 254 Hz; situado en la pretónica del segundo acento tonal.
- *Declinación*: Se inicia en el primer pico, produciéndose a partir de él un movimiento descendente progresivo del contorno.
- *Inflexión tonal*: descendente (tonema final en cadencia).

5 Si bien, en principio, el primer pico suele ser la primera vocal tónica del contorno, también en ocasiones constituye el primer segmento de una inflexión ascendente (Cantero, 2002: 158).

6 Como anunciamos con anterioridad, sólo el contorno entonativo resultante de los tres enunciados llanos presenta una pequeña diferenciación con respecto a los otros contornos ya que la curva sigue ascendiendo ligeramente hasta la tónica del segundo acento tonal, para inicial a partir de ella la declinación. Entendemos que no debe ser un rasgo de carácter generalizador sino casual, aunque habrá que cerciorarse.

- Campo tonal

	1	2	3	4	5	6	7	8	9	10	11	12	13
B	186	186	193	194	197	254	245	231	198	179	173	157	157

Tabla 2. Valores medios generales de cada vocal [en Hz].

2.1.2. Comparación interdialectal entre Sevilla (capital) y Don Benito (Badajoz)

Si comparamos los resultados hasta ahora recabados con los obtenidos en el análisis entonativo del habla urbana de Sevilla[7] (véase figura 3) podremos observar diferencias relativas tanto al *rango de frecuencias* -más extremo en subidas y descensos en el habla sevillana-, como a la *configuración de la curva* resultante; diferencias especialmente significativas en lo que afectan al núcleo o inflexión final:

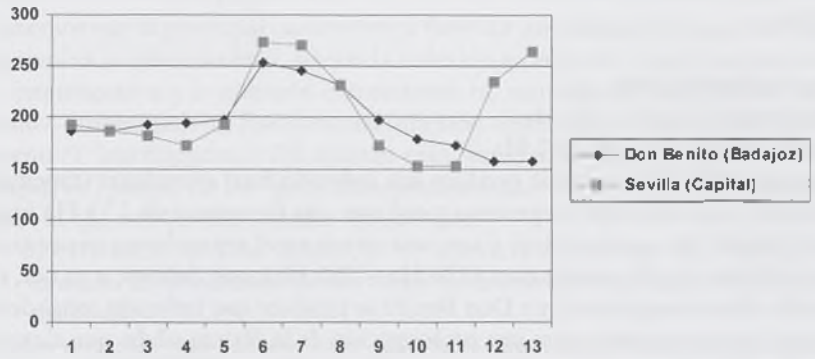


Figura 3. Medias generales de F0 de los enunciados interrogativos absolutos sin expansión [en Hz]

Anacrusis

- *Don Benito*: 186- 193- 194-197 Hz
- *Sevilla capital*: 192-186-182-172-193 Hz

Mientras que el contorno melódico de Don Benito se inicia con un ascenso suave y progresivo hasta la primera sílaba tónica (186-197), el de Sevilla capital comienza con un descenso también progresivo hasta la pretónica desde donde se inicia un ascenso hasta la tónica, alcanzando prácticamente el valor inicial (192-193). Ambas curvas comienzan con un tono de juntura inicial ligeramente por debajo de la media.

7 Datos tomados de Y. Congosto 2005: 225-246.

Primer Pico

- *Don Benito*: 254 Hz
- *Sevilla Capital*: 274 Hz

En ambas curvas el pico máximo está situado en la pretónica del segundo acento tonal, siendo el rango de frecuencia mayor en la curva sevillana que en la dombenitense.

Cuerpo

- *Don Benito*: 245- 231- 198- 179 Hz
- *Sevilla Capital*: 271- 231- 172- 153 Hz

En el contorno melódico de la informante sevillana la declinación comienza no en el primer pico, sino en el segmento siguiente, en este caso, la vocal tónica del segundo acento tonal. Por el contrario, en el contorno dombenitense la declinación se inicia justamente en el primer pico, produciéndose a partir de él un movimiento descendente progresivo del contorno. A ello hemos de añadir las diferencias existentes en cuanto al rango de frecuencias: más suave en Don Benito (245-179) y con mayor pendiente en Sevilla capital (271-153).

Núcleo o inflexión final

- *Don Benito*: 173-157-157 Hz
- *Sevilla capital*: 153-235-265 Hz

Mientras que en Sevilla capital se produce una inflexión final ascendente (tonema final en anticadencia), con un primer segmento tonal con una frecuencia de 153 Hz (que define el límite inferior del campo tonal) y con una altura en el segundo segmento tonal ligeramente inferior a la del primer pico (274 Hz – 265 Hz), que definen, a su vez, el límite superior de dicho campo tonal, en Don Benito se produce una inflexión tonal descendente (tonema final en cadencia) con una prolongación de la última sílaba que alarga la línea del tono horizontalmente, y que le aporta ese "suave dejo tendido" que les caracteriza.

Campo tonal

	1	2	3	4	5	6	7	8	9	10	11	12	13
DB	186	186	193	194	197	254	245	231	198	179	173	157	157
SC	192	186	182	172	193	274	271	231	172	153	153	235	265

Tabla 3. Valores medios generales de las vocales [en Hz]

2.1.3. Primeras conclusiones

El estudio reflexivo de los resultados obtenidos del análisis nos permite llegar a las siguientes conclusiones:

1. Ambas curvas comienzan con la melodía más frecuente en los contornos /+ interrogativos/, aquella en la que la primera vocal tónica (en este caso "xón") es el

primer segmento de una inflexión ascendente cuyo segundo segmento (en este caso "se") es realmente el primer pico máximo del contorno.

2. La altura relativa del primer pico (254Hz DB – 274Hz SC) condiciona, lógicamente, la pendiente de la declinación de cada curva.
3. En el contorno melódico dombenitense el primer pico es el primer elemento de la declinación mientras que en el contorno sevillano el primer pico se encuentra desligado de la declinación, al iniciarse ésta en la siguiente vocal.
4. Ambas curvas se diferencian radicalmente en el último segmento tónico del grupo, aquel de donde parte la inflexión tonal característica que determina el rasgo /± interrogativo/. En otras palabras, mientras que la curva de Sevilla capital responde al patrón /+ interrogativo/ tanto en su descripción cualitativa como cuantitativa, con una inflexión ascendente que incrementa el tono hasta 265 Hz., la curva de Don Benito responde al patrón /- interrogativo/ con una inflexión neutra, no marcado, que hace descender el tono hasta 157Hz.

Sabemos que la principal característica fonética (melódica) del rasgo fonológico /± interrogativo/ es la inflexión final, siendo la inflexión ascendente la que caracteriza el contorno /+ interrogativo/ y la inflexión descendente, no marcada o neutra, la que caracteriza el contorno /- interrogativo/. Rara vez una inflexión descendente caracteriza en castellano un contorno /+ interrogativo/. En nuestro caso, estaríamos ante un contorno terminal /- interrogativo/, propio de las "preguntas pronominales", pero no de las "interrogativas absolutas," lo que nos llevaría a poder afirmar que:

1. El extremeño de Don Benito, y me atrevería a decir que de Badajoz, ha uniformado la entonación de dos tipos de interrogativas: las pronominales y las absolutas.
2. No sería, así pues, en este caso, sólo la entonación el elemento que marcaría la modalidad oracional pregunta/afirmación.

Por otro lado, y de acuerdo con lo afirmado por F.J. Cantero (2002: 169-170), corroboramos con nuestro estudio que si bien la inflexión final es la principal característica fonética y por tanto melódica que define al rasgo fonológico /± interrogativo/, hay otras características melódicas que también contribuyen a ello: las relacionadas con el primer pico del contorno y el segmento tonal siguiente, ya que en los contornos no marcados, es decir, /- interrogativos/, la declinación comienza como hemos podido ver en el primer pico, mientras que en los contornos marcados, /+ interrogativos/, el pico se encuentra desligado de la declinación.

2.2. Análisis de la duración vocálica

2.2.1. Descripción general

La siguiente tabla (tabla 4) y el siguiente histograma (figura 4) recogen el resultado obtenido del estudio del comportamiento temporal de las vocales en los enunciados

interrogativos absolutos. Como podemos observar, se sigue manteniendo la tendencia generalizada que sostiene que la última vocal de cada sintagma es siempre la de mayor duración; y la última del enunciado la más larga de todas.

	1	2	3	4		5	6	7		8	9	10	11
DB	40	65	44	66		41	38	56		50	59	71	267

Tabla 4. Media general de duración [ms] de las vocales (enunciados interrogativos absolutos dombenitenses)

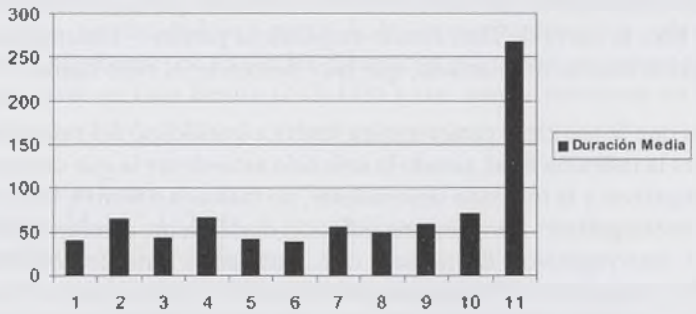


Figura 4. Media general de duración de las vocales [ms] (enunciados interrogativos absolutos dombenitenses)

Lo realmente significativo y caracterizador de las hablas extremeñas y, en concreto, dombenitense es el alargamiento prolongado que sufre la última vocal del enunciado, 267 ms., frente a la duración del resto de las vocales que oscila entre 38 ms. (tónica del segundo acento tonal), la más corta, y 71 ms. (penúltima del enunciado).

Sin embargo, no queremos dejar de llamar igualmente la atención sobre la duración que mantiene la segunda vocal (átona) del primer acento tonal, en este caso, la que corresponde a la primera sílaba de la palabra "saxofón".

Finalmente, constatamos un alargamiento progresivo de las vocales del último sintagma (8,9,10,11), que no se refleja en los dos sintagmas previos.

2.2.2. Comparación interdialectal entre Sevilla (capital) y Don Benito (Badajoz)

Si comparamos la media general de duración obtenida con la de la informante sevillana (figura 5) podremos ver las diferencias establecidas con mayor claridad:

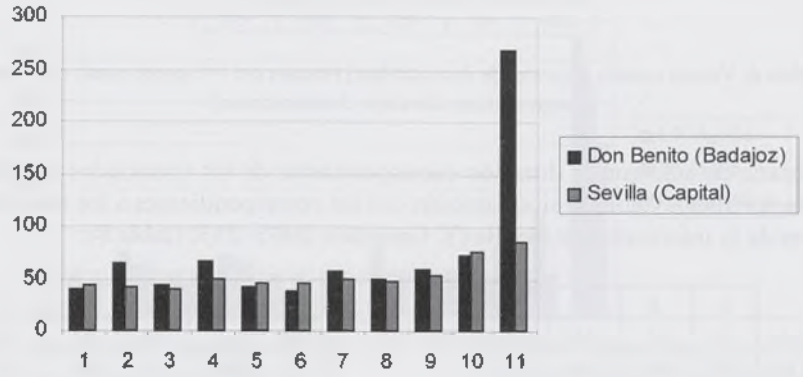


Figura 5. Medias generales de duración de las vocales [ms] (enunciados interrogativos absolutos dombenitenses y sevillanos)

La tabla que aparece a continuación (tabla 5) recoge con detalle los valores medios generales de cada vocal:

	1	2	3	4	5	6	7	8	9	10	11
DB	40	65	44	66	41	38	56	50	59	71	267
SC	43	41	40	50	46	46	50	47	53	76	85

Tabla 5. Valores medios generales de duración de las vocales [ms] (enunciados interrogativos absolutos dombenitenses y sevillanos)

En cuanto a la duración de la última vocal del enunciado, en efecto, frente a las 85 ms. de Sevilla capital están las 267 ms. de Don Benito, lo que supone más del triple del tiempo invertido en su realización. Diferencia temporal que además no sólo existe con la informante sevillana; los datos obtenidos en otros estudios realizados dentro del proyecto AMPER ponen de manifiesto lo excepcional del comportamiento temporal de las interrogativas absolutas sin expansión en las hablas extremeñas. Sin ir más lejos, en las hablas canarias (Dorta *et alt.* 2005: 98) la duración de la última vocal oscila entre 86 ms. y 167 ms. (100 ms. menos) y en las hablas castellano-leonesas (Zamora *et alt.* 2005: 135) la oscilación se mantiene entre 157 ms. y 182 ms. (138 ms. menos).

Del mismo modo, también las últimas vocales de los otros dos sintagmas (la 4ª y la 7ª) son ligeramente más largas en la informante dombenitense que en la sevillana (66-50, 56-50).

En lo que afecta a la segunda vocal del enunciado, hemos de centrar nuestra atención en la duración de todas las vocales del primer acento tonal dombenitense (tabla 6):

	1	2	3	4
DB	40	65	44	66

Tabla 6. Valores medios generales de duración [ms] (vocales del 1er. acento tonal; enunciados interrogativos absolutos dombenitenses)

y compararla no sólo con la duración correspondiente de los enunciados interrogativos absolutos sevillanos (tabla 7) sino también con los correspondientes a los enunciados declarativos de la informante de Sevilla (Y. Congosto 2005: 233) (tabla 8):

		1	2	3	4
DB	I.A	40	65	44	66
SC	I.A.	43	41	40	50

Tabla 7. Valores medios generales de duración [ms] (vocales del 1er. acento tonal; enunciados interrogativos absolutos de DB y SC)

		1	2	3	4
DB	I.A.	40	65	44	66
SC	Decl.	63	69	62	81

Tabla 8. Valores medios generales de duración [ms] (vocales del 1er. Acento tonal enunc. interr. abs. de DB y declar. de SC)

Como puede observarse en las tablas 7 y 8, el comportamiento temporal del primer acento tonal se corresponde más con el declarativo de SC que con el interrogativo de SC, constatando igualmente una mayor duración de esa segunda vocal. Y ello no sería llamativo si no fuera porque es un rasgo constante en nuestra informante dombetinense ya que alarga siempre dicha vocal en todos los enunciados (tabla 9):

	1	2	3	4
F. A	47	74	48	70
F. Ll	31	59	35	66
F. E.	43	63	49	62

Tabla 9. Valores medios de duración [ms] (vocales del 1er. acento tonal; enunciados interrogativos absolutos de DB)

2.2.3. Duración y acento

En el siguiente histograma (figura 6) se representan las medias generales de duración de los tres enunciados resultantes; recordemos que las frases analizadas responden al tipo SVO siendo su estructura acentual siempre oxítona en el SN1, paroxítona en el SV y oxítona, paroxítona y proparoxítona en el SN2, por lo que centraremos nuestra atención especialmente en el tercer acento tonal:

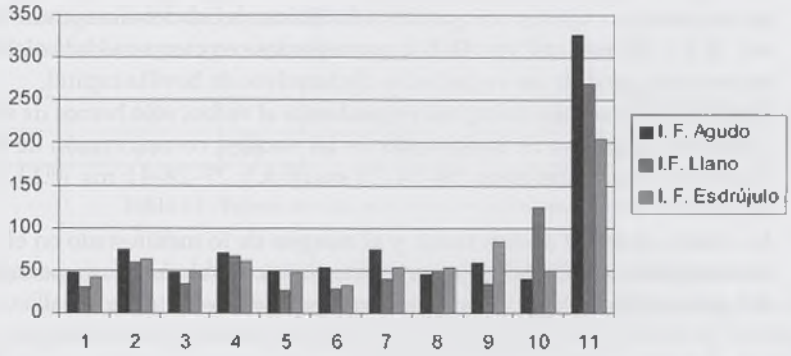


Figura 6. Medias de duración de las vocales [ms]

Los valores exactos son los siguientes (tabla 10):

	1	2	3	4		5	6	7		8	9	10	11
F. A	47	74	48	70		50	54	75		46	59	40	327
F. Ll	31	59	35	66		27	28	41		49	34	124	270
F. E.	43	63	49	62		47	32	54		54	85	50	205

Tabla 10. Valores medios de duración [ms]

Como podemos ver, el final agudo es el que presenta una mayor duración: 327 ms. frente a 270 ms. (F.LL.) y 205 ms. (F.E.), incremento temporal que queda justificado por la tonicidad de la última vocal del sintagma. Circunstancia que igualmente provoca que la tónica del final llano (10ª vocal) y del final esdrújulo (9ª vocal) se alarguen considerablemente con respecto a la misma vocal de los otros sintagmas finales: 124 ms. (F.LL) frente a 40 ms. (F.A.) y 50 ms. (F.E), y 85 ms. (F.E.) frente a 34 ms. (F.LL.) y 59 ms.(F.A.), sin llegar en ninguno de los dos últimos casos a superar en duración a la vocal final del enunciado.

2.2.4. Segundas conclusiones

Lo expuesto anteriormente nos lleva a las siguientes conclusiones:

1. En el habla de Don Benito y, por extensión, de Badajoz la duración de la última vocal de los enunciados interrogativos absolutos experimenta un aumento excepcional, lo que provoca un alargamiento del tono que oscila, en nuestra informante, entre 327 ms. y 205 ms. en función del contorno melódico terminal, lo que le confiere una cadencia entonativa característica.
2. Se confirma que la última vocal de cada sintagma suele ser siempre la de mayor duración y que la última del enunciado es la más larga de todas.

3. En el primer acento tonal destaca la duración de la segunda vocal que supera en ocasiones en valores temporales a la última vocal del sintagma: 74 ms.-70 ms. (F.A.), 63 ms.– 62 ms. (F.E.), asemejándose en cierta medida al comportamiento temporal de los enunciados declarativos de Sevilla capital.
4. En el segundo acento tonal, correspondiente al verbo, sólo hemos de señalar un progresivo aumento de la duración de las vocales, con excepción de los enunciados con final esdrújulos: 50-54-75 ms.(F.A.); 27-28-41 ms. (F.LL.); 47-32-54 ms.(F.E.).
5. En cuanto al tercer acento tonal, y al margen de lo manifestado en el punto 1, resaltar cómo la recaída del acento sintagmático sobre el acento paradigmático del grupo fónico alarga considerablemente la duración de la vocal.

2.3. Análisis de la intensidad vocálica

2.3.1. Descripción general

El histograma que presentamos a continuación (figura 7) recoge la intensidad media general de los nueve enunciados analizados:

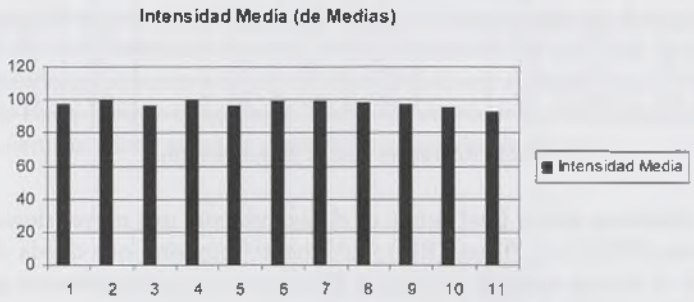


Figura 7. Medias generales de intensidad (en dB)

Como se puede observar, no hay grandes oscilaciones; la intensidad sonora es bastante estable a lo largo de toda la cadena hablada; la fluctuación máxima que llega a alcanzar es de 4 dB entre la tercera y cuarta vocal. Sin embargo, tales variaciones son significativas en el sentido de que se producen, por un lado, como es lógico, al iniciarse el tercer acento tonal, cuando la F0 baja de frecuencia, y de forma bastante paulatina (99-98-97-95-93 dB.). Por otro lado, creemos que también es relevante el hecho de que hasta la sexta vocal la intensidad sea más baja en las vocales impares (1, 3 y 5) y más alta en las pares (2, 4 y 6), donde se estabiliza (6-7) para ya comenzar el descenso final. Recordemos que las vocales 4 y 6 son vocales tónicas, y que la 2, siendo átona, ya presentaba peculiaridades de carácter temporal. Con ello no queremos decir que la intensidad esté, en este

caso, ligada al acento paradigmático o léxico (lo veremos en el siguiente apartado), pero sí que hay cierta uniformidad de comportamiento.

En la tabla 11 se recogen los valores puntuales de intensidad:

	1	2	3	4	5	6	7	8	9	10	11
I.A.	97	100	96	100	96	99	99	98	97	95	93

Tabla 11. Valores medios generales de intensidad (en dB)

2.3.2. Comparación interdialectal entre Sevilla (capital) y Don Benito (Badajoz)

El análisis comparativos de los datos obtenidos de ambas informantes revela, como es lógico, importantes diferencias, ya que los resultados de Don Benito se corresponden con los de un contorno /- interrogativo/ y los de Sevilla capital con los de un contorno /+ interrogativo/, aunque en ambos casos estemos comparando los mismos enunciados interrogativos absolutos.

Por otro lado, y a excepción de lo que sucede en la vocal 2 y 3, la amplitud de la onda sonora es mayor en SC que en DB. Mientras que se en DB los valores oscilan entre 93 y 100 dB, en SC lo hacen entre 95 y 103 dB.

La figura 8 y la tabla 12 que recogen los datos obtenidos del estudio comparativo de los dos ámbitos dialectales nos permiten observar con detalle las diferencias apuntadas con anterioridad:

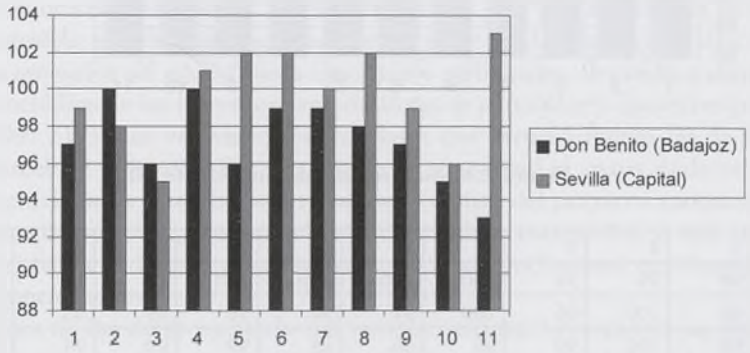


Figura 8. Medias generales de intensidad (en dB)

	1	2	3	4	5	6	7	8	9	10	11
DB	97	100	96	100	96	99	99	98	97	95	93
SC	99	98	95	101	102	102	100	102	99	96	103

Tabla 12. Valores medios generales de intensidad (en dB) de Don Benito y Sevilla capital

Como se puede apreciar, en DB las vocales más intensas son la 2 y la 4, seguida de la 6 y la 7; por el contrario, en SC la vocal más intensa es la 11, última del enunciado, la que recoge la inflexión ascendente, seguida de la 5, 6 y 8. En correspondencia, la vocal menos intensa en DB es, lógicamente, la 11, última del enunciado, que recoge a su vez la inflexión descendente, y en SC la 3 y la 10, penúltima del enunciado, de donde parte la inflexión ascendente.

2.3.3. Intensidad y acento

Nos centramos ahora en el tercer acento tonal que es el que presenta diferencia acentuales en un intento de establecer relaciones entre intensidad y acento, sin olvidar, por supuesto, el resto del enunciado.

La figura 9 y la tabla 13 nos ofrecen en este caso de forma comparativa los valores medios obtenidos de las tres repeticiones de cada frase con final agudo, llano y esdrújulo:

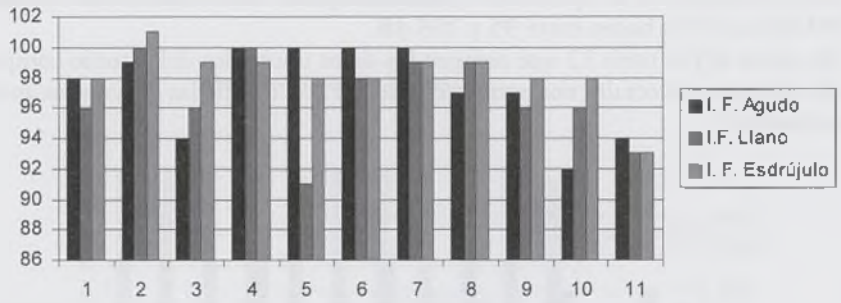


Figura 9. Medias generales de intensidad (en dB)

	1	2	3	4	5	6	7	8	9	10	11
F.A.	98	99	94	100	100	100	100	97	97	92	94
F. LL.	96	100	96	100	91	98	99	99	96	96	93
F. E.	98	101	99	99	98	98	99	99	98	98	93

Tabla 13. Valores medios de intensidad (en dB)

En la estructura con final agudo es interesante ver cómo la intensidad baja en la pretónica, vocal 10, para volver a subir, si bien ligeramente, en la tónica final (vocal 11). Y lo mismo ocurre en el primer acento tonal entre la vocal 3 (pretónica) y 4 (tónica). Por otro lado, es también interesante ver cómo de la vocal 4 a la vocal 7 se mantiene una intensidad estable, sin oscilaciones.

Sin embargo, en el final llano la intensidad disminuye más pausadamente, sin sobresaltos: (99-99), (96-96), (93). Es igualmente significativo advertir cómo nuevamente en la pretónica (en este caso la vocal 5, pretónica del segundo acento tonal, precedida a su vez de una tónica) la intensidad disminuye para volver a subir en la tónica siguiente, siendo la diferencia de intensidad realmente significativa (100-91-98 dB).

En el final esdrújulo se produce una pérdida de intensidad significativa en la última vocal del enunciado (93 dB), la más baja de todas. Por lo demás, la intensidad suele ser bastante estable a lo largo de la frase, sobre todo entre la vocal 3 y 10: (99-99), (98-98), (99-99), (98-98). En este caso, la informante no ha distinguido entre átonas y tónicas.

2.3.4. Terceras conclusiones

Sólo reseñar, por un lado, la estabilidad generalizada que presenta la media general resultante de los 9 enunciados; por otro lado, cómo el acento léxico o paradigmático influye, en cierta medida, en la intensidad; y en tercer lugar, marcar la diferencias existentes entre las interrogativas absolutas de DB y de SC en este sentido.

3. CONCLUSIÓN FINAL

Viene a ser casi un lugar común aceptar que las oraciones interrogativas absolutas tienden a terminar en español con inflexión ascendente, sin embargo, no es menos cierto que la riqueza y variedad lingüística del español traspasa también las fronteras de la fonética experimental, y que las interrogativas absolutas presentan en nuestra lengua otros tipos de contornos terminales. Ya Navarro Tomás en 1944 admitía en su *Manual de entonación española* más de una posibilidad y en 1999 J. M. Sosa llegó a describirnos en su obra *La entonación del español* hasta cinco tipos principales, llegando a afirmar que "el contorno melódico de las interrogativas absolutas es particular y distintivo para cada dialecto" (1999: 212). Las variedades peninsulares que estudió fueron las de Madrid, Pamplona, Barcelona y Sevilla. Por fortuna en la actualidad el mapa dialectal se está completando gracias a las contribuciones realizadas dentro del proyecto AMPER-ESPAÑA; entre ellas ésta que hoy aquí presentamos sobre las hablas extremeñas y, más concretamente, sobre el habla de Don Benito (Badajoz), y cuyas conclusiones nos disponemos a recapitular a continuación.

El análisis de los datos realizado nos permite ultimar brevemente las siguientes consideraciones respecto de los tres parámetros fundamentales que definen la entonación y que hemos analizado en este estudio: F0, duración e intensidad:

1. La curva melódica de los enunciados interrogativos absolutos dombenitenses corresponde en líneas generales al patrón entonativo /-interrogativo/, es decir, tanto su rango de frecuencias, como su declinación (iniciada en el primer pico máximo y con un descenso progresivo del tono en torno a los 100Hz), como su inflexión final en cadencia (L* L%) así lo manifiestan.
2. Es especialmente significativa y característica del habla de Don Benito y de Ba-

dajoz en general el comportamiento temporal de la última vocal del enunciado. El alargamiento prolongado que ésta soporta (entre 205 y 327 ms) confiere a los dombenitenses y a los pacenses en general un suave tono tendido que les caracteriza.

3. La intensidad, por otro lado, es bastante estable (entre 96 y 100 dB); sólo en la última vocal del enunciado se produce una fluctuación máxima de 4 dB. A ello hay que añadir cierta tendencia a realizar las vocales tónicas con más intensidad que las átonas.

De todo lo que llevamos dicho se infiere que:

1. O bien el extremeño de Don Benito y, en líneas generales, de Badajoz ha debido de uniformar la entonación de dos tipos de interrogativas: las pronominales y las absolutas.
2. O bien, dadas las semejanzas que presentan el contorno melódico de las oraciones interrogativas absolutas de Don Benito y el de las oraciones declarativas de Sevilla, la igualación se ha producido en este otro sentido, y de ser así nos preguntamos ¿dónde radica, pues, la diferencia entre la modalidad oracional pregunta/afirmación?

Las respuestas a estas preguntas se escapan de los límites y objetivos de este estudio ya que para su contestación se hace imprescindible ampliar el corpus que hemos analizado. Pero las respuestas no se harán esperar ya que serán el fundamento de la comunicación que presentaremos en las *III Jornadas Científicas del Proyecto AMPER* que se celebrarán en Tenerife (Islas Canarias) durante el próximo mes de Octubre.

Sin embargo, no está de más que finalicemos este estudio recordando ciertas sugerencias que hace F. J. Cantero (2002: 86-88) acerca de lo que él llama *entonación idiomática* y A. Quilis (1993: 453-489) *nivel sociolingüístico de la entonación* y que apuntan a que el "acento" característico de cada dialecto es un fenómeno que afecta a la manera de integrar los sonidos en el discurso, y por tanto al propio proceso de comprensión del discurso. No se trataría, pues, "de un fenómeno exclusivamente entonativo (melódico) sino que tiene que ver con la función prelingüística de la entonación: se trata de un fenómeno a la vez acentual y entonativo [...] Es decir, el *acento* dialectal tiene que ver con la propia integración de los sonidos en bloques fónicos (grupos rítmicos y grupos fónicos)". En definitiva, parece ser que es competencia de la entonación, del acento y del ritmo.

BIBLIOGRAFÍA

CANELLADA, M. J.: «Notas de entonación extremeña», *Revista de Filología Española*, XXV, 1941, págs. 79-91.

CANTERO, F. J.: *Teoría y análisis de la entonación*, Barcelona: Universitat, 2002.

CONGOSTO MARTÍN, Y.: «Una primera aproximación a la prosodia del habla urbana de Sevilla», *Estudios de Fonética Experimental*, XIV, 2005, págs. 225-246.

DORTA, J. - HERNÁNDEZ, B.: «Análisis prosódico de un corpus de habla experimental: interrogativas absolutas con expansión en el objeto vs. sin expansión», *Estudios de Fonética Experimental*, XIV, 2005, págs. 67-123.

NAVARRO TOMÁS, T.: *Manual de entonación española*, Nueva York, Hispanic Institute in the United States, 1944.

QUILIS, A.: *Tratado de fonología y fonética españolas*, Madrid: Gredos, 1993.

SOSA, J. M.: *La entonación del español. Su estructura fónica, variabilidad y dialectología*, Madrid: Cátedra, 1999.

ZAMORA, F.J. - CARRERA DE LA RED, M. - MELÉNDEZ, Mª del M.: «Variación geoprosódica en dos entonaciones de Castilla y León: análisis de frases declarativas e interrogativas sin expansión», *Estudios de Fonética Experimental*, XIV, 2005, págs.125-139.

ENTONACIÓN, ACENTO Y LÍMITES SINTAGMÁTICOS EN SUJETOS CON Y SIN EXPANSIÓN EN ASTURIANO CENTRAL[1]

MIGUEL CUEVAS ALONSO
Uniovi mcuevasalonso@gmail.com

LILIANA DÍAZ GÓMEZ
Uniovi lalilis@terra.es

CARMEN MUÑIZ CACHÓN
Uniovi carmu@uniovi.es

Mª JESÚS LÓPEZ BOBO
Uniovi mjlopez@uniovi.es

RUTH GONZÁLEZ RODRÍGUEZ
Uniovi gonzalezruth@uniovi.es

INTRODUCCIÓN

Este trabajo ha sido elaborado en el marco del proyecto AMPER-Astur, cuyo objetivo general es el estudio de las variedades geoprosódicas de Asturias. En este artículo nos centraremos en el análisis de la relación entre la entonación, el acento y los límites sintagmáticos.

Tradicionalmente algunos autores como Enríquez, Casado y Santos (*Apud.* Llisterri *et al.*, 1995) o Quilis han defendido que el acento ha de coincidir por lo general con una prominencia tonal, además de con una mayor duración e intensidad. Sin embargo, algunos estudios apuntan a una falta de sincronización entre el pico tonal y la vocal tónica que da como resultado una tendencia generalizada a la posrealización (Llisterri *et al.*, 1995; Amorós 2004). Esto no sólo sucede en español sino que encontramos resultados análogos para el alemán (Peters, 1999) e incluso en el inglés americano (Silverman y Pierrehumbert, 1990).

1 Este trabajo ha sido realizado gracias a la ayuda institucional proporcionada por la Universidad de Oviedo para AMPER-ASTUR (referencia, MB-04-501-1) y a una beca de colaboración concedida por la Academia de la Llingua Asturiana.

Algunos autores explican este fenómeno como consecuencia de la relación entre la regulación del pico tonal y la proximidad de los límites de las palabras así como de los límites entonacionales (Prieto *et al.* 1994) y otros aseguran que la fuerza de los límites sintácticos y la presencia o ausencia de pausas parecen ser factores que afectan a los cambios de los picos de la frecuencia fundamental, fonológicamente asociados con la sílaba léxicamente acentuada (Llisterri *et al.* 1995). Para otros, la verdadera alineación se produce entre picos tonales y fronteras morfológicas, sintácticas y semánticas (Pamies y Amorós 2005).

En nuestro trabajo el objetivo consiste en hallar la relación existente entre la estructura acentual y la estructura entonativa, así como analizar los desplazamientos que sufren los picos tonales en función de los tipos de barreras sintagmáticas, ya sean fuertes -entre el sujeto y el verbo- o débiles -entre el núcleo del sujeto y su expansión-. Para llevar a cabo esta tarea realizaremos una descripción de los patrones entonativos del asturiano en dos niveles, tal y como recogen Hirst y di Cristo (1998: 6 y 19): un primer nivel de análisis de las características globales que afectan a la entonación como totalidad; un segundo nivel en el que estudiaremos las características más concretas que afectan a puntos determinados de la secuencia ligadas al acento léxico.

También resulta de interés la observación de los movimientos de los picos acentuales en relación a los límites sintagmáticos, discriminando en cada caso la modalidad oracional, poniendo en relación el comportamiento de las secuencias asertivas e interrogativas. De este modo pudimos observar que el asturiano tiene un patrón entonativo diferente para las dos modalidades oracionales, con una preferencia muy marcada a retrasar los picos tonales en las secuencias asertivas, en tanto que las interrogativas ofrecen picos más adelantados (Gussenhoven, 2004 recoge la tendencia contraria para algunas variedades lingüísticas del sur de Italia). Ahora bien, como señala Ladd (1996), no debemos olvidar que la disposición de los picos tonales varía de una lengua o dialecto a otro e, incluso, dentro de una misma variedad lingüística de una modalidad oracional a otra. Esta circunstancia mostraría, en nuestra opinión, la existencia de un patrón entonativo particular para cada lengua, que trasciende el ámbito fonético.

METODOLOGÍA

Para el estudio de la entonación y el acento en relación con los límites sintagmáticos hemos tomado el corpus *ad hoc* adaptado por AMPER-Astur para el asturiano, siguiendo las directrices generales de AMPER. En este artículo hemos seleccionado las muestras que responden al patrón SVO y las que presentan expansión adjetiva en el sujeto, tanto en modalidad asertiva como interrogativa. Estas frases están compuestas por palabras oxítonas, paroxítonas y proparoxítonas en todas sus posibles combinaciones, con el fin de poder establecer comparaciones entre sus respectivos esquemas entonativos. Sin embargo, de momento, centramos nuestro estudio en el análisis del sujeto, tanto cuando está compuesto por una sola palabra como cuando lo integra un núcleo con su correspondiente expansión. Esto nos permitirá prestar atención a la entonación global y a aquellas

características de la entonación asociadas al acento léxico, poniéndolos en relación con dos límites sintagmáticos: un límite fuerte y uno débil.

Las grabaciones corresponden a cuatro informantes asturianas de la zona central, más concretamente de Mieres y Oviedo, entre 25-50 años, sin estudios superiores. Fueron grabadas en el Laboratorio de Fonética de la Universidad de Oviedo con una grabadora digital DAT y tratadas mediante los protocolos establecidos por AMPER a través de los programas *Goldwave 4.25* y *Matlab 4.0*.

El número de frases utilizadas para el experimento es de 216 sin expansión en el sujeto y 648 con expansión[2]. Para analizar con la máxima fidelidad la curva melódica se midió la frecuencia fundamental en tres puntos de cada vocal -al principio, en el medio y al final- y se diseñaron unas tablas en las que se anotaron cada uno de estos valores. Posteriormente se procedió a señalar los picos más prominentes de la curva desde su inicio hasta el verbo, comprobando la existencia de un solo pico en las secuencias sin expansión y de dos en aquellas que poseen expansión del núcleo en el sujeto. En relación con la caracterización global de la entonación, también se señaló el pico máximo de los sujetos analizados.

Una vez anotadas las tablas, se confeccionó una base de datos en SPSS en la que traspasamos los datos obtenidos, teniendo en cuenta la sílaba y el punto de mayor prominencia, así como los picos ligados al acento léxico. Esto nos permitió observar de un modo exhaustivo la tendencia al desplazamiento de los picos tonales y su desplazamiento en función de las fronteras sintagmáticas del enunciado.

RESULTADOS

Enunciados sin expansión en el sujeto

En los sujetos sin expansión analizados se evidencia por regla general una clara tendencia a la posrealización del pico tonal (overshooting) y a la aproximación al límite sintagmático. En las secuencias asertivas cuyo sujeto está compuesto por palabras oxítonas el pico alterna entre la sílaba tónica y la postónica. Sin embargo, en las interrogativas de estructura análoga la posrealización se impone al alineamiento y, aún más, a la prerrealización.

En el caso de las secuencias cuyo sujeto está compuesto por una palabra paroxítona la tendencia a la posrealización es clara tanto en asertivas como en interrogativas, siendo los casos de alineamiento muy esporádicos, así como los de coincidencia con la sílaba posterior a la postónica.

A medida que se aleja el acento léxico de la frontera sintagmática el pico fluctúa entre las sílabas que siguen a la tónica pero siempre con una tendencia clara a ocupar el

2 En el caso de las secuencias sin expansión, el lector ha de tener en cuenta que se trata de sujetos con cuatro sílabas (palabras trisílabas con artículo). La misma estructura presenta el núcleo de los enunciados con expansión (artículo + palabra trisílaba); la expansión está constituida sistemáticamente por palabras trisílabas. De este modo, las sílabas uno a cuatro corresponden al núcleo del sujeto, las sílabas cinco a siete a la expansión.

lugar de la sílaba postónica. En el caso de los enunciados asertivos cuyo sujeto está integrado por palabras proparoxítonas, solo observamos alineamiento del pico tonal en unos pocos casos mientras que se realiza mayoritariamente en la postónica. Frente al comportamiento constatado en las formas oxítonas y paroxítonas, la mayor distancia que separa el acento léxico de la frontera sintáctica permite que la posrealización alterne con una situación del pico tonal en la sílaba posterior a la postónica, que coincide con el límite sintagmático. En las interrogativas la situación es similar, con un alineamiento del pico algo mayor, posrealización en la mayoría de los casos y numerosas muestras en las que el pico tonal se sitúa en la sílaba posterior a la postónica.

También cabe destacar que el pico tonal suele coincidir en la mayoría de los casos con el punto central de los tres tomados para cada una de las vocales.

Secuencias sin expansión					
Estructura	Asertivas		Estructura	Interrogativas	
	Respecto acento léxico	%		Respecto acento léxico	%
Aguda	Pretónica	—	Aguda	Pretónica	2,8
	Tónica	48,6		Tónica	33,3
	Postónica	51,4		**Postónica**	63,9
	Postpostónica	—		Postpostónica	—
Llana	Pretónica	—	Llana	Pretónica	—
	Tónica	5,6		**Tónica**	5,6
	Postónica	86,1		**Postónica**	91,7
	Postpostónica	8,3		Postpostónica	2,8
Esdrújula	Pretónica	—	Esdrújula	Pretónica	—
	Tónica	2,8		Tónica	5,6
	Postónica	75		**Postónica**	77,8
	Postpostónica	22,3		Postpostónica	16,7

Tabla 1. Distribución del pico tonal respecto al acento léxico
(en negrita se consignan los valores más relevantes)

ENUNCIADOS CON EXPANSIÓN EN EL SUJETO

Comportamiento de la primera palabra (núcleo)

Una vez vista la situación que ofrece el sujeto de los enunciados sin expansión, pasamos a analizar los resultados que presentan las tablas 2 y 3, correspondientes al análisis de los enunciados con expansión en el sujeto, haciendo especial hincapié en la relación entre límites sintagmáticos, acento léxico y patrón entonativo.

En líneas generales, observamos que en el caso de las palabras oxítonas se produce una clara tendencia al desplazamiento del pico a la sílaba postónica, más acusado en las interrogativas llegando incluso a alcanzar el 100% en la estructura aguda-llana. El número de casos de alineamiento tonal es también destacable frente al comportamiento

que presentan paroxítonas y proparoxítonas, dado que la tónica constituye en este caso el límite sintagmático débil.

En el caso de las secuencias cuyo núcleo -primera palabra- está formado por una palabra paroxítona también se observa una clara tendencia al desplazamiento del pico una sílaba hacia la derecha, tanto en secuencias asertivas como interrogativas, con escasa presencia de alineamiento tonal. En estos casos, el desplazamiento de una sílaba hacia la derecha alterna con el desplazamiento de dos sílabas, que presenta un elevado número de casos. En nuestra opinión, esto obedece al espacio existente hasta la frontera sintagmática débil.

Por lo que se refiere a los enunciados que comienzan por palabra proparoxítona, el alejamiento de la frontera sintagmática origina la escasa presencia de casos de alineamiento tonal, que sólo se constata excepcionalmente en la estructura esdrújula-esdrújula. Del mismo modo que señalábamos en los enunciados sin expansión, el pico tonal fluctúa entre la postónica y la posterior a la postónica, en una proporción más equilibrada en el caso de las asertivas. En las secuencias interrogativas parece mucho más marcada la tendencia a la postónica.

Parece pues que, si bien en las secuencias formadas por núcleo paroxítono como proparoxítono la realización en la primera postónica alterna con la segunda postónica, siendo el alineamiento tonal prácticamente esporádico, en el caso de las oxítonas la frontera sintáctica parece frenar el desplazamiento a la segunda postónica, incrementándose notablemente los casos de alineamiento tonal.

Ahora bien, un análisis detenido de la situación que presentan los picos tonales del núcleo pone de manifiesto la existencia de una relación con el acento léxico de la segunda palabra -expansión-. Así, independientemente del acento léxico del núcleo, a medida que los acentos léxicos de la expansión se sitúan más cerca del límite sintagmático débil hay un menor número de realizaciones en la sílaba postpostónica. En este sentido, cuando en la expansión del sujeto aparece una palabra aguda, el número de realizaciones con desplazamiento de dos sílabas hacia la derecha en el núcleo es mayor que en el caso de los núcleos llanos y, en éstos, mayor que en el de los esdrújulos. Se exceptúan las secuencias asertivas que comienzan por palabra esdrújula, en las que hay un desplazamiento equilibrado hacia la postónica y la sílaba posterior a la postónica.

Comportamiento de la segunda palabra (expansión)

En líneas generales, los datos presentan una gradación en función de la estructura acentual de la palabra y del mayor o menor alejamiento de la frontera sintagmática. Así, la proximidad del límite sintagmático fuerte en el caso de secuencias que disponen de una expansión oxítona, produce una clara tendencia al alineamiento tonal; este comportamiento es similar al constatado para los núcleos oxítonos. Esta coincidencia es más acentuada en las secuencias asertivas, en tanto que en las interrogativas alterna en mayor medida con una realización en la postónica. Por otro lado, es importante destacar la presencia de un número elevado de realizaciones del pico tonal en la sílaba pretónica en las secuencias interrogativas.

Pico tonal primera palabra (núcleo)				
Estructura	Asertivas		Interrogativas	
	Respecto acento léxico	%	Respecto acento léxico	%
Aguda-aguda	Tónica	17,6	Tónica	8,6
	Postónica	64,7	**Postónica**	74,3
	Postpostónica	17,6	Postpostónica	17,1
Aguda-llana	Tónica	32,1	Tónica	--
	Postónica	53,6	**Postónica**	100
	Postpostónica	14,3	Postpostónica	--
Aguda-esdrújula	Tónica	31,4	Tónica	41,2
	Postónica	62,9	**Postónica**	55,9
	Postpostónica	5,7	Postpostónica	2,9
Llana-aguda	Tónica	2,8	Tónica	--
	Postónica	27,8	**Postónica**	45,7
	Postpostónica	63,9	**Postpostónica**	54,3
	Post. a postpostónica	5,6	Post. a postpostónica	--
Llana-llana	Tónica	--	Tónica	--
	Postónica	56,4	**Postónica**	57,1
	Postpostónica	41,0	**Postpostónica**	42,9
	Post. a postpostónica	2,7	Post. a postpostónica	--
Llana-esdrújula	Tónica	11,4	Tónica	--
	Postónica	60,0	**Postónica**	91,2
	Postpostónica	28,6	Postpostónica	8,8
Esdrújula-aguda	Tónica	--	Tónica	--
	Postónica	44,4	**Postónica**	61,1
	Postpostónica	38,9	Postpostónica	33,3
	Post. a postpostónica	16,7	Post. a postpostónica	--
Esdrújula-llana	Tónica	2,9	Tónica	2,8
	Postónica	54,3	**Postónica**	66,7
	Postpostónica	40,0	Postpostónica	30,6
	Post. a postpostónica	2,9	Post. a postpostónica	--
Esdrújula-esdrújula	Tónica	5,6	Tónica	5,6
	Postónica	52,8	**Postónica**	69,4
	Postpostónica	41,7	Postpostónica	25,0

Tabla 2. Distribución del pico tonal de la primera palabra (núcleo) respecto al acento léxico

El mayor alejamiento de la frontera sintáctica que presentan las secuencias con expansión paroxítona explica la existencia de un equilibrio entre la realización con alineamiento y el desplazamiento de una sílaba a la derecha que, en este caso, coincide con la frontera fuerte. Sólo en el caso de la estructura esdrújula-llana, hemos podido constatar algunos casos de posrealización de dos sílabas.

El último eslabón de esta gradación se presenta, de nuevo, en las palabras proparoxítonas, en las que la tendencia es la posrealización de una sílaba, pero con un

mayor número de casos de desplazamiento de dos sílabas que en las combinaciones anteriores.

Pico tonal segunda palabra (expansión)				
Estructura	Asertivas		Interrogativas	
	Respecto acento léxico	%	Respecto acento léxico	%
Aguda-aguda	Pretónica	20,0	Pretónica	38,9
	Tónica	74,3	**Tónica**	55,6
	Postónica	5,7	Postónica	5,6
Llana-aguda	Prepretónica	--	Prepretónica	2,9
	Pretónica	--	Pretónica	22,9
	Tónica	88,6	**Tónica**	54,3
	Postónica	11,4	Postónica	20,0
Esdrújula-aguda	Prepretónica	8,6	Prepretónica	8,3
	Pretónica	--	Pretónica	5,6
	Tónica	91	**Tónica**	58,3
	Postónica	--	Postónica	27,8
Aguda-llana	**Tónica**	37,9	**Tónica**	40,0
	Postónica	58,6	**Postónica**	60,0
	Postpostónica	3,4	Postpostónica	--
Llana-llana	Pretónica	2,4	Pretónica	2,4
	Tónica	42,9	**Tónica**	51,2
	Postónica	54,8	**Postónica**	46,3
Esdrújula-llana	Pretónica	--	Pretónica	5,6
	Tónica	48,6	**Tónica**	50,0
	Postónica	48,6	**Postónica**	38,9
	Postpostónica	2,9	Postpostónica	5,6
Aguda-esdrújula	**Tónica**	17,6	Tónica	22,6
	Postónica	58,8	**Postónica**	51,6
	Postpostónica	23,5	Postpostónica	25,8
Llana-esdrújula	Tónica	20,6	Tónica	12,5
	Postónica	58,8	**Postónica**	56,3
	Postpostónica	20,6	Postpostónica	31,3
Esdrújula-esdrújula	**Tónica**	50,0	**Tónica**	48,6
	Postónica	36,1	**Postónica**	40,0
	Postpostónica	13,9	Postpostónica	11,4

Tabla 3. Distribución del pico tonal de la segunda palabra (expansión) respecto al acento léxico

Comportamiento del pico máximo

Como se ha indicado en la introducción de este trabajo, la situación que presenta el pico máximo en la secuencia puede tener relación con algunos cambios morfosintácticos. Una vez señaladas aquellas características entonativas locales, vinculadas a caracterís-

ticas morfológicas, presentamos a continuación los datos obtenidos a partir del análisis de la situación en que se encuentra el pico máximo en las dos modalidades oracionales.

Como se puede observar en el gráfico 1, es importante destacar que el pico máximo se sitúa en la segunda palabra en las secuencias asertivas y de forma complementaria en las interrogativas, en las que es la primera palabra la que coincide con el pico máximo.

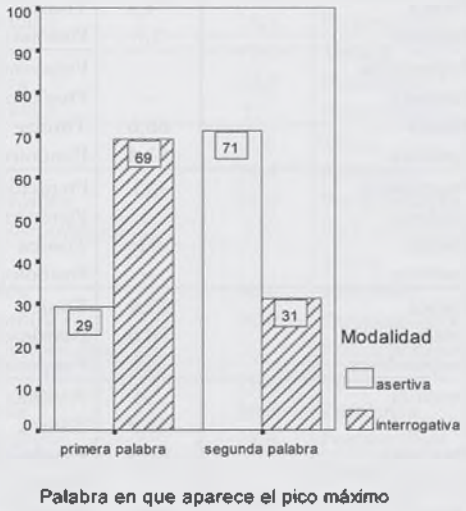


Ilustración 1. Distribución del pico máximo por modalidad

Pico máximo				
Estructura	Asertivas		Interrogativas	
	% Pal. 1	% Pal. 2	% Pal. 1	% Pal. 2
Aguda-aguda	35,3	64,7	57,1	42,9
Aguda-llana	10,7	89,3	50,0	50,0
Aguda-esdrújula	25,7	74,3	33,3	66,7
Llana-aguda	40,0	60,0	79,4	20,6
Llana-llana	15,0	85,0	82,9	17,1
Llana-esdrújula	8,8	91,2	42,9	57,1
Esdrújula-aguda	60,0	40,0	94,1	5,9
Esdrújula-llana	28,6	71,4	94,4	5,6
Esdrújula-esdrújula	36,1	63,9	74,3	25,7

Tabla 4. Distribución del pico máximo dependiendo de la estructura y de la palabra en la que aparece

CONCLUSIONES

El estudio realizado para comprobar la relación entre acento y fronteras sintagmáticas pone de manifiesto que los picos asociados con el acento léxico tienden a la posrealización de una o dos sílabas excepto en los casos en que el límite sintagmático fuerte -entre sujeto y verbo- lo establece una palabra oxítona. En tal circunstancia se observa un elevado número de casos de alineamiento tonal y un descenso en el número de casos de posrealización, más acusado en secuencias con expansión. Este hecho se explica porque el límite sintagmático frena el avance del pico a la postónica, tendencia natural entonativa como se constata cuando existe margen de una o dos sílabas para el aumento del F0 como índice acentual.

El límite sintagmático débil, esto es, entre el núcleo del sujeto y su expansión, también deja sentir su influencia pero de manera más moderada. Si bien no anula el avance del pico a la postónica cuando el núcleo del sujeto es una palabra oxítona, sí que parece limitado al desplazamiento de una sílaba a la derecha, en tanto que paroxítonas y proparoxítonas gozan de mayor espacio y aumentan considerablemente los porcentajes de desplazamiento hasta dos sílabas después de la tónica. En este sentido, se observa el mismo comportamiento en los enunciados con expansión que en aquellos en que el sujeto lo constituye una única palabra

La influencia de los límites sintagmáticos sobre la manifestación del acento tonal no parece fuertemente ligada a la modalidad, ya que la observación de los porcentajes en enunciados asertivos y en interrogativos no arroja diferencias reseñables, como se observa en los datos del apartado precedente, de igual modo en secuencias con y sin expansión.

El intento de sistematizar la relación entre los parámetros que venimos señalando nos ha llevado a fijarnos en los picos asociados al acento léxico, cuantificables en la altura en Hz que alcanza el F0. Conscientes de que el único parámetro para identificar la entonación es también la evolución del F0 -acento y entonación están íntimamente ligados pero no son lo mismo-, hemos buscado entre los picos acentuales el pico máximo como responsable genuino de la melodía entonativa.

Este pico máximo sufre los movimientos derivados de la presión de las barreras sintagmáticas -en cuanto a su vinculación con una sílaba determinada- al igual que los picos condicionados por el acento, ya que el pico máximo coincide siempre con uno de los picos acentuales. Por consiguiente, podemos afirmar que las barreras sintagmáticas también ejercen una influencia notable en relación con la entonación, observable en desplazamientos de una o dos sílabas según se manifieste en palabras de naturaleza oxítona, paroxítona o proparoxítona.

Parece también destacable la relación entre el pico máximo y la modalidad. En descripciones previas (R. González Rodríguez et al. 2005; M. J. López Bobo et al. 2005; C. Muñiz Cachón et al. 2005) hemos señalado las diferencias existentes en el pretonema entre enunciados asertivos e interrogativos en asturiano central, dado que en ambas modalidades el tonema es descendente. La observación de los picos máximos nos ha permitido constatar que en los sujetos con expansión la mayor prominencia se sitúa en el núcleo

en la modalidad interrogativa (en un 69% en la primera palabra y un 31% en la segunda) y se desplaza a la expansión -segunda palabra del sujeto- en las asertivas (en un 71% en la primera palabra y un 29% en la segunda).

BIBLIOGRAFÍA

AMORÓS CÉSPEDES, M. C. (2004) "Sincronización entre pico tonal y acento: resultados según posición métrica y morfológica" *EFE* XIII, págs. 203-223

CALLEJA AZPIAZU, N. (2004) "Alineamiento fonético de acentos tonales en el castellano de Vitoria" *EFE* XIII, págs. 39-63.

CRUTTENDEN, A. (1981) "Falls and rises: meanings and universals" *Journal of Linguistics* 17, págs. 77-91.

ESTEBAS-VILAPLANA, E (2003) "The modelling of prenuclear accents in Central Catalan declaratives" *Catalan Journal of Linguistics*, 2, págs. 97-114

FACE, TIMOTHY L. (2002) "Spanish evidence for pitch-accent structure," *Linguistics* 40-42, págs. 319-345.

FERNÁNDEZ REI, E., A. ESCOURIDO PERNAS, M. CAAMAÑO VARELA y L. XUNCAL PEREIRA (2005) "A entoación dunha fala de Santiago: fronteira prosódica entre suxeito e predicado" *EFE* XIV, págs. 141-165.

GONZÁLEZ RODRÍGUEZ, R., M. CUEVAS ALONSO, L. DÍAZ GÓMEZ, C. MUÑIZ CACHÓN y M. LÓPEZ BOBO (2005) "La entonación y el conxuntu de rasgos prosódicos na carauterización d'una llingua", *Lletres Asturianes*, nº 88, págs. 7-21.

GUSSENHOVEN, C. (2004) *The Phonology of Tone and Intonation*, Cambridge, Cambridge University Press.

HIRST, D. y A. DI CRISTO (1998) "A survey of intonation systems", en D. Hirst y A. di Cristo (eds.) *Intonation systems. A survey of twenty languages*. Cambridge, Cambridge University Press, págs. 1-44.

LADD, D. R. (1983) "Phonological Features of intonational peaks" *Language*, 59, pág. 4.

LADD, D. R. (1996) *Intonational Phonology*. Cambridge, Cambridge University Press.

LLISTERRI, J., R. MARÍN, C. DE LA MOTA y A. RÍOS (1995) "Factors affecting F0 peak displacement in Spanish" en J. M. Pardo, E. Enríquez, J. Ortega, J. Ferreiros, J. Macías, F. J. Valverde (eds.) *Eurospeech'95 Proceedings. ESCA 4th European Conference on Speech Communication and Technology. Madrid. 18-21 Septiembre, 1995. Vol. 3* págs. 2061-2064.

LLISTERRI, J., M. J. MACHUCA, A. RÍOS, C. DE LA MOTA, M. RIERA (2003) "Algunas cuestiones en torno al desplazamiento acentual en español" en Z. Herrera, E. P. Martín Butragueño (Eds.) *La tonía: dimensiones fonéticas y fonológicas*. México: El Colegio de México, Centro de Estudios Lingüísticos y Literarios (Cátedra Jaime Torres Bodet, Estudios de Lingüística, 4) págs. 163-185.

LÓPEZ BOBO, M. J., R. GONZÁLEZ RODRÍGUEZ, M. CUEVAS ALONSO, L. DÍAZ GÓMEZ y C. MUÑIZ CACHÓN (2005) "Rasgos prosódicos del centro

de Asturias: comparación Oviedo-Mieres", *Estudios de fonética experimental*, XIV, págs. 167-199.

MUÑIZ CACHÓN, C., M. J. LÓPEZ BOBO, R. GONZÁLEZ RODRÍGUEZ, M. CUEVAS ALONSO, L. DÍAZ GÓMEZ (2005) "Algunas notas acerca de la entonación en asturiano" *Actas del XXIV CILPR 2004*, Aberystwyth (RU), en prensa.

PAMIES BELTRÁN, C. y M. C. AMORÓS (2005) "Pico tonal, acento y fronteras morfosemánticas: experimento con hablantes granadinos" *EFE XIV*, págs. 201-223.

PETERS, JÖRG (1999) "The timing of nuclear high accents in German dialects". *Proceedings of the 14th International Congress of Phonetic Sciences,* San Francisco, August 1-7, 1999, Vol. 3, págs. 1877-1880

PRIETO, P. (2004): "The Search for Phonological Targets in the Tonal Space: Evidence from Five Sentence-Types in Peninsular Spanish", *Laboratory Approaches to Spanish Phonology*, ed. Timothy Face, págs. 29-59. Mouton de Gruyter, The Hague.

QUILIS, A. (1993) *Tratado de fonología y fonética españolas.* Madrid, Gredos.

SILVERMAN, K y J. PIERREHUMBERT (1990) "The timing of prenuclear high accents in English", en Kingston, J. Beckman, M. (Eds.) *Papers in Laboratory Phonology.* Cambridge, Cambridge University Press, págs. 72-106.

CONFIGURACIÓN MELÓDICA Y CAMPO TONAL DE LAS INFLEXIONES DE LAS INTERROGATIVAS NO PRONOMINALES

JOSEFA DORTA
Laboratorio de Fonética de la Universidad de La Laguna

I. INTRODUCCIÓN

Tradicionalmente se ha destacado (véase, por ejemplo, Navarro Tomás 1948 o Quilis 1993) que la configuración melódica de las interrogativas no pronominales o absolutas del castellano se inicia en un tono medio que asciende luego, en torno a la primera sílaba acentuada, hasta la postónica; a partir de ésta se inicia un progresivo descenso hasta la penúltima o última donde se registra el tono más grave para, finalmente, volver a ascender de manera que su patrón final es, según Quilis, /12↑/, es decir, "juntura terminal ascendente precedida de un nivel tonal medio" (1993: 429).

En lo que respecta al ascenso inicial, Sosa señala que es una característica general de todo el español pues afirma: "Nuestra conclusión ya expresada es que el escalonamiento ascendente (*upstep*) del tono de juntura inicial H%, que hace que la sílaba inicial del enunciado y las subsiguientes acentuadas se eleven de cierto número constante de hercios, es una característica general de todo el español, no solamente del dialecto castellano según las descripciones existentes" (1999:198).

En cambio, en lo que respecta al tonema final, el mismo Quilis, al analizar las interrogativas de Puerto Rico, observa una diferencia importante respecto de la configuración del castellano pues, en lugar del final ascendente característico de esa variedad, aquéllas se caracterizan por un pico o cima que "se sitúa sobre una vocal tónica, que preferentemente es la penúltima, en un enunciado de estructura acentual tanto paroxítona como proparoxítona. Este movimiento circunflejo, a veces, abarca toda la frase" (Quilis 1985: 155; también 1989: 59).

En el mismo sentido, García Riverón destaca para el español cubano que estas interrogativas si bien comienzan con un ascenso de la F0 a partir de la primera sílaba para luego mantenerse en suspensión o con cierta inmovilidad en el cuerpo del enunciado y

continuar ascendiendo normalmente hasta antes de la última vocal tónica, en la última sílaba acentuad a sucede un brusco descenso que continua en el segmento postónico (1996: 79).

Más recientemente, Sosa comprueba que en el tipo de interrogativas que consideramos se dan "variaciones muy importantes en la configuración de los tonemas" (1999: 198). Así, en el español porteño de Buenos Aires, en el bogotano y en el de Ciudad de Méjico el final es ascendente con un tono de juntura H% como en castellano; en cambio, en el español caribeño, insular y continental, de San Juan de Puerto Rico, Caracas o La Habana los tonemas finales son descendentes con predominio de la entonación circunfleja (1999: 203 y sigtes) con lo cual el tono de juntura es en estos casos bajo (L%) precedido de un tono alto en la sílaba acentuada (bitonal H +H* o simple H*).

En definitiva, Sosa concluye que "En general, podemos decir que el contorno melódico de las interrogativas absolutas es particular y distintivo para cada dialecto, estableciéndose las diferencias entre ellos a nivel del tonema" (1999: 212). La configuración característica de los dialectos caribeños es la "descendente desde el punto más alto del enunciado (tonema circunflejo), efecto de la secuencia H+H* L%" (1999: 212).

Pero este patrón H o H+H* L% no es exclusivo de los dialectos caribeños. Quilis, en su estudio sobre las interrogativas de Las Palmas de Gran Canaria, comprobó un comportamiento muy frecuente (76,78%) similar al de Puerto Rico, esto es, las interrogativas absolutas tienen un movimiento circunflejo que si bien a veces abarca toda la frase, es más común al final de la oración: se inicia en la última sílaba tónica y termina "en la mayoría de los casos, por encima o al mismo nivel del cuerpo del enunciado" (1989: 59).

Dorta, asimismo, comprobó en la isla de La Palma que cuando la extensión de las interrogativas es de 5-6 sílabas, se da un movimiento circunflejo que abarca toda la oración: la F0 asciende hasta alcanzar un pico máximo en la penúltima sílaba tónica para luego descender abruptamente situándose su final por debajo del inicio de la oración (Dorta 2000: 61); en cambio, cuando tienen mayor extensión, salvo alguna excepción, tienen dos picos destacados ya que la F0 inicia su ascenso hasta alcanzar el PMx1 (pico máximo 1) en la primera sílaba acentuada o en alguna de las adyacentes, luego desciende hasta el nivel inicial manteniéndose con escasas variaciones a lo largo de la oración para, por fin, como en las frases más cortas, volver a subir en la sílaba tónica en donde se da el PMx2; a partir de aquí, como en aquéllas, la F0 desciende bruscamente situándose por debajo del inicio de la curva (Dorta 2000: 63-64).

Por otra parte, en los diferentes estudios sobre las interrogativas absolutas se suele aludir a las diferencias entre los picos tonales de las oraciones interrogativas o a las existentes entre los de éstas y los de otras de diferente modalidad (declarativa vs interrogativa vs exclamativa); sin embargo, se suele obviar, o al menos no tratar de manera específica, el campo tonal de las inflexiones que se dan en la melodía, esto es, "el intervalo existente entre el valle y el pico de una inflexión ascendente o entre el pico y el valle de una inflexión descendente" (Prieto 2003: 27).

En relación con dicho intervalo, Pierrehumbert (1980) y algunos otros partidarios del modelo métrico-autosegmental consideran que la ampliación o reducción del mismo

es una cuestión meramente fonética ya que depende del énfasis que se pone al hablar, es decir, a mayor énfasis mayor será el campo tonal de una inflexión determinada. En otras palabras, el campo tonal depende de la expresividad del mensaje que se transmite y, por tanto, no tendría implicaciones de tipo fonológico, semántico o gramatical.

No obstante, sin bien parece evidente que el campo tonal se relaciona en mayor o en menor medida con el componente fonético, algunos estudiosos ponen en duda que sólo se dé tal vinculación. Así, Pilar Prieto advierte que en algunos estudios se observa que en determinados casos "el incremento del campo tonal de un acento tonal comporta una interpretación distinta del enunciado" (2003: 28). Es el caso de Ward y Hirschberg (1985) y Hirschberg y Ward (1992) que, dice Prieto, "demostraron mediante pruebas de percepción donde se manipuló artificialmente el intervalo tonal del acento nuclear que un mismo contorno del inglés admite dos interpretaciones posibles (incertidumbre o incredulidad) y que la aparición de una u otra depende del intervalo del campo tonal: así, una excursión tonal menor refleja un sentido de incertidumbre y una excursión tonal mayor uno de incredulidad" (2003: 28).

A la luz de estos datos no es extraño que Pilar Prieto y otros autores, como Robert Ladd, concluyan que el estatus del campo tonal no está aún resuelto, pues no está claramente definida su función en la entonación de las lenguas. En lo que respecta al español ello es evidente por lo que faltan estudios sistemáticos y experimentales que aclaren dicho estatus. Estos últimos me parecen imprescindibles para la consecución de tal objetivo pues sólo mediante la manipulación artificial del intervalo tonal es posible comprobar hasta qué punto la prominencia tonal repercute, por ejemplo, en la discriminación auditiva de una modalidad.

II. OBJETIVO

El presente trabajo tiene como objetivo hacer una breve descripción de las interrogativas no pronominales con el propósito de centrarme luego, ciñéndome a los datos ofrecidos por el análisis acústico, en comprobar si se dan diferencias en el campo tonal (CT) de las inflexiones de dichas oraciones y, en ese caso, ver si se relacionan con diferencias fonéticas (Pierrehumbert 1980) o de otro tipo. En estudios posteriores tendré ocasión de comparar las diferencias entre el campo tonal de las inflexiones de las interrogativas con el de otras modalidades y, asimismo, la funcionalidad de tales diferencias.

III. CORPUS, INFORMANTES Y METODOLOGÍA

El *corpus* corresponde al que hemos clasificado como experimental leído en el marco del proyecto AMPER-España y que analizamos en el proyecto 2002/058 *Prosodia de las oraciones declarativas e interrogativas del español de Canarias*[1] siguiendo la metodología

1 Este Proyecto que dirijo como Investigadora Principal, está subvencionado por la Dirección General de Universidades e Investigación de la Consejería de Educación, Cultura y Deportes del Gobierno de Canarias.

del macroproyecto AMPER (*Atlas Multimedia de la Prosodia del Espacio Románico*). Se ha considerado, por una parte, oraciones interrogativas no pronominales de 11 sílabas del tipo sintáctico SVO, concretamente SN (Sintagma Nominal) + SV (Sintagma Verbal) + SPrep (Sintagma Preposicional) (v. gr. *La cítara se toca con obsesión*)[2]. Como puede observarse en el ejemplo, el SN y el SPrep aparecen en los extremos sin expansiones de ningún tipo y su núcleo está formado por palabras trisílabas de diferente tipología acentual; en definitiva, las oraciones contemplan las diversas combinaciones con agudas, llanas y esdrújulas en la posición inicial de la oración (SN) y en la posición final (SPrep). En el SV siempre aparece una forma verbal llana.

Puesto que en el marco de nuestro proyecto y de acuerdo con el consenso general de AMPER-España hemos comenzado por el análisis de voz femenina, este primer *corpus* fue emitido por cuatro mujeres: una de La Gomera rural (Gr, 27 años), dos de Gran Canaria, una de zona rural (GCr, de 31 años) y otra urbana (GCu, de 26 años) y, por último, una tinerfeña de zona urbana (TFu, de 43 años), todas sin estudios superiores.

Siguiendo la metodología de AMPER, cada una de las oraciones fue repetida tres veces evitando la lectura monótona según unas pautas determinadas como, por ejemplo, que cada frase se presentara en una ficha individual y que las tres repeticiones no fueran sucesivas. Se obvió así las variaciones de tipo gramatical o léxico-semántico y que el contorno melódico de cada frase fuera ocasional con lo cual se neutralizó bastante la entonación de las emisiones. En definitiva, este primer *corpus* estuvo constituido por 108 oraciones interrogativas (9 x 3 repeticiones x 4 islas).

Por otra parte, considerando la extensión de la frase, he tenido en cuenta, además, el análisis del mismo tipo de oraciones pero con expansión en el objeto, es decir, oraciones en las que el SPrep consta de doble estructura acentual ya que se combinan los tres tipos de acentos mencionados como, por ejemplo, en *La cítara se toca con obsesión práctica* en donde el Sprep consta de la doble estructura acentual esdrújula-esdrújula por lo que la extensión de estas oraciones aumenta a 14 ó 15 sílabas. En este caso, la lectura corresponde a las mismas informantes ya mencionadas representativas del habla urbana de TF y de GC, sin estudios superiores, de 43 y 26 años, respectivamente. De esta manera, el análisis se centró en un conjunto de 162 oraciones, es decir, 81 interrogativas con expansión en el objeto en cada isla (27 x 3 repeticiones x 2 islas)[3].

Una vez digitalizado el *corpus* con el programa Goldwave (versión 4.25) los ficheros wav se analizaron con el programa MatLab[4] que nos permitió obtener el oscilograma de cada frase y a partir de él la segmentación de las vocales concebidas como bloques discretos a los que se asignó tres valores tonales (inicial, medio y final), así como su duración e intensidad. No obstante, para los objetivos del presente estudio fue suficiente la consideración de los puntos centrales de cada vocal, por ser más estables, además del inicio y final absolutos.

---

2 Véase la relación de este *corpus* y la identificación de cada frase en el Anexo 1 que aparece al final.

3 Véase el *corpus* en el Anexo 2 del final.

4 Nº de licencia 256105.

IV. BREVE DESCRIPCIÓN DE LA CONFIGURACIÓN MELÓDICA DE LAS INTERROGATIVAS SIN EXPANSIÓN Y CON EXPANSIÓN EN EL OBJETO

En el gráfico 1 se puede observar la configuración media de la F0 de todas las interrogativas no pronominales sin expansión y con expansión en el objeto obtenida, en ambos casos, a partir de la media de los valores centrales de F0 de cada vocal. El primer valor y el último corresponden, respectivamente, a la media de los inicios y finales absolutos.

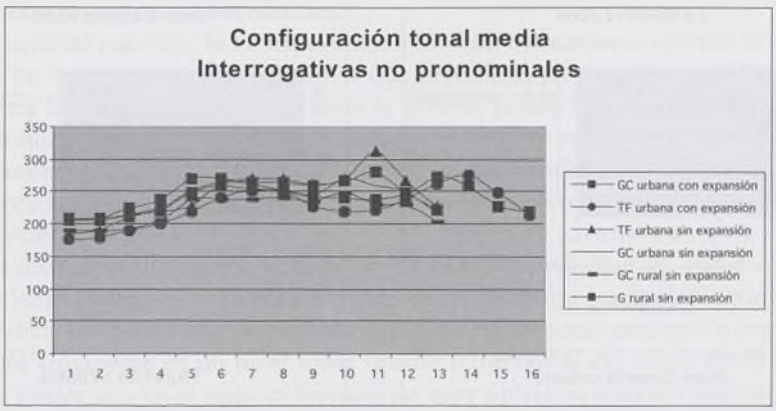


Gráfico 1: Configuración tonal media de las interrogativas no pronominales sin expansión y con expansión en el objeto

Como puede verse, según los valores medios de todas las oraciones emitidas existen pocas diferencias en la configuración melódica de este tipo de interrogativas, sin expansión o con ella. En ambos casos, tal como había señalado Sosa para el español general (1999: 198), se caracterizan en su inicio por un escalonamiento ascendente (*upstep*) pues la F0 sube de manera relativamente progresiva hasta alcanzar una primera cima tonal o pico (PMx1); sin experimentar cambios bruscos, o se mantiene en el nivel tonal alcanzado o desciende suavemente (más en las interrogativas con expansión) para elevarse de nuevo en el núcleo donde se localiza la prominencia tonal máxima (PMx2); desde ahí vuelve a descender más o menos bruscamente dándose así un característico movimiento circunflejo final muy acusado por lo que, tal como se había señalado para los dialectos caribeños y para el español de Las Palmas de Gran Canaria, se termina con un tono de juntura bajo (L%) precedido de un tono L+H. Como puede apreciarse, el pico nuclear se registra al final de la oración por lo que en las interrogativas más largas recae en la expansión del Sprep.

Ahora bien, teniendo en cuenta que en el *corpus* analizado intervienen diferentes estructuras acentúales, veamos a continuación la configuración tonal de este tipo de oraciones considerando dichas estructuras. La diferencia, pues, respecto del gráfico 1 es que

ahora se obtienen las medias agrupando las oraciones según el tipo de estructura prosódica en el SN y en el Sprep. El primer valor y el último siguen correspondiendo, respectivamente, a la media de los inicios (I) y finales absolutos (F). Véase la configuración de las interrogativas sin expansión en los Gráficos 2-5 y la de las interrogativas con expansión en el objeto en los gráficos 6 y 7.

INTERROGATIVAS SIN EXPANSIÓN

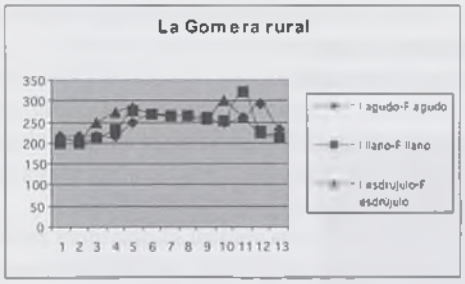


Gráfico 2

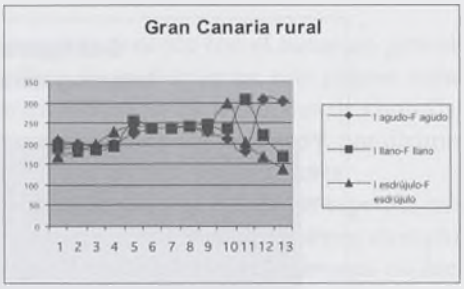


Gráfico 3

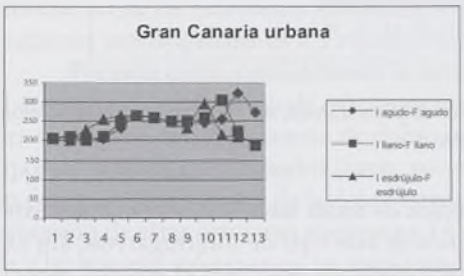


Gráfico 4

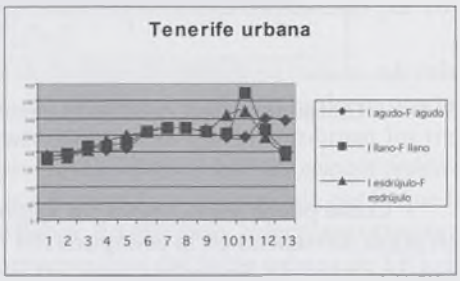


Gráfico 5

INTERROGATIVAS CON EXPANSIÓN EN EL OBJETO

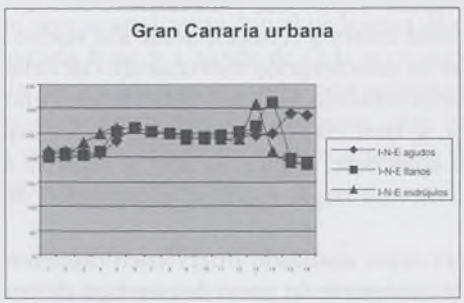


Gráfico 6

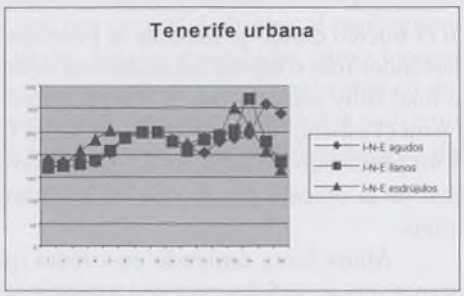


Gráfico 7

Como puede observarse, la diferente estructura acentual no altera sustancialmente la configuración inicial de las interrogativas con expansión en el objeto y sin expansión: en todos los casos y con ligeras variantes, la F0 comienza a ascender hasta llegar a la primera cima o pico tonal que nunca se da en la sílaba tónica pues lo antes que se localiza es en la última sílaba átona del SN (sílaba 4) estableciéndose así una débil frontera con el verbo[5]. En otros casos se retrasa hasta la primera sílaba átona del SV (número 5)[6], hasta la tónica del verbo –número 6- sea cual sea el tipo de estructura inicial[7] o, incluso, hasta la postónica del verbo[8]. Por otra parte, en todos los casos el primer pico alcanza e incluso supera el tono medio de las informantes.

A partir de esa cima, la F0 experimenta un ligero descenso cuyo valle se sitúa normalmente en la pretónica o antepretónica que precede al pico nuclear para luego volver a subir hasta la tónica final, esto es, hasta la última, penúltima o antepenúltima según se trate de finales agudos, llanos o esdrújulos[9] en donde se registra el tono más elevado (acento nuclear) por lo que, en todos los casos, los picos máximos marcan un escalonamiento ascendente (*upstep*) debido a que la frecuencia del pico nuclear (PMx2) es superior a la del primer pico (PMx1).

Por último, la F0 desciende de nuevo hasta el final dándose así, como dijimos ya, un movimiento circunflejo característico que ahora, al separar los finales según la estructura prosódica, se observa que es muy acusado y prácticamente igual en los finales llanos y esdrújulos situándose en un nivel tonal similar o inferior al del inicio medio; en cambio, en los finales agudos se sitúa en frecuencias muy próximas a las del pico máximo por lo cual el movimiento circunflejo o es prácticamente inexistente o es muy poco acusado.

Esto último implica, además, que la pendiente I-F es siempre *ascendente* a veces muy pronunciada cuando interviene una estructura oxítona en el SN y en el Sprep (entre 5-7 semitonos)[10]. En cambio, cuando en esos mismos sintagmas aparecen estructuras llanas y esdrújulas, las pendientes I-F son muy poco acusadas (entre 0-3 semitonos); por otra parte, son siempre *descendentes* en GC rural y urbana con la diferencia de que en las interrogativas sin expansión es algo más pronunciada (2-3 semitonos) que en las interrogativas con expansión (1 semitono). En Tenerife, por el contrario, son *ascendentes* como en los agudos aunque, a diferencia de éstos, muy poco pronunciadas (1-2 semitonos)[11].

---

5 Sucede así en las interrogativas sin expansión con SN llano y esdrújulo de La Gomera rural y GC rural (Gráficos 2 y 3) y con SN esdrújulo de GC urbana (Gráfico 4); también en las interrogativas con expansión con inicio esdrújulo de TF urbana (Gráfico 7).

6 Véanse las interrogativas sin expansión con SN agudo de La Gomera rural y GC rural (Gráficos 2 y 3) y las que tienen SN agudo y llano de GC urbana (Gráfico 4); asimismo sucede en todas las interrogativas con expansión en el objeto de GC urbana (Gráfico 6).

7 Véanse las interrogativas sin expansión de TF (Gráfico 5).

8 Sucede así en las interrogativas con expansión de TF urbana que tienen inicio agudo o llano (Gráfico 7).

9 Se exceptúa sólo que en las interrogativas sin expansión con final esdrújulo de TF el pico recae en la postónica.

10 Se exceptúa sólo que en La Gomera la pendiente I-F, aunque ascendente en términos absolutos, es relativamente neutra (0 semitonos).

11 Se exceptúa sólo que en las interrogativas con expansión es descendente en los esdrújulos aunque también poco pronunciada (1semitono).

Por último, las oraciones gomeras se aproximan a las tinerfeñas en que la pendiente es siempre ascendente aunque relativamente menos acusada (0-1 semitonos)

V. EL CAMPO TONAL DE LAS INFLEXIONES MELÓDICAS

A partir del análisis acústico efectuado, he procedido a delimitar el campo tonal de los dos picos tonales referidos, esto es, el intervalo de F0 entre el V1 (valle 1) y PMx1 (primer pico) y el que se da entre el V2 (valle 2) y el PMx2 (segundo pico o pico nuclear). El valor medio absoluto (en Hz) de los puntos señalados puede verse en las Tablas 1 y 2; además, con el propósito de mostrar la caída de F0 desde el PMx2, he incluido en dichas tablas el valor medio del final.

	V1				PMx1				V2				PMx2				F			
	GCr	Gr	GCu	TFu	GCr	Gr	GCu	TFu	GCr	Gr	GCu	TFu	GCr	Gr	GCu	TFu	GCr	Gr	GCu	TFu
A	198	214	208	204	241	269	265	271	183	250	249	241	305	294	325	298	304	233	274	290
LL	185	199	202	214	253	276	265	271	237	251	251	255	307	322	307	374	170	210	188	200
E	179	208	204	186	243	284	265	271	234	257	234	266	297	303	296	320	139	224	187	187

Tabla 1: Campo tonal de las inflexiones ascendentes de las interrogativas sin expansión

	V1		PMx1		V2		PMx2		F	
	GCu	TFu	Gcu	TFu	GCu	TFu	Gcu	TFu	Gcu	TFu
A	208	184	259	250	238	203	290	309	286	290
LL	204	177	259	250	241	209	314	323	189	183
E	206	180	259	251	238	225	318	301	186	165

Tabla 2: Campo tonal de las inflexiones ascendentes de las interrogativas con expansión en el objeto

Partiendo de dichos valores realicé una nueva estilización para mostrar gráficamente la diferencia entre los intervalos de ambos picos. Véanse los gráficos 8-13.

INTERROGATIVAS SIN EXPANSIÓN

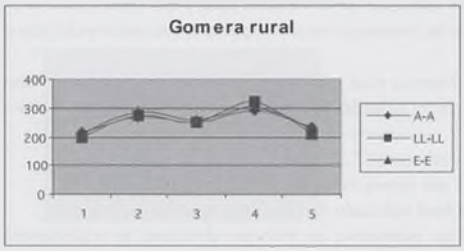


Gráfico 8

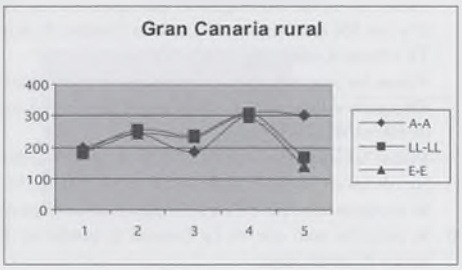


Gráfico 9

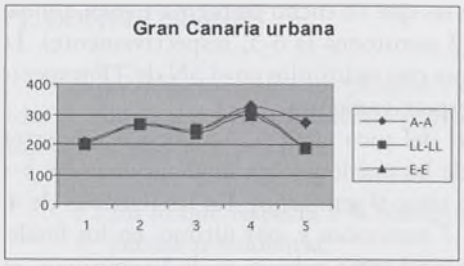


Gráfico 10

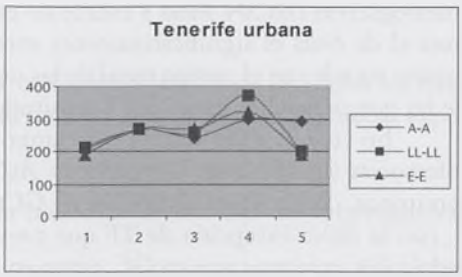


Gráfico 11

INTERROGATIVAS CON EXPANSIÓN

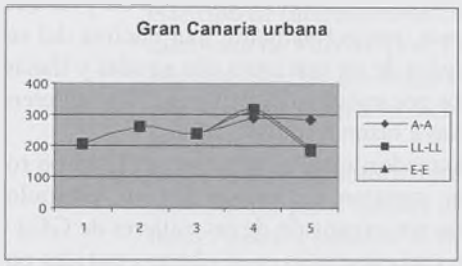


Gráfico 12

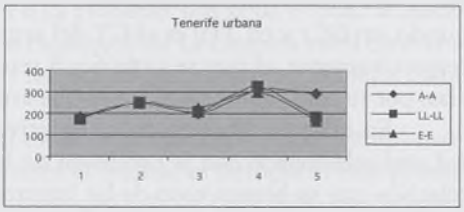


Gráfico 13

En la tabla 3 se muestran los valores relativos (en semitonos) de los campos tonales de las inflexiones de las oraciones interrogativas analizadas.

	INTERROGATIVAS SIN EXPANSIÓN							
	Gr		GCr		GCu		TFu	
	V1-PMx1	V2-PMx2	V1-PMx1	V2-PMx2	V1-PMx1	V2-PMx2	V1-PMx1	V2-PMx2
A	4	3	3	9	4	4	5	4
LL	5	4	6	5	4	4	4	7
E	5	3	5	5	4	5	6	2

Tabla 3: Valores relativos de las interrogativas sin expansión

Como puede observarse, estos valores relativos indican que el campo tonal del primer pico, salvo excepciones no sistemáticas, es de 4-5 semitonos lo que implica que se dan pocas diferencias significativas entre las diferentes informantes rurales y urbanas de las distintas islas. Por otra parte, teniendo en cuenta las diferentes estructuras acentúales, lo único que merece destacarse es que, a diferencia de lo que sucede en GCu donde dicho campo tonal es siempre de 4 semitonos, en GCr se diferencia significativamente el de las

interrogativas con SN llano y esdrújulo del de las que en dicho sintagma tienen agudas pues el de éstas es significativamente menor (3 semitonos *vs* 6-5, respectivamente). Lo mismo sucede con el campo tonal de las oraciones con esdrújulos en el SN de TF respecto de las que tienen SN llano (6 y 4 semitonos, respectivamente).

En cuanto a los valores del campo tonal del pico nuclear, tampoco se advierten diferencias significativas sistemáticas. Así, el de las oraciones con final agudo es de 3-4 semitonos, con la única excepción de GCr que tiene 9 semitonos. En las llanas es de 4-5, con la única excepción de TF que presenta 7 semitonos y, por último, en los finales esdrújulos, mientras que en GC, tanto en zona rural como urbana, es de 5 semitonos, en Gr desciende a 3 y en TF a 2 semitonos. Tales resultados no evidencian, por tanto, que la diferente estructura acentual afecte sistemáticamente al campo tonal del pico nuclear.

Lo mismo sucede si comparamos los valores de los dos campos tonales en cada informante. Así, mientras que en las oraciones con estructuras esdrújulas de Gr y de TFu el CT del primer pico se sitúa en 2 y 4 semitonos, respectivamente, por encima del segundo, en GCr y en TFu es el CT del segundo pico de las oraciones con agudas y llanas, respectivamente, el que se sitúa 6 y 3 semitonos por encima del primero. Tales diferencias, por su asistematicidad, habrá que atribuirlas a razones individuales.

En relación con otro de los objetivos planteados, esto es, observar si el campo tonal está relacionado con la extensión de la frase, mostraré a continuación los resultados relativos que ya hemos visto de las interrogativas sin expansión de las mujeres de GCu y de TFu y, junto a ellos, los de las interrogativas con expansión en el objeto emitidas por esas mismas mujeres. Son los que se pueden ver en la Tabla 4.

Puede observarse, en primer lugar, que en GCu el campo tonal de los picos de las interrogativas sin expansión y con expansión en el objeto no presenta diferencias significativas pues los valores oscilan siempre entre 4-5 semitonos. TF, en cambio, es menos monótono pues en las interrogativas con expansión aguda y esdrújula, el CT del segundo pico supera en 4 y 3 semitonos, respectivamente al de las interrogativas sin expansión, mientras que en las interrogativas con expansión llana es el CT del primer pico el que supera en 2 semitonos al correspondiente de las oraciones sin expansión.

	GCu				Tfu			
	Sin expansión		Con expansión		Sin expansión		Con expansión	
	V1-PMx1	V2-PMx2	V1-PMx1	V2-PMx2	V1-PMx1	V2-PMx2	V1-PMx1	V2-PMx2
A	4	4	4	4	5	4	6	8
LL	4	4	4	5	4	7	6	7
E	4	5	4	5	6	2	6	5

Tabla 4: Valores relativos de las interrogativas sin expansión y con expansión en el objeto

Por último, si comparamos las dos islas observamos que, salvo excepciones, los valores del CT son más altos en TFu, aunque las diferencias más significativas y constantes se dan entre las interrogativas con expansión en el objeto.

VI. CONCLUSIONES

El estudio realizado nos permite extraer varias conclusiones de las cuales las referidas al campo tonal las consideraré provisionales a falta de más análisis y pruebas experimentales. Son las siguientes:

1ª) En lo que respecta a la configuración melódica de este tipo de interrogativas isleñas se evidencia:
   a) las que tienen un final agudo son las que más se asemejan a las descripciones realizadas para el castellano y otras variedades del español (*v. gr.* el bonaerense, bogotano o mexicano) debido a que en su final, o no se da el movimiento circunflejo característico de los otros dos finales que relaciona a este tipo de interrogativas con las del español caribeño insular y continental (San Juan de Puerto Rico, Caracas o La Habana), o si se da es poco acusado. No obstante, las interrogativas sin expansión de La Gomera rural constituyen una excepción a lo dicho pues los finales agudos son circunflejos como los otros.
   b) Lo común en las informantes es el escalonamiento ascendente (*upstep*) marcado por los picos tonales de los cuales el nuclear es siempre el más prominente; destaca siempre el de las interrogativas con final llano con la única excepción de las de Gran Canaria urbana. Con todo, la diferencia respecto de los otros finales no es relativamente significativa.
   c) El primer pico no actúa como índice de frontera entre el SN y el SV pues incluso cuando se da al final del trisílabo inicial la diferencia respecto de la altura tonal del verbo es muy pequeña.
   d) La mayor prominencia tonal del núcleo se asocia con la sílaba que lleva el acento léxico.

2ª) En lo que respecta al campo tonal de los picos, la conclusión más importante es que las diferencias observadas parecen estar más relacionadas con diferencias individuales que con diferencias lingüísticas pues no se dieron hechos sistemáticos que prueben lo contrario. Se exceptúa sólo que la diferente extensión de la frase, si bien no influye de manera decisiva en la informante urbana de Gran Canaria, sí lo hace en la informante de Tenerife aunque habría que corroborar en el futuro si tal hecho es sistemático o meramente ocasional.

3ª) En definitiva, parece confirmarse en buena medida la tesis de Pierrehumbert (1980) respecto de que la ampliación o reducción del campo tonal es una cuestión meramente fonética. No obstante, el presente estudio habrá de servir para comparar los resultados obtenidos con los de otras modalidades, como la declarativa, que permitirán ratificar si, en efecto, las diferencias observables en el campo tonal pueden estar relacionadas con razones diferentes a las meramente fonéticas como pueden ser la modalidad de la frase o su extensión.

BIBLIOGRAFÍA CITADA

DORTA, Josefa (2000): "Entonación hispánica: interrogativas no pronominales *vs.* pronominales", *Lingüística Española Actual* XXII, 1, 51-76.

GARCÍA RIVERÓN R., (1996), *Aspectos de la entonación hispánica*, T II *Análisis acústico de muestras del español de Cuba*, Cáceres: Servicio de Publicaciones de la Universidad de Extremadura.

NAVARRO TOMÁS, Tomás (1948): *Manual de entonación española.* Madrid: Guadarrama, 1974.

PIERREHUMBERT, Janet (1980): *The phonology and phonetics of English intonation.* Tesis Doctoral. Cambridge, Massachusetts: MIT.

PRIETO, Pilar (Ed.) (2003): *Teorías de la entonación.* Barcelona: Ariel.

QUILIS, Antonio (1985): "Entonación dialectal hispánica", *Lingüística Española Actual* VII, 145-190.

QUILIS, Antonio (1989): "La entonación de Gran Canaria en el marco de la entonación española", *Lingüística Española Actual* XI, 55-87.

QUILIS, Antonio (1993): *Tratado de fonología y fonética españolas.* Madrid: Gredos.

SOSA, Juan Manuel (1999): *La entonación del español. Su estructura fónica, variabilidad y dialectología.* Madrid: Cátedra.

ANEXO 1: *Interrogativas sin expansión*

1. La guitarra se toca con paciencia (twt)
2. La guitarra se toca con obsesión (twk)
3. La cítara se toca con pánico (pwp)
4. La cítara se toca con paciencia (pwt)
5. El saxofón se toca con obsesión (kwk)
6. La guitarra se toca con pánico (twp)
7. La cítara se toca con obsesión (pwk)
8. El saxofón se toca con paciencia (kwt)
9. El saxofón se toca con pánico (kwp)

ANEXO 2: *Interrogativas con expansión en el objeto*

1. ¿El saxofón se toca con obsesión y con amor? (kwg)
2. ¿La guitarra se toca con obsesión y con amor? (twg)
3. ¿La cítara se toca con obsesión y con amor? (pwg)
4. ¿El saxofón se toca con paciencia y con amor? (kwd)
5. ¿La guitarra se toca con paciencia y con amor? (twd)

6. ¿La cítara se toca con paciencia y con amor? (pwd)
7. ¿El saxofón se toca con pánico y con amor? (kwb)
8. ¿La guitarra se toca con pánico y con amor? (twb)
9. ¿La cítara se toca con pánico y con amor? (pwb)
10. ¿El saxofón se toca con obsesión finita? (kwx)
11. ¿La guitarra se toca con obsesión finita? (twx)
12. ¿La cítara se toca con obsesión finita? (pwx)
13. ¿El saxofón se toca con paciencia finita? (kws)
14. ¿La guitarra se toca con paciencia finita? (tws)
15. ¿La cítara se toca con paciencia finita? (pws)
16. ¿El saxofón se toca con pánico finito? (kwf)
17. ¿La guitarra se toca con pánico finito? (twf)
18. ¿La cítara se toca con pánico finito? (pwf)
19. ¿El saxofón se toca con obsesión práctica? (kwj)
20. ¿La guitarra se toca con obsesión práctica? (twj)
21. ¿La cítara se toca con obsesión práctica? (pwj)
22. ¿El saxofón se toca con paciencia práctica? (kwz)
23. ¿La guitarra se toca con paciencia práctica? (twz)
24. ¿La cítara se toca con paciencia práctica? (pwz)
25. ¿El saxofón se toca con pánico práctico? (kwv)
26. ¿La guitarra se toca con pánico práctico? (twv)
27. ¿La cítara se toca con pánico práctico? (pwv)

APROXIMACIÓN A LA PROSODIA DEL CASTELLANO DE ARABA: ORACIONES SIN Y CON EXPANSIÓN EN EL SINTAGMA VERBAL

Ana Elejabeitia, Alexander Iribar, Rosa Miren Pagola
Laboratorio de Fonética. Universidad de Deusto

1. PRESENTACIÓN

Este trabajo pertenece al proyecto AMPER-Eusk, que a su vez forma parte del proyecto AMPER (Atlas Multimedia de la Prosodia del Espacio Románico).[1] En esta ocasión, se ha analizado la prosodia de una informante de la zona urbana de Araba (concretamente de su capital, Vitoria-Gasteiz), con el objeto de examinar además la influencia prosódica de la expansión del sintagma verbal. Para ello, se analizarán, de acuerdo con el protocolo AMPER, dos conjuntos de oraciones, sin expansión y con ella, y se efectuarán las oportunas comparaciones.

Los estudios acerca de la prosodia del castellano hablado en la zona lingüística vasca son muy escasos. Recientemente, CALLEJA (*2004*) ha examinado el alineamiento fonético de los acentos tonales del castellano de Vitoria-Gasteiz, siguiendo el modelo utilizado por FACE (*2002*) para Madrid, con cuya habla encuentra semejanzas.

2. ANÁLISIS DE LA PROSODIA DE ARABA (ZONA URBANA)

El análisis de la prosodia del castellano de la zona urbana de Araba se basa en el material producido por la informante 381: mujer, 47 años, nivel socio-cultural medio, natural de Vitoria-Gasteiz, monolingüe castellana.

---

1 Puede encontrarse abundante información sobre este proyecto en el número XIV de la revista *Estudios de Fonética Experimental*, dedicado monográficamente a los últimos resultados de AMPER-España. En ELEJABEITIA-IRIBAR-PAGOLA (*2004*), publicado en ese volumen, se da cuenta general del proyecto AMPER-Eusk.

2.1. Entonación de las oraciones sin expansión

Las oraciones sin expansión son 18, correspondientes a las tres repeticiones de oraciones enunciativas e interrogativas de acento agudo, grave y esdrújulo. Se han manejado los valores tonales de cada una de sus once sílabas, obtenidos en el punto central de cada núcleo silábico. En el caso de las frases con acento final agudo, se han colocado dos valores para la última sílaba: el central y el final.[2]

En general, las oraciones interrogativas presentan una mayor altura tonal que las oraciones afirmativas, especialmente aquéllas con acento de frase agudo, cuyo rango tonal, además, es el menor. Ello se debe fundamentalmente a que estas oraciones no presentan el acusado descenso final que aparece en las oraciones interrogativas de acento llano y esdrújulo. Los datos aparecen en la siguiente tabla:

	Medias tonales			Rangos tonales		
	Enunciativas	Interrogativas	Todas	Enunciativas	Interrogativas	Todas
Agudas	231,89	283,33	257,61	179	143	181
Llanas	239,82	273,64	256,73	180	185	185
Esdrújulas	230,33	274,94	252,64	163	184	194
Todas	233,95	277,48	255,72	190	186	194

Las frases enunciativas comienzan ligeramente por debajo del valor tonal medio y finalizan en un tono aún más bajo, por lo que su pendiente Inicio-Final es descendente. Las frases interrogativas comienzan muy por debajo de su respectivo valor tonal medio, pero su valor final difiere según el acento de frase: asciende en las frases agudas y desciende –aunque menos que las afirmativas– en los otros dos casos. Los valores de las pendientes, en semitonos, aparecen en la siguiente tabla:

	Enunciativas	Interrogativas
Frases agudas	-3,05	+4,37
Frases llanas	-3,51	-2,58
Frases esdrújulas	-3,57	-2,87

Examinemos seguidamente la influencia de los acentos en la curva entonativa, por medio de los siguientes gráficos:

2 Todas las gráficas aparecen en el Anexo.

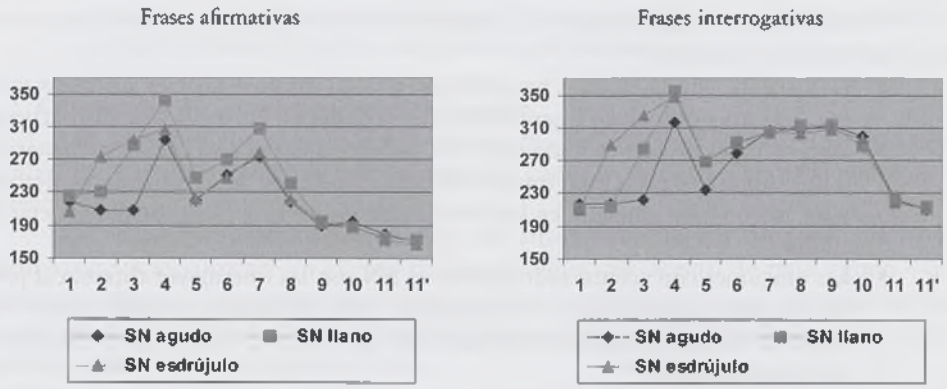


Fig. 1.- Afirmativas e interrogativas agrupadas por el último acento[3]

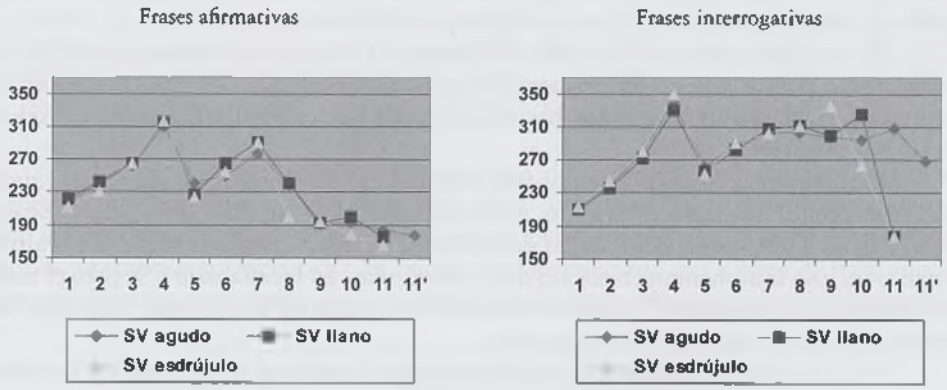


Fig. 2.- Afirmativas e interrogativas agrupadas por el primer acento[4]

En todas las oraciones, tanto afirmativas como interrogativas, la curva asciende hasta alcanzar el primer pico tonal en la sílaba 4, es decir, en la frontera de palabra y de SN. Esta sílaba constituye en todos los casos la cima tonal de la oración. Además, pueden señalarse algunas particularidades:

- La altura tonal del pico es mayor en las oraciones interrogativas.
- La mayor subida tonal intersilábica en el tramo inicial se registra en todos los casos en la sílaba tónica: la segunda con SN esdrújulo (*la cítara...*), la tercera con SN llano (*la guitarra...*) y la cuarta con SN agudo (*el saxofón...*).

3 En este caso, se agrupan las frases de la siguiente manera: "el saxofón se toca con obsesión/paciencia/pánico", "la guitarra se toca con obsesión/paciencia/pánico", "la cítara se toca con obsesión/paciencia/pánico".

4 En este otro caso, se agrupan las frases de la siguiente manera: "el saxofón/la guitarra/la cítara se toca con obsesión", "el saxofón/la guitarra/la cítara se toca con paciencia", "el saxofón/la guitarra/la cítara se toca con pánico".

Si examinamos separadamente la influencia del acento léxico en este primer tramo, observamos lo siguiente:

- Con acento agudo, apenas hay diferencia tonal durante las tres primeras sílabas. La curva es incluso ligeramente descendente en las oraciones afirmativas.
- Las oraciones de acento llano en el SN, independiente de su carácter afirmativo o interrogativo, presentan una altura tonal similar en las dos primeras sílabas, y un ascenso pronunciado en las dos siguientes, hasta la cima tonal de la cuarta sílaba.
- Las oraciones con acento esdrújulo en el SN son las que mayor diferencia presentan entre afirmativas e interrogativas. Esta diferencia no atañe tanto a la forma de la curva, cuanto a la altura tonal que alcanza, notablemente mayor en las interrogativas.

A partir de la cima tonal de la sílaba 4, la curva desciende bruscamente en la sílaba 5 (algo menos en las interrogativas que en las afirmativas), y asciende seguidamente hasta llegar a la segunda cima tonal de la frase, que se produce en la sílaba 7 (frontera de palabra y límite entre el verbo y su sintagma preposicional).

El comportamiento de la sílaba 8 diferencia las oraciones afirmativas de las interrogativas, puesto que en las primeras se produce un acusado descenso (de entre 2,2 semitonos, en el caso de las agudas, hasta 5,2, en el de las esdrújulas), que no se produce en las interrogativas.

El tramo final de la curva entonativa, en torno al último acento, diferencia fundamentalmente los dos tipos de oración: en las afirmativas, se produce un descenso brusco en las sílabas 8 y 9, y muy débil en las dos últimas; las interrogativas mantienen la altura tonal -o incluso la aumentan en el caso de las esdrújulas- en las sílabas 8 y 9, pero el tonema desciende bruscamente. La sílaba tónica se destaca más en la oraciones interrogativas que en las afirmativas. Más concretamente:

- Con acento final agudo, en las oraciones afirmativas se produce un descenso acusado y escalonado desde la sílaba 7 a la 9 (más de 5 semitonos); se mantiene en la sílaba 10 y desciende muy ligeramente en la final (tónica). En las oraciones interrogativas, el tono desciende ligeramente hasta la sílaba 10, asciende en el primer valor de la sílaba 11 (tónica) y cae 2 semitonos en su segundo valor.
- Con acento final llano, el tono desciende 5,6 semitonos en las oraciones afirmativas desde la sílaba 7 hasta la pretónica (9), después asciende ligeramente en la tónica y sigue descendiendo (2 semitonos) en la final. En las oraciones interrogativas, se mantiene en una altura tonal más o menos constante hasta la silaba pretónica, asciende en la tónica (más que en las afirmativas) y desciende bruscamente (7,5 semitonos) en la sílaba final.
- Con acento final esdrújulo, en las oraciones afirmativas se produce un descenso continuado del tono desde la sílaba 7 hasta la final; el mayor descenso se produce en la sílaba pretónica (5,2 semitonos) y el menor (medio semitono), en la tónica (9). En las oraciones interrogativas, el tono asciende hasta la sílaba

tónica (1,6 semitonos desde la sílaba 7) y desciende de forma abrupta en las dos últimas sílabas (7,8 semitonos hasta la sílaba final).

2.2. Entonación de las oraciones con expansión

Las oraciones con expansión en el sintagma verbal son 54, porque la expansión triplica el número de posibilidades acentuales, como se muestra detalladamente en el Anexo.

Las oraciones interrogativas tienen una altura tonal media superior a la de las afirmativas, especialmente en el caso de las oraciones con acento de frase agudo,[5] que presentan además el menor de los rangos tonales. Como en el caso de las oraciones sin expansión, esto se debe a que no presentan el brusco descenso final de llanas y esdrújulas. Los datos aparecen en la siguiente tabla:

	Medias tonales			Rangos tonales		
	Enunciativas	Interrogativas	Todas	Enunciativas	Interrogativas	Todas
Agudas	225,30	279,70	252,50	148	135	163
Llanas	235,68	275,41	255,55	146	153	170
Esdrújulas	229,84	267,97	248,90	136	172	175
Todas	230,16	274,48	252,32	154	172	180

Todas las oraciones comienzan por debajo de su respectivo valor medio, ligeramente en las enunciativas y notablemente en las interrogativas. Todas finalizan en un tono aún más bajo que el inicial, por lo que la pendiente Inicio-Final es descendente. Hay, no obstante, una excepción: las oraciones interrogativas con acento de frase agudo, cuya pendiente es ascendente y muy acusada. Los datos se recogen en la tabla siguiente:

	Enunciativas	Interrogativas
Frases agudas	- 2,09	+ 5,46
Frases llanas	- 2,67	- 1,60
Frases esdrújulas	- 2,34	- 2,29

La influencia de los acentos en la línea melódica puede examinarse en la gráfica siguiente:[6]

---

5 La diferencia tonal entre los dos tipos de oración es de 3,24 semitonos en el caso de las agudas, frente a 2,40 y 2,37 en las llanas y las esdrújulas.

6 En esta ocasión no presentamos la gráfica que agrupa las oraciones por el último acento, porque se ha comprobado que no genera comportamientos muy diferenciados, salvo en el caso de las agudas (tal vez porque, como se explica más detalladamente en el Anexo, no contiene el mismo número de datos). Por otra parte, el objetivo de este trabajo es examinar la expansión en el sintagma verbal.

Así pues, cada curva de la gráfica corresponde a: "el saxofón/la guitarra/la cítara se toca con obsesión/paciencia/pánico y con amor"; "el saxofón/la guitarra/la cítara se toca con obsesión/paciencia/pánico finito"; "el saxofón/la guitarra/la cítara se toca con obsesión/paciencia/pánico práctico"

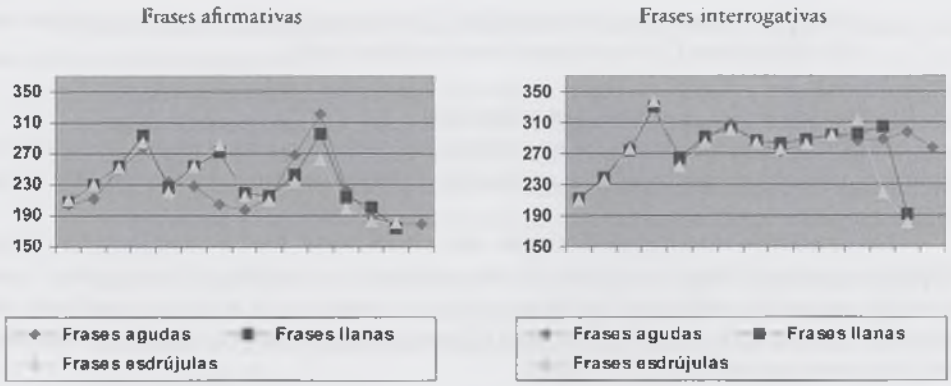


Fig.3.- Afirmativas e interrogativas agrupadas por los primeros acentos

La curva melódica asciende en todos los casos hasta la cima tonal de la sílaba 4, en la frontera de palabra y de SN. Esta sílaba constituye la cima oracional en todas las oraciones interrogativas, pero únicamente en un caso de las oraciones afirmativas (las de acento oracional esdrújulo).

Después de esta cima, la curva desciende bruscamente en la siguiente sílaba (3,4 semitonos en las afirmativas y 3,7 en las interrogativas). Tras este valle, se inicia un nuevo ascenso hasta la siguiente cima, situada en la sílaba 7 (frontera entre el verbo y su sintagma preposicional). Esta cima es la menos relevante de las tres que se registran en las oraciones, con la excepción de las oraciones interrogativas con acento de frase agudo.

No es la anterior la única particularidad melódica de las oraciones con acento de frase agudo. Precisamente, tanto las afirmativas como las interrogativas oxítonas presentan un comportamiento diferente del resto. En lo que respecta a la señalada cima tonal de la sílaba 7, ésta no se produce en las afirmativas agudas, cuya curva melódica continúa descendiendo desde el valle de la sílaba 5 hasta el siguiente valle (sílaba 8).

A partir de la cima tonal señalada en la sílaba 7, la curva melódica se comporta de distinta manera en afirmativas e interrogativas. En las primeras (salvo las agudas), se produce un acusado descenso tonal (3,5 semitonos), que da lugar al segundo valle oracional, que ocupa las sílabas 8 y 9. En las oraciones interrogativas, la depresión del valle es muy poco significativa, aunque se prolonga a lo largo de más sílabas, puesto que asciende progresivamente hasta el tonema.

Por otra parte, hay que hacer notar que la curva correspondiente a las oraciones agudas no es del todo exacta, puesto que el acento del SN verbal origina distintas silabificaciones: cuando es agudo ("con obsesión y con amor") hay una sílaba más que en los otros casos ("con paciencia y con amor" y "con pánico y con amor"). Por tanto, la curva general debería contener un punto más en el primer caso. Además, en la gráfica actual, el punto 14' corresponde al segundo valor de la sílaba final en dos casos (llanos y esdrújulos), y al valor inicial de la sílaba final en el otro (agudo). Hemos optado por esta presentación conjunta por una mayor claridad, dado que se ha comprobado que distinguir todos los casos no aporta una gran información.

En las oraciones afirmativas, la curva melódica asciende abruptamente (4,5 semitonos) desde la sílaba 9 hasta la 11 (frontera de palabra anterior a la expansión), donde se produce, como se ha señalado antes, la cima oracional, salvo en las oraciones con acento de frase esdrújulo.[7] Seguidamente, se produce otro descenso igualmente abrupto (4,6) en la sílaba 12, a partir de la cual el tonema experimenta un suave y continuado descenso, aminorado en cada caso en la sílaba tónica correspondiente: 12 en las esdrújulas, 13 en las llanas y 14 en las agudas.[8]

En las oraciones interrogativas, el comportamiento de las frases agudas vuelve a destacarse en relación al resto. En llanas y esdrújulas, la curva asciende progresivamente desde la sílaba 9 hasta la última tónica (12 y 13, respectivamente), a partir de la cual desciende de modo brusco (6,1 y 6,9 semitonos, respectivamente). Las oraciones agudas, debido probablemente al mayor número de sílabas que separa la sílaba 9 del tonema, presentan una inflexión en la sílaba 11 que no aparecía en los otros casos. En el tonema, la sílaba tónica también coincide con un leve pico tonal, tras el cual se produce el habitual descenso final, en este caso mucho menor que en los anteriores (2 semitonos, aproximadamente).

2.3. Comparación de las dos entonaciones

La siguiente gráfica muestra con claridad el comportamiento general de las curvas entonativas, afirmativas e interrogativas, en las dos modalidades analizadas: sin expansión y con expansión. Nótese que, en las oraciones sin expansión, la curva presenta una discontinuidad en los valores centrales.

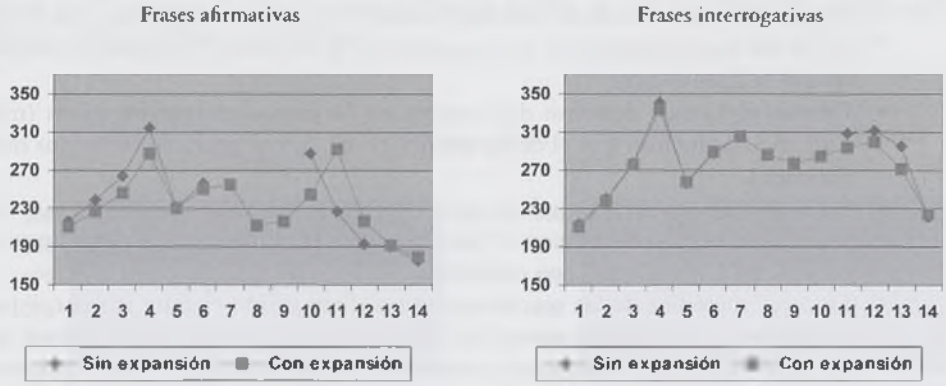


Fig. 4.- Oraciones enunciativas e interrogativas, con y sin expansión

7 ¿La diferencia sintáctica de la expansión aguda (dos sintagmas prepositicionales frente a uno solo, con adjetivo) influirá en la mayor altura tonal de su cima en la sílaba 11?

8 Por lo señalado en la nota 6, la gráfica no permite ver la pequeña inflexión que origina en el tonema el acento agudo. En el Anexo, como ilustración, se incluye una gráfica de las oraciones afirmativas que amplía las cinco últimas sílabas, teniendo en cuenta la diferente silabificación.

La gráfica muestra que la expansión gramatical del sintagma verbal no genera diferencias importantes en las curvas entonativas, siempre que se comparen adecuadamente:

- Hay que comparar entre sí las curvas correspondientes al sintagma nominal y al verbo (S1 a S7), idénticas en ambos casos.
- Dentro del sintagma verbal, hay que comparar las cuatro últimas sílabas, correspondientes al sintagma preposicional de las oraciones sin expansión (S8 a S11), con las otras cuatro últimas con expansión (S11 a S14).
- De lo anterior resulta la discontinuidad señalada en la gráfica.

De esta manera, se comprueba la semejanza entre las dos curvas melódicas, especialmente en las oraciones interrogativas.

En las afirmativas, la semejanza entre el tramo final de las dos curvas abarca una sílaba más (S10 a 14), puesto que la cima tonal de oración no ocupa la misma sílaba en los dos casos: es la pretonemática en las oraciones con expansión, y la anterior a ésta en las de sin expansión.

En definitiva, la expansión del sintagma verbal genera un segundo valle, entre el segundo y tercer pico, más pronunciado que el primero en las afirmativas, y menos en las interrogativas.

3. CONCLUSIONES

Del análisis y la comparación de las curvas entonativas en la oraciones afirmativas e interrogativas, con y sin expansión en el sintagma verbal, se desprende lo siguiente:

- El patrón entonativo de las oraciones enunciativas e interrogativas es distinto en las dos modalidades (con y sin expansión). Esta distinción aumenta a medida que la frase avanza.
- Destaca el brusco descenso del tonema en las oraciones interrogativas (más pronunciado incluso que el de las enunciativas), que se produce en las dos modalidades.
- Las oraciones con acento de frase agudo presentan un comportamiento más diferenciado: no desciende tanto el tonema de las interrogativas, y tiene un pico menos en las afirmativas con expansión.
- La curva melódica de las oraciones con y sin expansión es muy similar, especialmente en las interrogativas. La expansión del sintagma verbal genera un segundo valle, entre el segundo y tercer pico, más pronunciado que el primero en las afirmativas, y menos en las interrogativas.

BIBLIOGRAFÍA

DORTA, J.; HERNÁNDEZ, B.: "Análisis prosódico de un corpus de habla experimental: interrogativas absolutas con expansión en el objeto vs. sin expansión", *Estudios de Fonética Experimental*, XIV, 2005, en prensa.

ELEJABEITIA, A.; IRIBAR, A.; PAGOLA, R.M.: "Notas sobre la prosodia del castellano en Bizkaia", *Estudios de Fonética Experimental*, XIV, 2005, en prensa.

CALLEJA, N.: "Alineamiento fonético de acentos tonales en el castellano de Vitoria", *Estudios de Fonética Experimental*, XIII, 2004, pp. 39-63.

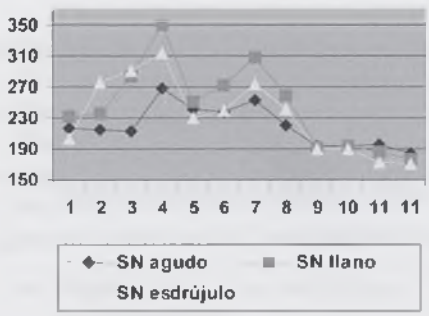
FACE, T.: Intonational marking of contrastive focus in Madrid Spanish, Berlin, Lincom-Europa, 2002.

TORRES ÁLVAREZ, M.ª I.: "Oraciones cortas vs. oraciones largas: diferencias en el patrón entonativo", *Revista de Filología de la Universidad de La Laguna*, 18, 2000, pp. 383-390.

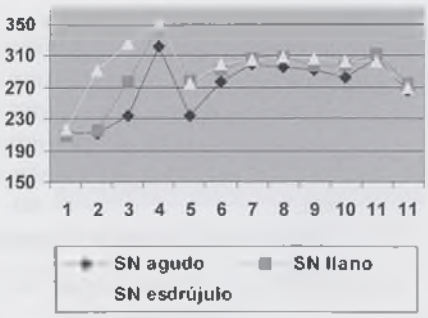
ANEXO

En la Figura 5 aparecen las curvas entonativas medias de las 18 frases sin expansión, correspondientes a las tres repeticiones de oraciones enunciativas e interrogativas de acento agudo, grave y esdrújulo. La columna izquierda recoge las oraciones enunciativas y la derecha, las interrogativas; las tres filas corresponden a los tres acentos de frase. En cada gráfica aparecen las tres curvas correspondientes a las tres posibilidades de acento inicial de cada frase. Los once puntos marcados en cada curva entonativa corresponden a los valores tonales de las once sílabas, obtenidos en el punto central de cada núcleo silábico. En el caso de las frases con acento final agudo, se han colocado dos valores para la última sílaba: el central y el final.

Frases agudas



Frases afirmativas



Frases interrogativas

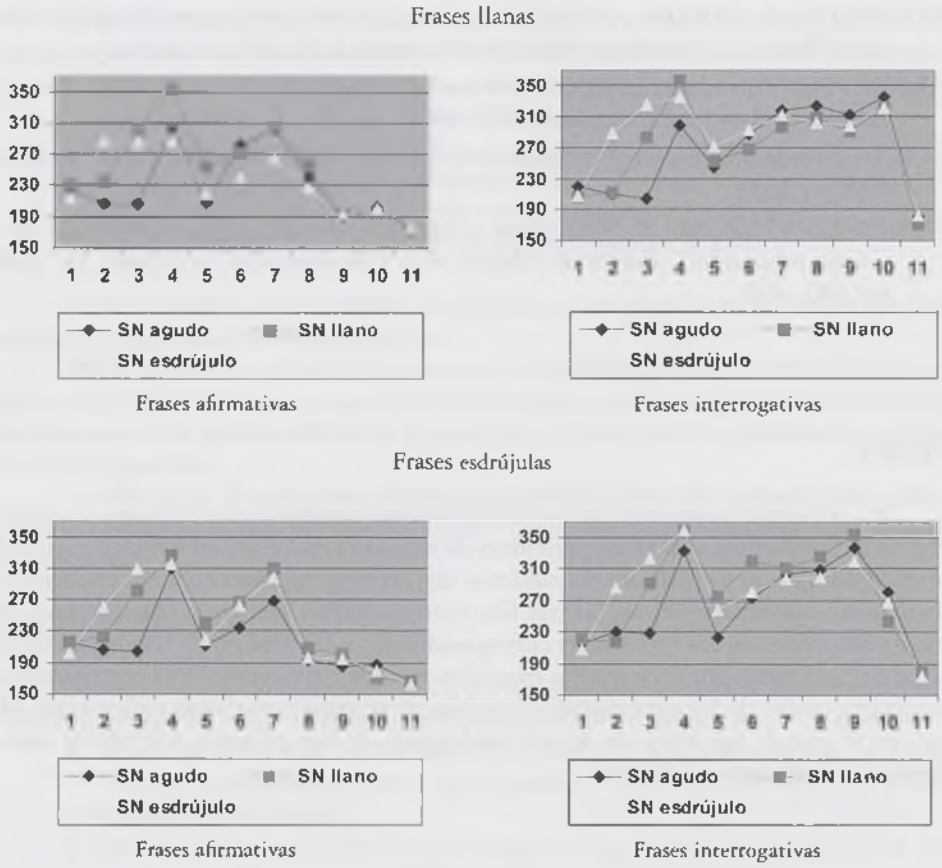


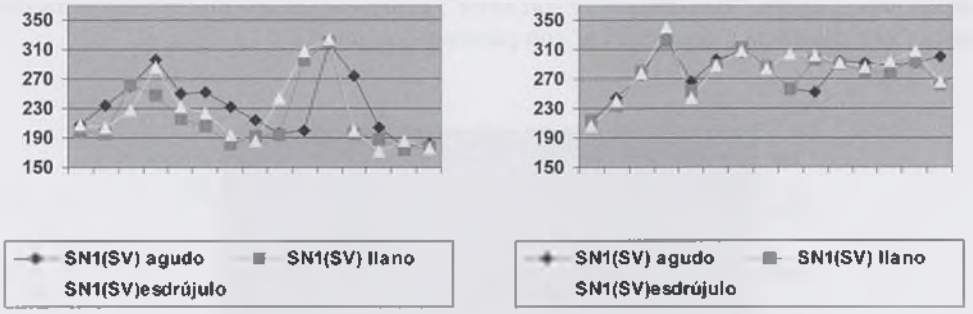
Figura 5.- Las 18 curvas sin expansión

En la Figura 6 aparecen las curvas entonativas medias de las 54 frases con expansión, agrupadas en 6 gráficos. Cada uno de ellos corresponde a 9 oraciones:

- "todo (el saxofón, la guitarra, la citara) se toca con obsesión/paciencia/pánico y con amor", afirmativas e interrogativas.[9]
- "todo se toca con obsesión/paciencia/pánico finito", afirmativas e interrogativas.
- "todo se toca con obsesión/paciencia/pánico práctico", afirmativas e interrogativas.

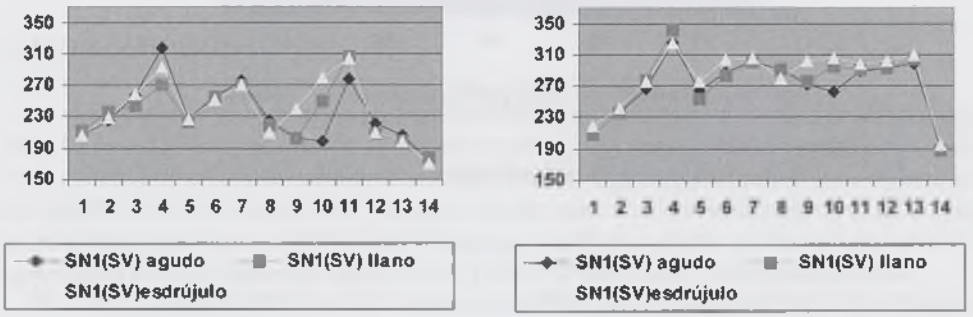
9 El primer gráfico contiene únicamente información de 6 oraciones, puesto que las tres restantes fueron anuladas durante el análisis (dos de primer acento de sintagma verbal llano y una de esdrújulo), al presentar ciertos problemas que aconsejaban su eliminación y una nueva grabación.

Frases agudas



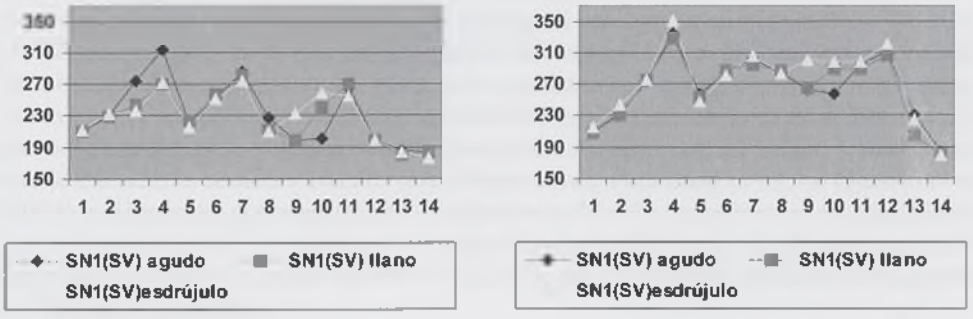
Frases afirmativas Frases interrogativas

Frases llanas



Frases afirmativas Frases interrogativas

Frases esdrújulas



Frases afirmativas Frases interrogativas

Fig. 6.- Las 18 curvas con expansión

La figura 7 muestra la ampliación de las últimas sílabas en las oraciones afirmativas con expansión aguda, que permite observar que las oraciones cuyo sintagma nominal expandido es agudo ("con obsesión y con amor") genera una sílaba más que los otros dos casos ("con paciencia y con amor" y "con pánico y con amor").

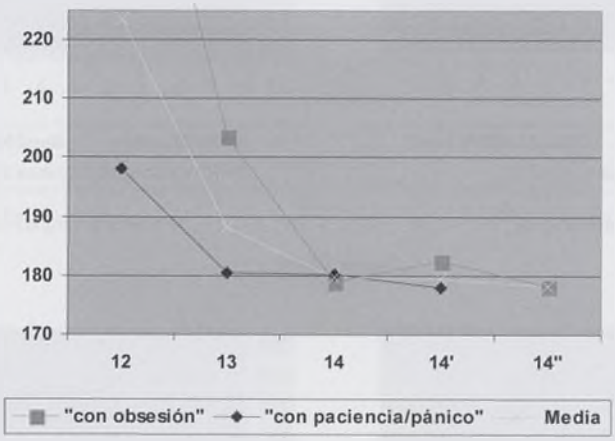


Fig. 7.- Ampliación de las últimas sílabas en las oraciones agudas afirmativas con expansión

APROXIMACIÓN AO ESTUDO DA FOCALIZACIÓN NO GALEGO

Ana Belén Escourido Pernas
Instituto da Lingua Galega

1. INTRODUCIÓN

Chomsky (1971) divide o discurso en dúas partes complementarias: *presupposition* e *focus*. O foco defínese, por tanto, como o elemento "non presuposto" da oración ou, máis ben, a parte que o falante considera descoñecida polo seu interlocutor. O foco é portador da información nova do discurso. No noso estudo, para determinar nas oracións do corpus cal é a parte presuposta e cal é o foco seguimos o método pregunta/resposta, do estilo do empregado noutros traballos como Frota (1998 e 2000) para o portugués europeo.

Unha visión tradicional da fonética e da fonoloxía considera que a distribución dos acentos tonais dunha cadea falada depende da súa estrutura sintáctica e vén asignada por unha serie de regras de acentuación involuntarias, como a formulada por Chomsky e Halle (1968) e coñecida como *Nuclear Stress Rule*, que determina como acento principal o último da oración. Este acento principal marcaría a información nova ou non referida no discurso, que xeralmente, no inglés, está asociada á posición final. No entanto, Bolinger (1972) advirte que en moitos casos o patrón de acentos dun enunciado pode depender da intencionalidade dos falantes, neste caso a posición que ocupan estes acentos é significativa; vén dada pola súa importancia semántica e pola elección do falante. Por tanto, a prosodia é empregada voluntariamente para poñer en relevo certas palabras do discurso que se consideran semanticamente máis importantes. Para Ladd (1983) e Gussenhoven (1984) a asignación dos acentos non é unha mera elección entre marcar ou non marcar prosodicamente unha ou varias palabras do discurso, senón que vén determinada pola presenza ou ausencia do trazo semántico [±foco], que determina a aparición dos acentos na realización prosódica.

Dende a perspectiva da fonética acústica trataremos de ver neste estudo en que medida os parámetros de frecuencia fundamental, duración e enerxía nas oracións reflic-

ten o realzado prosódico dun determinado constituínte e como contribúen a diferenciar este tipo de enunciados das oracións neutras ou non focalizadas.

2. OBXECTIVOS

Ata o de agora non existen investigacións sobre fonética galega centradas na manifestación prosódica do foco, fenómeno que si chamou a atención a estudosos doutras linguas como o inglés, o español ou o portugués. No entanto, son escasos os estudos especializados na focalización noutras linguas, e a maioría son mencións máis ou menos exhaustivas en traballos máis amplos sobre entoación.

Dada a brevidade do estudo e o escaso número de informantes empregadas, a nosa intención era, *a priori*, facer unha primeira achega á entoación marcada por focalización e non facer unha verdadeira descrición da realización do foco no galego. Por tanto, os datos achegados só pretenden ser rexistros de tendencias comúns que poidan servir de base a posteriores investigacións máis extensas sobre a focalización.

3. METODOLOXÍA

Gravamos dúas mulleres novas de entre 20 e 30 anos, nacidas na Comarca de Bergantiños, que están realizando ou xa completaron os seus estudos superiores.

Empregamos unha gravadora dixital portátil –DAT-, modelo *TCD-D10 PRO II* e un micrófono *Shure* de cabeza, externo e unidireccional, propiedade do Instituto da Lingua Galega. Para a dixitalización das gravacións e o tratamento do sinal acústico utilizamos o programa GoldWave 4.25 e para as análises MatLab 6.5.

Figura 1: Ficha que corresponde á oración *O médico falaba ca*[1] *rapaza*

Mostramos a cada informante seis fichas de debuxos diferentes (figura 1). Cada ficha representa unha mesma oración que as informantes debían entoar estimuladas por tres preguntas que lles realizamos previamente.

1 Empregamos como representación gráfica da preposición e o artigo a forma máis próxima á realización da área dialectal á que pertencen as informantes e non a normativa do idioma, que sería *coa*.

Estas preguntas buscan tres tipos de contextos diferentes. Por un lado, a entoación non marcada ou neutra, denominada *broad focus* ou foco ancho, onde a totalidade da oración constitúe información nova ou non presuposta polo oínte. Por outro lado, a entoación marcada como información nova, *narrow focus* ou foco estreito, onde unicamente un constituínte da oración se presenta como elemento novo ou non presuposto polo oínte. Por último, a entoación marcada por oposición, que é un tipo de foco estreito onde un constituínte da oración confirma ou opón a información coñecida coa non coñecida.

As fichas de debuxos foron elaboradas tendo en conta, por un lado, a localización do constituínte focalizado -que ocupara a posición inicial ou final da oración-, e por outro lado, a configuración acentual oxítona, paroxítona e proparoxítona do elemento marcado da oración.

ESQUEMA A	O +	MÉDICO SOLDADO CAPITÁN	+ fala ca rapaza
ESQUEMA B	A rapaza encontrou un	+	CÁNTARO PENDENTE ALACRÁN

Así, empregamos un Corpus formado por oracións sinxelas, enunciativas, coa estrutura *Suxeito+Verbo+Complemento*, que seguen dous esquemas diferentes: o A que focaliza o suxeito e o B que focaliza o complemento.

No cadro 1 e 2 mostramos a totalidade das oracións recollidas no noso Corpus coa pregunta contextualizadora que lle corresponde. As maiúsculas sinalan o foco na oración.

Suxeito	1. Que pasa no debuxo?	O médico xoga ca rapaza
	2. Quen xoga ca rapaza?	O MEDICO xoga ca rapaza
Proparoxítono	3. O pai xoga ca rapaza?	O MÉDICO xoga ca rapaza
Suxeito	1. Que pasa no debuxo?	O soldado xoga ca rapaza
	2. Quen xoga ca rapaza?	O SOLDADO xoga ca rapaza
Paroxítono	3. O pai xoga ca rapaza?	O SOLDADO xoga ca rapaza
Suxeito	1. Que pasa no debuxo?	O capitán xoga ca rapaza
	2. Quen xoga ca rapaza?	O CAPITÁN xoga ca rapaza
Oxítono	3. O pai xoga ca rapaza?	O CAPITÁN xoga ca rapaza

Cadro 1: Corpus con focalización no suxeito

Complemento Proparoxítono	1. Que pasa no debuxo?	A rapaza encontrou un cántaro
	2. Que encontrou a rapaza?	A rapaza encontrou un CÁNTARO
	3. A rapaza encontrou un cordel?	A rapaza encontrou un CÁNTARO
Complemento Paroxítono	1. Que pasa no debuxo?	A rapaza encontrou un pendente
	2. Que encontrou a rapaza?	A rapaza encontrou un PENDENTE
	3. A rapaza encontrou un cordel?	A rapaza encontrou un PENDENTE
Complemento Oxítono	1. Que pasa no debuxo?	A rapaza encontrou un alacrán
	2. Que encontrou a rapaza?	A rapaza encontrou un ALACRÁN
	3. A rapaza encontrou un cordel?	A rapaza encontrou un ALACRÁN

Cadro 2: Corpus con focalización no complemento

4. DATOS ACADADOS

Observaremos primeiro o comportamento da frecuencia fundamental, onde encontramos os datos máis relevantes, e atenderemos posteriormente ás tendencias mostradas na duración e na enerxía. Os datos que mostramos nos seguintes gráficos representan realizacións concretas das informantes analizadas. Para elaborar os gráficos de F0, fixemos unha única medición por cada sílaba. No eixo horizontal móstrase o número de sílaba e no eixo vertical o valor da frecuencia en *hertz*. Nos gráficos de duración, o eixo horizontal representa o número de sílaba e o vertical o valor da duración en milisegundos.

4.1. Frecuencia fundamental

Como a intención do noso estudo é mostrar condutas comúns sobre a focalización e non unha descrición exhaustiva, interésanos comezar cos exemplos onde encontramos as realizacións máis relevantes e peculiares do fenómeno. É por iso que, neste epígrafe, presentamos primeiramente o foco contrastivo, aínda que sexa un tipo de foco estreito, porque nos permite describir con maior claridade as diferentes tendencias. Móstranse posteriormente os datos sobre o foco estreito e en que medida varían dos achegados para o contrastivo. Por último, observaremos cal é o comportamento das realizacións neutras, se seguen a tendencia xeral do galego e en que se diferencian das oracións marcadas.

4.1.1. Foco contrastivo

Na figura 2 mostramos o gráfico correspondente aos enunciados con foco contrastivo ao inicio, xa que é no suxeito onde se atopa a información nova que se opón á información dada na pregunta. Son aquelas oracións que, como indicamos no punto 3, presentan o esquema A:

O +	MÉDICO SOLDADO CAPITÁN	+ fala ca rapaza

Este tipo de foco presenta un único pico tonal relevante nas primeiras sílabas átonas do suxeito. A partir deste máximo tonal da oración prodúcese un descenso intenso da F0 nas últimas sílabas do suxeito. O predicado da oración non mostra alteracións na liña descendente, senón máis ben unha inmobilidade final, que parece reforzar a atención sobre a parte inicial focalizada da oración. Non aparece, por tanto, o último acento do enunciado que nas oracións non marcadas está asociado á última sílaba tónica.

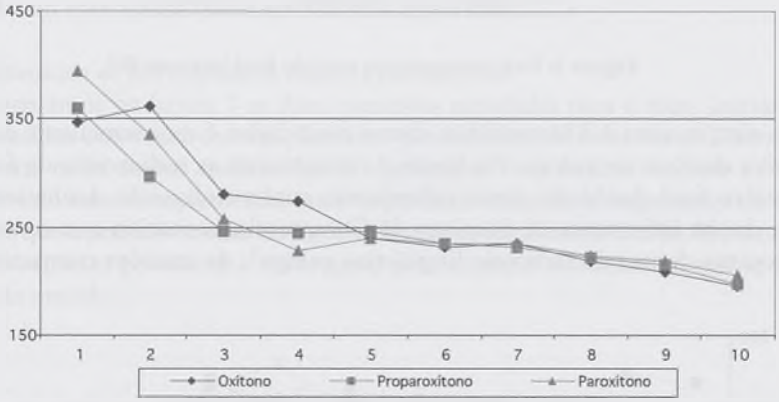


Figura 2: Foco contrastivo en posición inicial (esquema A)

O comportamento do foco contrastivo(figura 3 e 4) é semellante naquelas oracións con esquema B, nos que a parte focalizada se encontra no complemento:

A rapaza encontrou un	+	CÁNTARO PENDENTE ALACRÁN

Este tipo de realizacións comeza nunha frecuencia elevada que se mantén sen movementos tonais relevantes ata o final. No complemento prodúcese un pequeno ascenso nas sílabas pretónicas e posteriormente sobre a tónica ten lugar unha caída abrupta alcanzando o baixo ton final da oración.

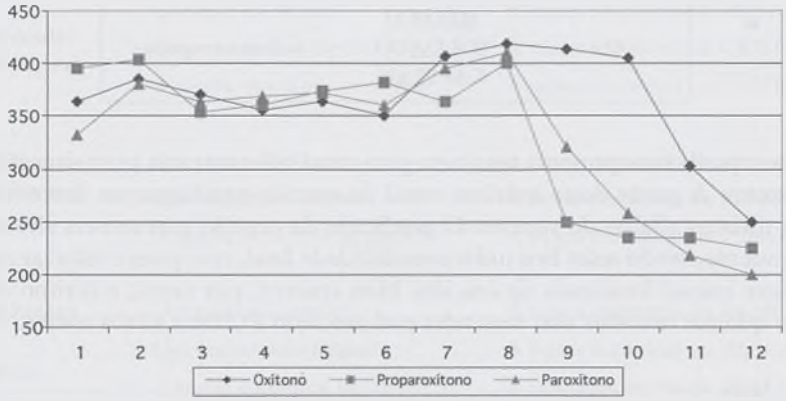


Figura 3: Foco contrastivo en posición final (esquema B)[2]

A simple vista a liña melódica destes enunciados é moi semellante á dunha interrogativa absoluta en galego. Na figura 4 comparamos as realizacións de foco estreito e contrastivo final dunha das nosas informantes cunha realización dunha interrogativa absoluta dunha informante de Santiago de Compostela, entoación que consideramos a máis frecuente dentro do territorio lingüístico galego[3]. As oracións comparadas son di-

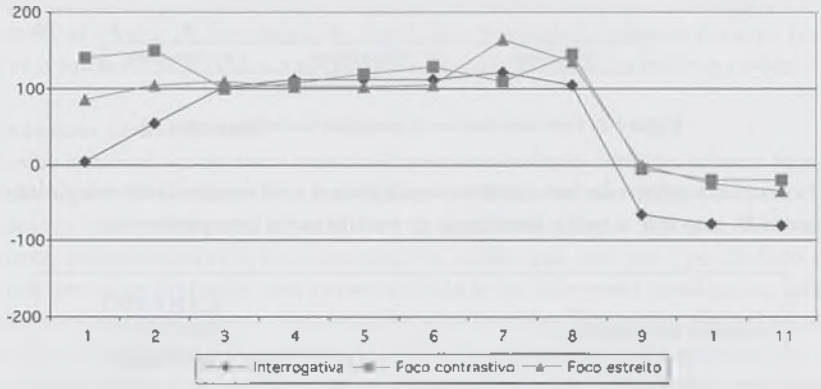


Figura 4: Comparación entre unha interrogativa e o foco contrastivo e estreito en posición final[4]

2 Como sílaba 12, indicamos o valor do último punto medido na vogal final do complemento, dada a súa relevancia no complemento oxítono onde se produce un marcado descenso.

3 Datos publicados en Fernández et alii. (2005a).

4 O gráfico débese tomar con certa cautela, posto que comparamos informantes con diferentes frecuencias fundamentais

ferentes no predicado pero respectamos a configuración acentual (suxeito paroxítono e complemento proparoxítono). A interrogativa de Santiago corresponde ao enunciado *A rapaza xogaba co dálmata* mentres que a oración dos dous tipos de foco é *A rapaza encontrou un cántaro*.

As tres realizacións teñen un comportamento similar dentro do predicado, pero no suxeito si encontramos diferenzas significativas. A interrogativa comeza nun ton máis baixo que as focalizadas, ascende nas sílabas do suxeito e alcanza certa altura que mantén ata o complemento. As oracións marcadas tanto con foco estreito como contrastivo iníciánse nun ton moi alto que manteñen ou ben diminúen lixeiramente ata o final. No complemento prodúcese nas tres realizacións unha caída brusca de ton sobre a sílaba tónica, porén na modalidade interrogativa este descenso é algo maior. A vista dos datos sería interesante facer un test perceptivo con sons sintetizados das tres mostras e ver en que medida son entendidos como realizacións diferentes.

4.1.2. Comparación do foco en posición inicial e posición final

Comparamos na figura 5 as dúas posicións estudadas para o foco, inicial e final, naquelas oracións con suxeito e complemento paroxítono. Na informante 1 vemos unha clara diferenza entre ambas realizacións: a oración con foco ao final iníciase e mantense nun ton elevado ata o complemento, onde se produce unha caída importante de frecuencia, mentres que a oración co suxeito focalizado comeza nun nivel alto de frecuencia, pero xa dentro do suxeito prodúcese a caída tonal para alcanzar o rexistro no que se mantera ata o final da oración.

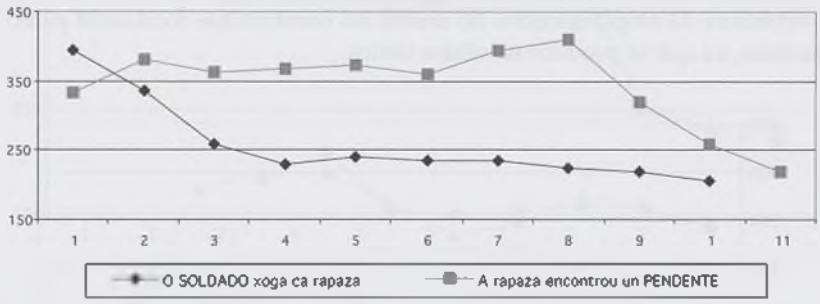


Figura 5: Foco contrastivo no suxeito e no complemento

medias (FFM). Para paliar este problema igualamos a cero as súas FFM para contrastar os movementos da liña melódica na mesma escala.

4.1.3. Foco estreito

No foco estreito tanto en posición inicial como en posición final a liña melódica segue o mesmo deseño observado para o foco contrastivo.

Na figura 6 mostramos un exemplo dunha oración con foco no suxeito oxítono (esquema A), onde as frecuencias máis elevadas están asociadas ás primeiras sílabas átonas do suxeito. Prodúcese posteriormente a caída de frecuencia na terceira sílaba (segunda pretónica) e a inmobilidade tonal do predicado, coa desaparición do acento final do enunciado.

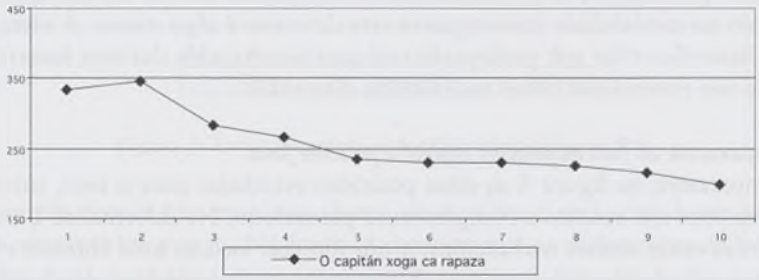


Figura 6: Foco estreito en posición inicial (esquema A)

En posición final (esquema B), como se aprecia na figura 7, a caída brusca de frecuencia prodúcese dentro do complemento, despois dun pequeno ascenso de ton nas súas sílabas pretónicas. O emprazamento do acento no constituínte focalizado parece influír neste descenso, xa que se produce na sílaba tónica.

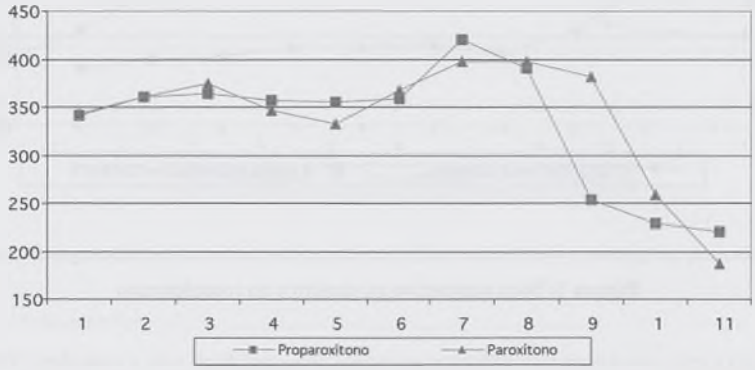


Figura 7: Foco estreito en posición final (esquema B)

4.1.4. Comparación entre foco estreito e foco contrastivo

Nas figuras 8 e 9, onde comparamos o foco estreito coa súa variante contrastiva, vemos como a liña de entoación é practicamente igual tanto na focalización do suxeito (figura 8) como na do complemento (figura 9).

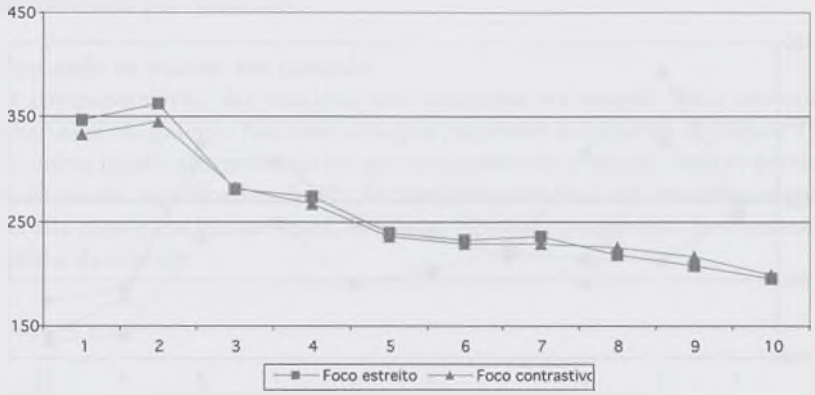


Figura 8: Foco estreito e contrastivo en posición inicial (suxeito oxítono)

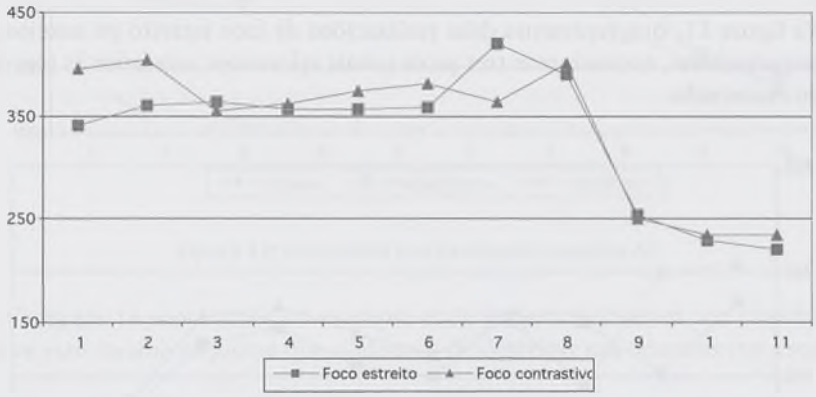


Figura 9: Foco estreito e contrastivo en posición final (complemento proparoxítono)

4.1.5. Outras realizacións do foco

Dentro do foco en posición inicial, rexistramos outro tipo de exemplos que non incluímos en epígrafes anteriores porque obedecen a intencións comunicativas diferentes por parte das falantes e non ao modelo esperable que buscabamos coas preguntas contextualizadoras. Estas realizacións presentan dúas ou incluso tres elevacións tonais, focali-

zando o complemento ou o verbo máis o complemento da mesma maneira que o suxeito (que era a información nova que se solicitaba a partir das preguntas: *Quen xoga ca rapaza?* ou *O pai xoga ca rapaza?*).

Na figura 10, onde mostramos tres realizacións de foco contrastivo, vemos como se producen dous picos tonais relevantes, un no suxeito e outro no complemento.

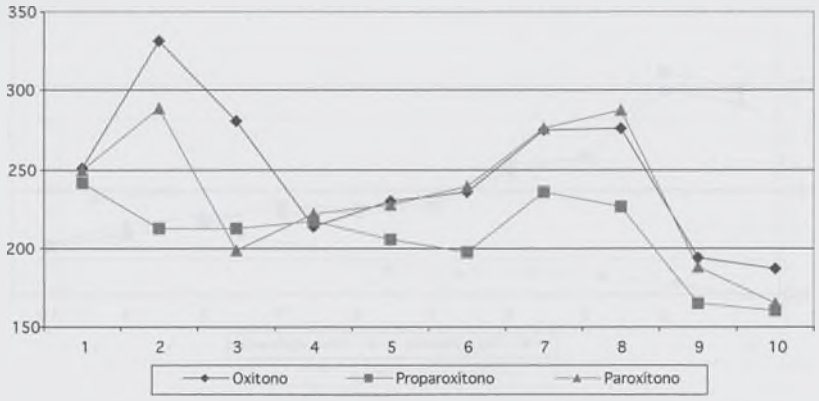


Figura 10: Foco contrastivo en posición inicial (esquema A)

Na figura 11, que representa dúas realizacións de foco estreito en suxeito paroxítono e proparoxítono, encontramos tres picos tonais relevantes, asociados ás tres palabras léxicas do enunciado.

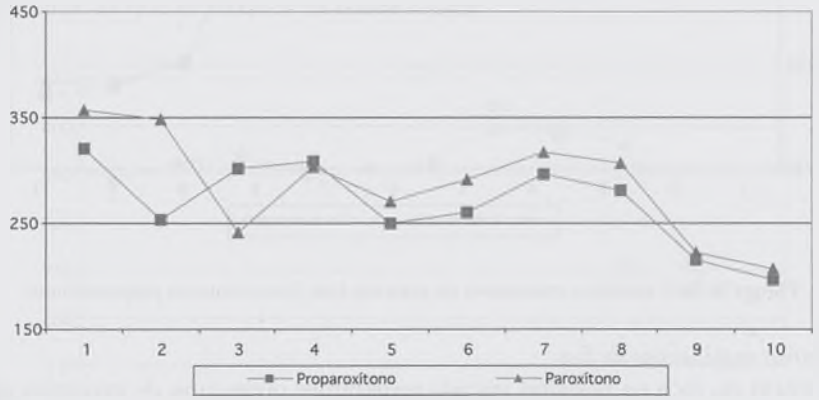


Figura 11: Foco estreito en posición inicial (esquema A)

A percepción auditiva que nos producen estas mostras é de reiteración ou imitación dun mesmo esquema inicial. A informante parece evitar con esta entoación posteriores malentendidos no seu interlocutor, marcando prosodicamente a información xa coñecida. Entendemos que a focalización que se produce no complemento (figura 10) ou en verbo máis complemento (figura 11) é un tipo de foco contrastivo, que funciona por confirmación e non por oposición.

4.1.6. Foco ancho ou oracións non marcadas

O comportamento das oracións non marcadas ou neutras (foco ancho) segue o modelo habitual do galego. Nas oracións que presentan o esquema A (figura 12), a frecuencia comeza nunha altura media na que se mantén ata a tónica, onde se produce unha primeira elevación significativa da F0. O complemento final da oración, independentemente da súa configuración acentual, inicia un descenso progresivo de frecuencia cara á última sílaba da oración.

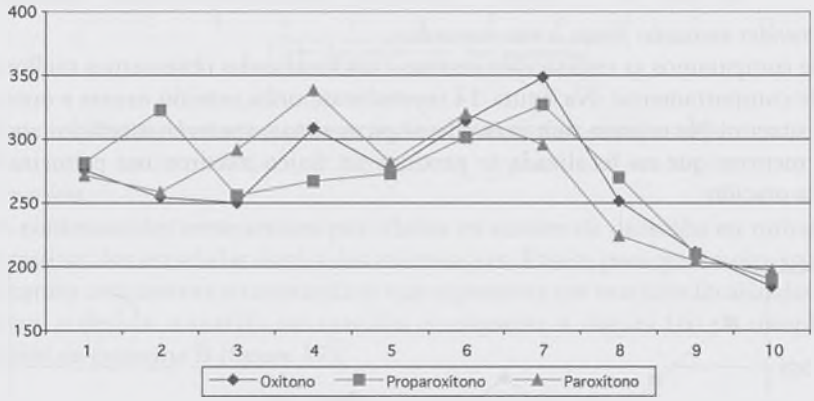


Figura 12: Enunciados non focalizados (esquema A)

Na figura 13 mostramos un exemplo onde podemos observar con claridade como se cumpre este mesmo esquema que acabamos de describir nas oracións con esquema B.

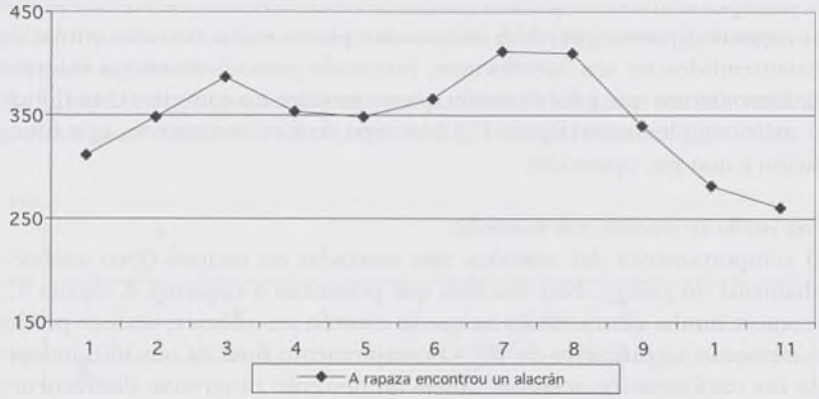


Figura 13: Enunciados non focalizados (esquema B)

4.1.7. Oracións marcadas fronte a non marcadas:

Se comparamos as realizacións neutras coas focalizadas observamos mellor as diferenzas de comportamento. Na figura 14 represéntase unha oración neutra e outra focalizada no suxeito. Na oración non marcada os picos tonais aparecen asociados aos acentos léxicos, mentres que na focalizada se produce un único máximo nas primeiras sílabas átonas da oración.

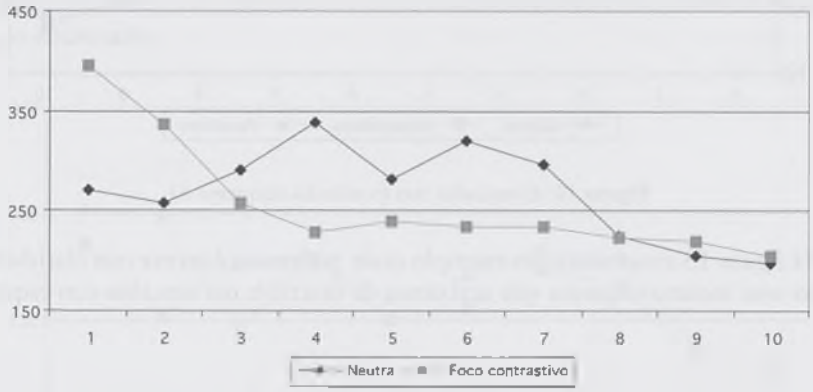


Figura 14: *O soldado xoga ca rapaza* non marcada e con foco inicial

Na figura 15, onde comparamos a realización dunha oración neutra con outra con foco ao final, encontramos diferenzas importantes tanto no suxeito coma no complemento. No inicio da oración non marcada a frecuencia, que comeza a unha altura relativa-

mente baixa, debe ascender, mentres que na focalizada a tendencia da F0 é manter o ton inicial alto da oración. No complemento prodúcese un descenso da tonalidade nas dúas oracións, pero só na oración focalizada parece influír a configuración acentual do complemento, porque a sílaba tónica asóciase a un descenso brusco da F0.

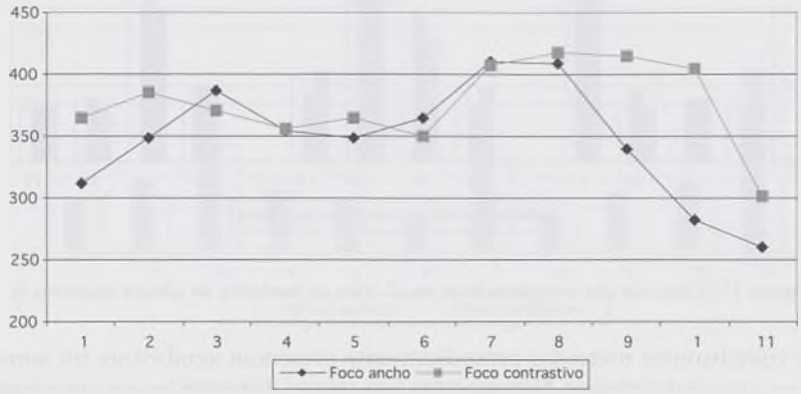


Figura 15: *A rapaza encontrou un alacrán* non marcada e con foco final

4.2. Duración

A continuación, presentamos por sílabas os valores da duración en milisegundos nas tres realizacións estudadas dunha das informantes. Porén, para que resulte máis gráfico, mostramos unicamente o constituínte que representa nas oracións focalizadas a información non coñecida: o suxeito nas oracións co esquema A (figura 16) e o complemento nas oracións co esquema B (figura 17).

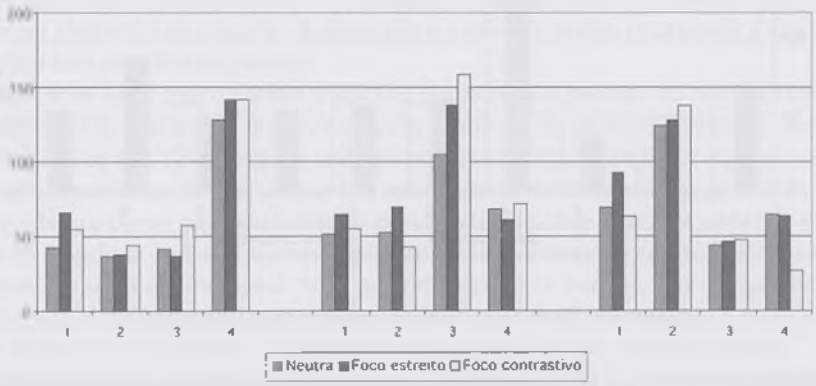


Figura 16: Duración dos suxeitos: *o capitán*, *o soldado* e *o médico* (esquema A)

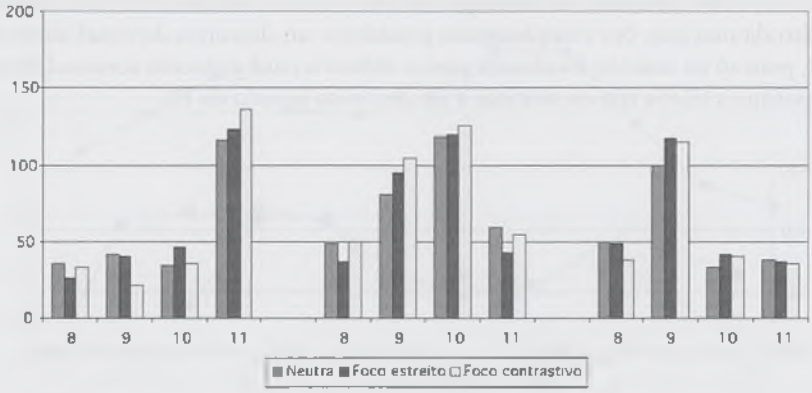


Figura 17: Duración dos complementos: *un alacrán*, *un pendente* e *un cántaro* (esquema B)

Os constituíntes marcados prosodicamente presentan xeralmente un aumento de duración na súa sílaba tónica. Nas oracións neutras, a duración marca un acento nesta sílaba, como é esperable, pero é menor ao das oracións focalizadas.

Por outro lado, polo que respecta a toda a oración, a sílaba tónica do constituínte focalizado en posición inicial é sempre a de maior duración, superando incluso á última tónica do enunciado. Vémolo exemplificado na figura 18, na que representamos a duración de todas as sílabas nas oracións con suxeito oxítono focalizado.

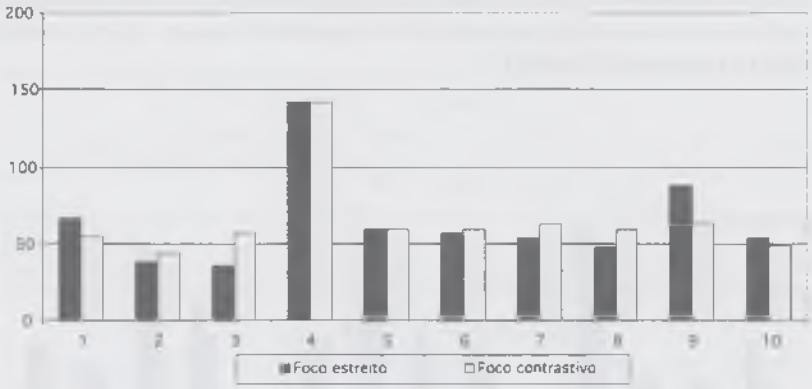


Figura 18: Duración en *O capitán xoga ca rapaza* (esquema A)

En posición final, a sílaba tónica do constituínte focalizado, asociada ademais ao núcleo tonal da oración, é tamén a máis longa de todo o enunciado. Mostramos como

exemplo, na figura 19, a duración de todas as sílabas das oracións con foco en complemento oxítono.

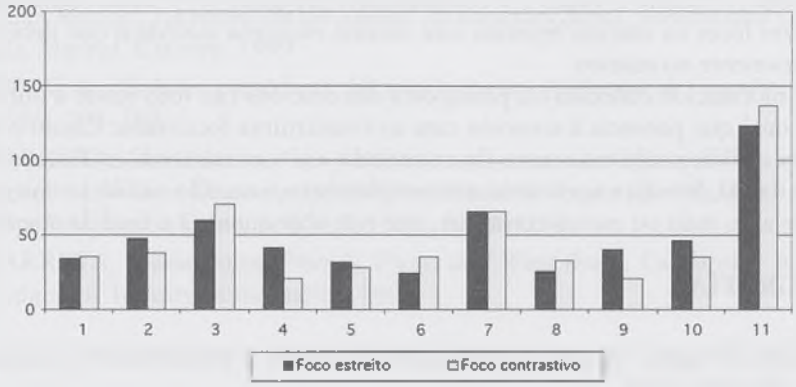


Figura 19: Duración en *A rapaza encontrou un alacrán*

4.3. Enerxía

Nos enunciados que analizamos non parece relevante o parámetro da enerxía para marcar os enunciados con foco. Nas tres entoacións estudadas (enunciados non focalizados, enunciados con foco estreito e enunciados con foco contrastivo) as sílabas que presentan maior intensidade xeralmente son aquelas que tamén mostran un aumento de frecuencia, e pola contra, as menos intensas son as que se realizan nunha tonalidade baixa.

5. CONCLUSIÓN

A frecuencia fundamental é o parámetro máis relevante nos casos analizados para focalizar un elemento da oración. A duración e a enerxía poden contribuír á súa identificación pero non son determinantes.

Podemos dicir que o patrón tonal das enunciativas neutras ou non focalizadas das dúas informantes analizadas segue o modelo habitual do galego, acorde co descrito en Fernández Rei et alii. (2005a) para unha estrutura sintáctica similar á aquí estudada. As variacións importantes de frecuencia nas enunciativas están asociadas ás sílabas tónicas, tanto no suxeito como no complemento non focalizado. Nas oracións neutras, na tónica do suxeito é onde se inicia o primeiro pico tonal importante da oración, alcanzando habitualmente o seu máximo tonal na(s) postónica(s), se as houber. No complemento, hai unha segunda elevación tonal para marcar o último acento de enunciado e caída progresiva final de ton que non parece depender da configuración acentual da palabra.

Porén o constituínte focalizado, tanto se se encontra ao inicio coma ao final da oración, caracterízase por un aumento de frecuencia nas sílabas pretónicas, inmediatamente

seguido por unha caída importante da tonalidade na segunda ou terceira sílaba -tónica ou átona-. Esta conduta é a mostrada tamén naqueles exemplos nos que, sendo o suxeito unicamente a información que se requiría como nova, púñase en relevo de igual xeito o complemento e mesmo o verbo e o complemento (epígrafe *4.5*). Estas realizacións con dous ou tres focos na oración repetían este mesmo esquema melódico que debía encontrarse unicamente no suxeito.

A información coñecida ou presuposta das oracións con foco tende a unha inmobilidade tonal que potencia a atención cara ao constituínte focalizado. Cando o foco vai no suxeito anúlase o último acento do enunciado e o ton mantense en frecuencias moi baixas ata o final. Se o foco se encontra no complemento, a oración sitúase no inicio nunha tonalidade alta, máis ou menos constante, que non abandona ata o final da oración.

BIBLIOGRAFÍA

BOLINGER, Dwight, "Accent is predictable (If your're a mind-reader)", *Language* 48, 3, 633-644, 1972.

BOLINGER, Dwight e Marion Hodap, "Acento melódico. Acento de intensidad", *Boletín de Filología*, Tomo XIII, 1961, 35-50.

CHOMSKY, Noam e Morris Halle, *The sound pattern of English*, New York, Harper & Row, 1968.

CHOMSKY, Noam, "Deep structure, surface structure and semantic interpretation", in Steinberg, D. D. e L.A. Jakobivits (eds.): *Semantics: an interdisciplinary reader in philosophy, linguistics and psychology*, UK: Cambridge University Press, 1971, 183-216.

FERNÁNDEZ, Elisa et alii., "Achega á entoación dunha fala do centro de Galicia. Contribución para o AMPER", *Géolinguistique*, Hors de Série, 3, 2005a, 87-102.

FERNÁNDEZ, Elisa et alii., "A entoación dunha fala de Santiago: fronteira prosódica entre suxeito e predicado", *Estudios de Fonética Experimental*, vol. XIV, 2005b, 141-165.

FROTA, Sónia, *Prosody and Focus in European Portuguese*, tese de doutoramento, Facultade de Letras da Universidade de Lisboa, 1998 (inédita).

FROTA, Sónia, *Prosody and Focus in European Portuguese*, New York & London, Garland Publishing, 2000.

GUSSENHOVEN, Carlos, *On the Grammar and Semantic of Sentence Accents*, Dordrecht, Foris Publications, 1984.

LADD, D. Robert, "Even, Focus, and Normal Stress", in Journal of *Semantics*: 2, 2, 1983, 157-170.

LADD, D. Robert, *Intonational phonology*, Gran Bretaña, Cambridge University Press, 1996.

NESPOR, Marina e Irene Vogel, *Prosodic Phonology*, Dordrecht, Foris Publications, 1986. Tradución española: *La Prosodia*, Madrid, Visor,1994.

PRIETO, Pilar, *Entonació. Models, teoria, mètodes*, Barcelona, Ariel Lingüística, 2002.

SOSA, Juan Manuel, *La entonación del español. Su estructura fónica, variabilidad y dialectología*, Madrid, Cátedra, 1999.

TOLEDO, G.A., "Señales prosódicas del foco", *Revista Argentina de Lingüística*: 5, 1-2, 1989, 205-230.

ZUBIZARRETA, María Luisa, "Las funciones informativas: tema y foco", in Bosque, Ignacio e Violeta Demonte (eds): *Gramática de la lengua española*, Real Academia Española/Espasa Calpe, 3, §64.2, 1999, 4218-4242.

ZUBIZARRETA, María Luisa, *Prosody, Focus, and Word Order*, Cambridge & London, Linguistic Inquiry Monographs, 1998.

PRODUCTION AND PERCEPTION OF WORD-EDGE TONES IN CATALAN AND SPANISH

EVA ESTEBAS-VILAPLANA
Universidad Nacional de Educación a Distancia
PILAR PRIETO
ICREA-UAB

1. INTRODUCTION

The aim of this paper is to clarify the role of word-edge tones in prenuclear rising accents (LH) in Catalan and Spanish broad focus declaratives. The modelling of prenuclear rises in different languages within the Autosegmental-Metrical approach of intonational analysis has shown that whereas the F0 valley presents a stable alignment and scaling, the anchoring of H is much more unprecise and prompts different phonological interpretations. Spanish prenuclear accents have been classified differently in the literature: H* with a peak delay (Prieto et al. 1995), L*+H (Sosa 1999, Face 1999 and Beckman et al. 2002) or (L+H)* (Hualde 2002). In Catalan (Estebas-Vilaplana 2000), prenuclear rises have been described as instances of a low pitch accent (L*) followed by a word edge tone (H). In this study a production test and a perceptual test were designed for both Catalan and Spanish so as to check whether the anchoring of H to the end of the word is consistently produced by speakers to disambiguate the location of the word boundary and also whether this H alignment is a helpful perceptual cue to identify word boundaries.

2. PRODUCTION TEST

The aim of the production test is to examine whether H anchoring is used to disambiguate sentences. The hypothesis is that if an H word edge tone is present the F0 peak should be aligned at the end of words regardless of the number of postaccentual syllables.

2.1. Experimental procedure

The data used in the two production tests consisted of 20 pairs of potentially ambiguous sentences which had the same segmental and stress composition and were *only* distinguished by word boundary location. Spanish sentences included words with three stress

distributions: *oxytones*, *paroxytones* and *proparoxytone*, as for example, *ve bovinos* '(s)he sees cows' vs. *bebo vinos* '(s)he drinks wines' or *sube Melino* 'Melino goes up' vs. *súbeme lino* 'bring up the linen for me'. In Catalan only oxytones and paroxytones were examined, as in *compra ventalls* '(s)he bought fans' vs. *compraven talls* 'they bought pieces'. Three speakers for each language read the utterances four times (360 utterances for each language). Sentences were analysed by means of *Praat*. Measurements were made on simultaneous displays of speech wave, wide-band spectrogram and F0 tracks. For each sentence, the following segmental and pitch landmarks were manually placed in the two test syllables: 1) Segmental landmarks: a) beginning of onset and vowel on the target accented syllable, b) beginning of onset, vowel, and coda (whenever present) on the target postaccentual syllable, c) end of the postaccentual syllable, d) end of the onset of the following syllable and e) phrase-final point. 2) Pitch landmarks: a) L1, valley of the first pitch accent (in Hz) and b) H1, peak of the first pitch accent (in Hz). Figure 1 shows the Praat windows with the segmentation of the Spanish utterances *ve bovinos* and *bebo vinos*. The first two boxes displayed in each graph show the speech waveform and a spectrogram with an overlapped F0 trace. The last two boxes show the segmental landmarks with the segmental boundaries of the two test syllables and the pitch landmarks.

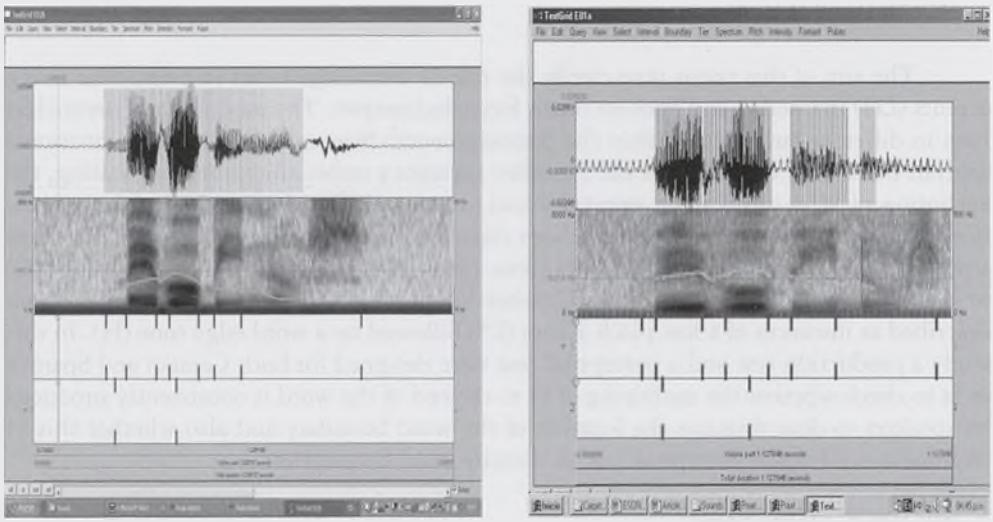


Figure 1. Diplays of speech waveform, spectrogram and F0 trace, and the segmental and pitch landmarks for the Spanish utterances *Ve bovinos* '(s)he sees cows' vs. *Bebo vinos* 'I drink wines'.

2.2. Results

2.2.1. H peak delay as a function of the duration of the accented syllable

The graphs displayed in Figures 2 and 3 plot H peak delay (or distance between the onset of the accented syllable and the F0 peak in ms) as a function of the duration of

the accented syllable for the three Catalan and the three Spanish speakers respectively. The regression lines summarize the strong positive correlations found between the two variables both in Catalan (correlation coef. 0,67-0,82) and in Spanish (correlation coef. 0,42-0,91).

Figure 2. Peak delay as a function of the duration of the accented syllable in ms in the two stress conditions (word-final vs. word-medial stress) in Catalan for speakers AG, PG, and PP.

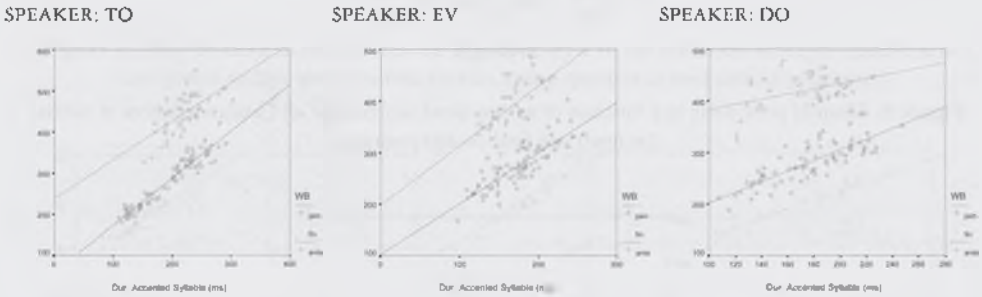


Figure 3. Peak delay as a function of the duration of the accented syllable in ms in three stress conditions (words with final, penultimate and antepenultimate stress) in Spanish for speakers TO, EV, and DO.

The results of Figures 2 and 3 also show that there is a consistent difference in H delay depending on the prosodic condition in both languages as the data are visually separated into two groups in Catalan and into three groups in Spanish. Hs in words with a final accent are less delayed than Hs in words with a penultimate accent. Similarly, in Spanish, Hs in words with a penultimate accent are less delayed than Hs in words with an antepenultimate accent. This seems to indicate that the location of the H is highly influenced by within-word position.

2.2.2. Effects of within-word position on H location

The graphs displayed in Figures 4 and 5 plot the mean H peak delay or distance in ms between the H peak relative to the beginning of the accented syllable in words with final (fin) and medial (med) stress for Catalan (Figure 4) and in words with fin, penultimate (pen) and antepenultimate (ante) stress for Spanish (Figure 5). The bars represent standard errors.

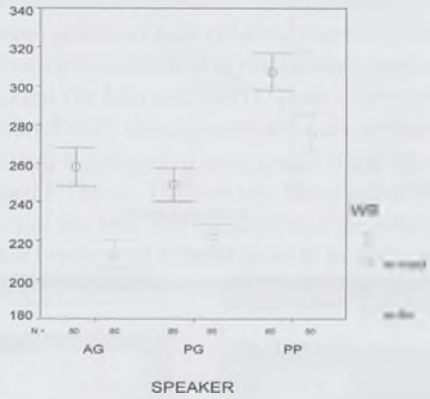


Figure 4. Mean H peak delay as a function of within-word position for all Catalan speakers in medial (w-med) and final (w-fin) position.

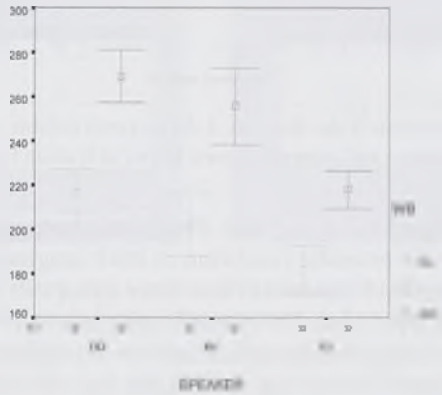


Figure 5. Mean H peak delay as a function of within-word position for all Spanish speakers: fin vs pen (graph on the left) and pen vs ante (graph on the right).

For both languages, the data reveal consistent effects of within-word position on H delay. For the three Catalan speakers, peak delay is significantly shorter ($p<0,001$) in word-final syllables (w-fin) than in word-medial syllables (w-med). For the Spanish data, peaks are less displaced in words with a final accent than in words with an accent on the penultimate (left graph) or antepenultimate syllables (right graph). Thus, F0 peaks are more displaced in paroxytones than in oytones and even more displaced in proparoxytones than in paroxytones.

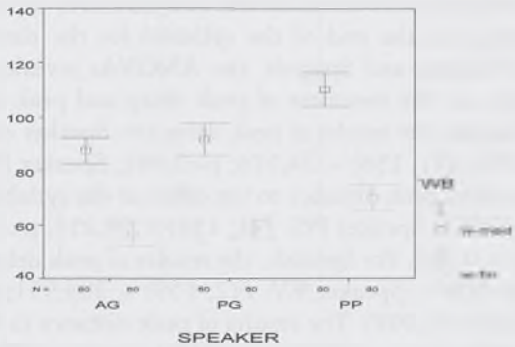


Figure 6. Mean distance in ms between the H peak relative to the end of the accented syllable as a function of within-word position for all Catalan speakers in med and fin positions.

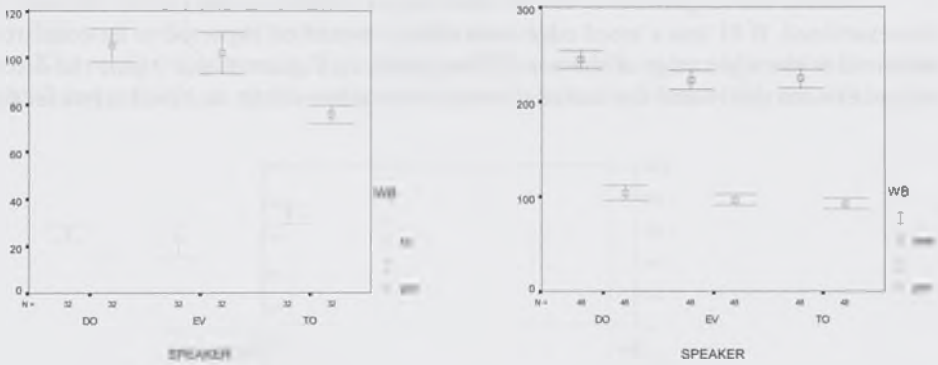


Figure 7. Mean distance in ms between the H peak relative to the end of the accented syllable as a function of within-word position for all Spanish speakers: fin vs pen (graph on the left) and pen vs ante (graph on the right).

The results in Figures 6 and 7 show the distance between the location of the H relative to the end of the accented syllable in words with fin and med stress for Catalan (Figure

6) and in words with fin, pen and ante stress for Spanish (Figure 7). Taking the 0 value as the end of the accented syllable, the plots show that all F0 peaks are displaced to the postaccentual syllable/s, since all peaks are located beyond the 0 value. For the three Catalan speakers, peak distance to the end of the syllable is significantly shorter ($p<0.0001$) in word-final position than in word-medial position. The same is true for the Spanish data (left graph). The results also show that in Spanish words with antepenultimate stress F0 peaks are even more displaced than in words with penultimate stress (right graph). These results reveal that there is no strict anchoring of the H at the end of the word.

Separate ANOVAs were done for the two measures of H alignment (relative to either the beginning and the end of the syllable) for the three Catalan and Spanish speakers. Both for Catalan and Spanish, the ANOVAs revealed a significant effect of word boundary both for the measures of peak delay and peak distance to the offset of the syllable. For Catalan, the results of peak delay are: Speaker AG: $F(1, 158) = 50{,}088$; $p<0{,}001$; Speaker PG: $F(1, 158) = 24{,}574$; $p<0{,}001$; Speaker PP: $F(1, 158) = 22{,}434$; $p<0{,}001$). The results of peak distance to the offset of the syllable are: Speaker AG: $F(1, 158) = 84{,}230$; $p<0{,}001$; Speaker PG: $F(1, 158) = 98{,}253$; $p<0{,}001$; Speaker PP: $F(1, 158) = 103{,}399$; $p < 0{,}001$. For Spanish, the results of peak delay are: Speaker TO: $F(2, 159) = 269{,}505$; $p<0{,}001$; Speaker EV: $F(2, 159) = 204{,}951$; $p<0{,}001$; Speaker DO: $F(2, 159) = 348{,}952$; $p<0{,}001$). The results of peak distance to the offset of the syllable are: Speaker TO: $F(2, 159) = 726{,}453$; $p<0{,}001$; Speaker EV: $F(2, 159) = 629{,}163$; $p<0{,}001$; Speaker DO: $F(2, 159) = 780{,}654$; $p < 0{,}001$. Thus all speakers in both languages show that peak location is affected by within-word position in a way that the peak systematically shifts backwards as the end of the word is closer.

Finally, the alignment of the H tonal target relative to the end of the word was also examined. If H was a word edge tone then it would be expected to be consistently anchored at the right edge of the word. The graphs in Figures 8 and 9 plot the distance in ms between the H and the end of the word in words with fin and med stress for Cata-

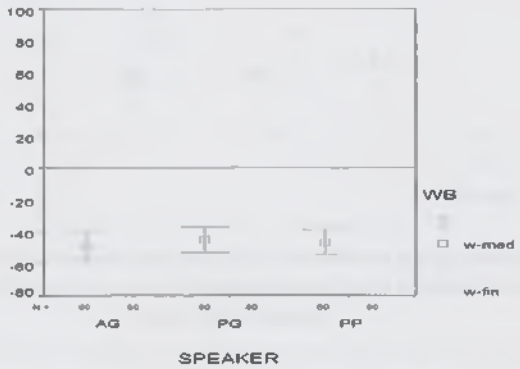


Figure 8. Mean values of distance between H peak location relative to the right edge of the word (in ms) as a function of within-word position for all Catalan speakers in med and fin positions.

lan (Figure 8) and in words with fin, pen and ante stress for Spanish (Figure 9). Taking the 0 value as the end of the word, the results show that in words with a final accent, the H is placed after the end of the word both in Catalan and in Spanish (left graph). Alternatively, in words with a penultimate accent, the H is located before the end of the word. Similarly, in the Spanish data, peaks in words with an antepenultimate accent are also located before the end of the word (right graph). These results confirm the idea that there are no strict word-anchoring effects in any of the two languages.

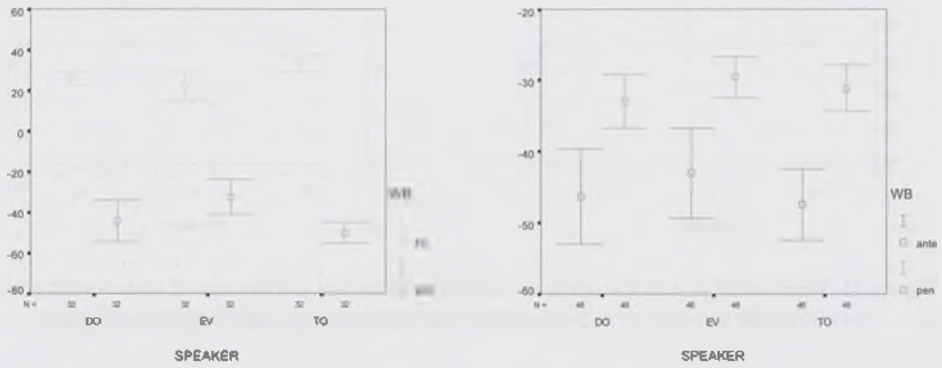


Figure 9. Mean values of distance between H peak location relative to the right edge of the word (in ms) as a function of within-word position for all Catalan speakers: fin vs pen (left graph) and pen vs ante (right graph).

2.2.3. Effects of within-word position on syllable duration

The results displayed in Figures 10 and 11 show the mean values of duration of the accented syllable in words with fin and med stress for Catalan (Figure 10) and in

Figure 10. Mean values of duration of the accented syllable as a function of within-word position for all Catalan speakers in med and fin positions.

words with fin, pen and ante stress for Spanish (Figure 11). It is expected that accented syllables before a word-boundary will be longer than accented syllables in medial position (Lehiste 1960, Turk and White 1999).

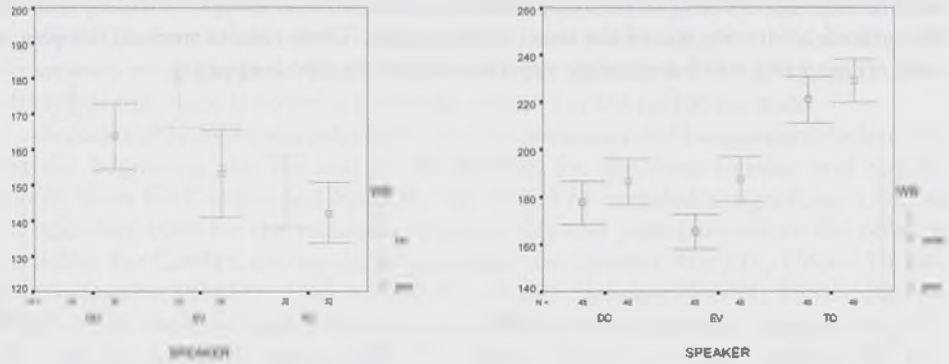


Figure 11. Mean values of duration of the accented syllable (in ms) as a function of within-word position for all Spanish speakers: fin vs pen (graph on the left) and pen vs ante (graph on the right).

The results show no clear duration effects. Word-final accented syllables are slightly longer for all speakers in both languages except for AG (Catalan). However, these differences were only significant ($p=0.004$) for speaker EV (Spanish). This seems to indicate that duration differences as a cue to indicate word boundary location are optional.

The results displayed in Figure 12 and 13 show the mean values of duration of the postaccentual syllable in words with fin and med stress for Catalan (Figure 12) and in

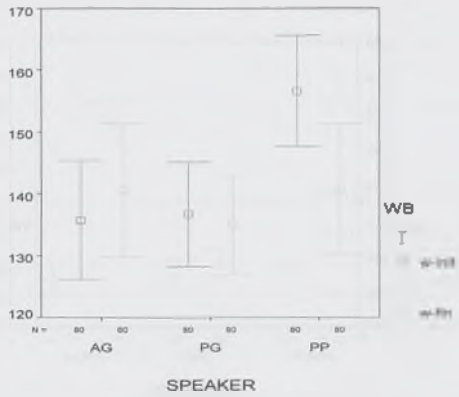


Figure 12. Mean values of duration of the postaccentual syllable (in ms) as a function of within-word position for all Catalan speakers speakers in medial (w-med) and final (w-fin) position.

words with fin, pen and ante stress for Spanish (Figure 13). It is expected that postaccentual syllables before a word-boundary will be longer than non-accented syllables in medial position. However, the results show no clear duration effects. Postaccentual syllables in non-word-final position are slightly longer than postaccentual syllables in word-final position except for speakers AG (Catalan) and TO (Spanish). These differences are not statistically significant for any of the speakers (at p>0,05) except for speaker PP (Catalan). Thus, the hypothesis that the postaccentual syllable will be longer before a word boundary is not confirmed by these data.

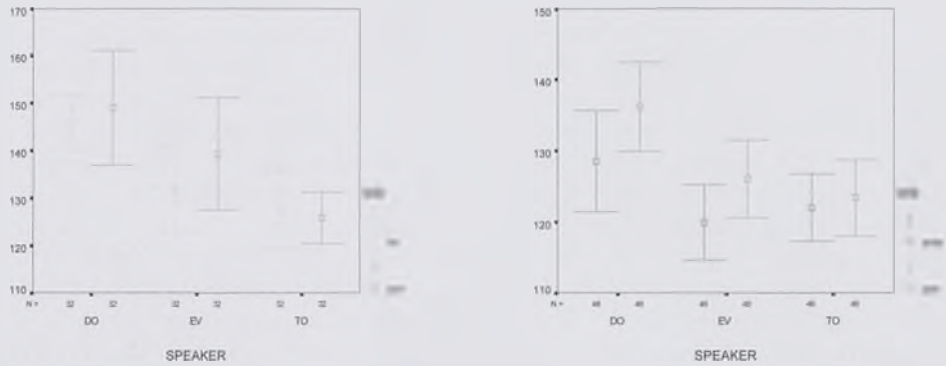


Figure 13. Mean values of duration of the postaccentual syllable (in ms) as a function of within-word position for all Spanish speakers: fin vs pen (graph on the left) and pen vs ante (graph on the right).

Separate ANOVAs were performed for the two dependent variables (namely, the duration of the accented syllable and the duration of the postaccentual syllable) for all speakers. In Catalan, for speaker PG, the ANOVAs revealed that none of the variables showed a significant effect of word boundary, $F(1, 158)=3{,}054$; $p=0{,}082$ (accented syllable), and $F(1, 158)=0{,}088$; $p=0{,}767$ (postaccentual syllable). For speaker AG, the ANOVA for the duration of the accented syllable showed a significant effect of word boundary, $F(1, 158)=7{,}631$; $p<0{,}005$, but not for the duration of the postaccentual syllable, $F(1, 158)=0{,}450$; $p=0{,}503$. Conversely, for speaker PP, the effect of word boundary was only significant for the duration of the postaccentual syllable ($F(1, 158)= 5{,}221$; $p=0{,}024$), bot not for the duration of the accented syllable ($F(1, 158)=2{,}903$; $p=0{,}024$). In Spanish, speakers EV and DO showed no significant effect of word boundary for none of the variables. Speaker EV: $F(2, 159)=1{,}791$; $p=0{,}170$ (accented syllable), and $F(2, 159)=3{,}996$; $p=0{,}02$ (postaccentual syllable); Speaker DO: $F(2, 159)=0{,}007$; $p=0{,}993$ (accented syllable), and $F(2, 159)=4{,}321$; $p=0{,}015$ (postaccentual syllable). For speaker TO, the ANOVA for the duration of the accented syllable showed a significant effect of word boundary, $F(2, 159)=27{,}773$; $p<0{,}001$, but not for the duration of the postaccentual syllable, $F(2, 159)=0{,}939$; $p=0{,}393$. Syllable duration measures thus indicate

that syllable lengthening is not systematic neither in word-final nor in word-initial syllables.

3. PERCEPTION TESTS

The main goal of the perception tests was to check whether H alignment differences could be used to identify word boundaries between otherwise identical sentences.

3.1. Identification test

The materials used in the identification test consisted of 10 ambiguous utterances from the production test. 12 Catalan and 20 Spanish listeners heard the ambiguous utterances a maximum of three times and had to identify the sentence in a two-choice task. The results of the identification test are presented in Figure 14 for Catalan and 15 for Spanish which show the frequency of correct responses for all listeners for each sentence type: Type A includes oxytonic words, Type B paroxytonic words and Type C proparoxytonic words. The results show a low rate of identification for the Catalan sentences (35 to 66% of correct responses) and a variable rate of identification for Spanish sentences (25% to 95%). Words with an antepenultimate or penultimate accent tend to be better identified than words with final stress. These results indicate that the clear H displacement towards the right edge of the word in paroxytones and proparoxytones helps the hearer to a better identification. In oxytones, H placement is more ambiguous and hence contributes to a more difficult identification.

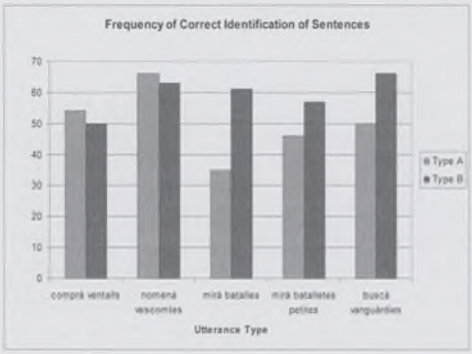


Figure 14. Frequency of correct identification of the 10 ambiguous Catalan sentences.

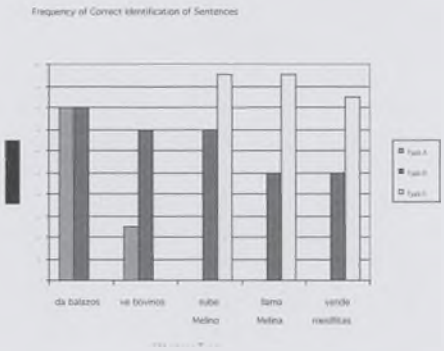


Figure 15. Frequency of correct identification of the 10 ambiguous Spanish sentences.

3.2. Discrimination test

The discrimination test investigates whether listeners could identify sentences in a better way when heard in pairs. The stimuli used were the same 10 sentences as before but presented in pairs. The same listeners as for the identification test heard the pairs of

sentences for a maximum of three times. Subjects had to decide in which order sentences were presented.

The results of the discrimination test are presented in Figure 16 for Catalan and 17 for Spanish. The graph plots the frequency of correct identification responses for each pair. The results show a variable rate of identification for both Catalan (30% to 90% of correct responses) and Spanish (55% to 95%). Overall, the results show some improvement on the identification rates when sentences are heard in pairs.

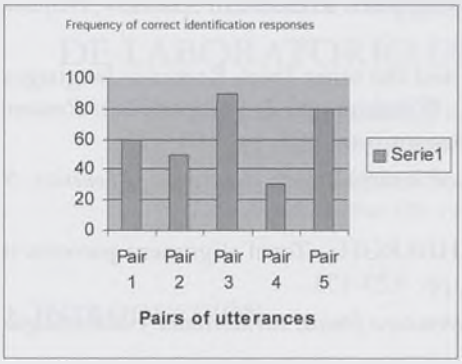


Figure 16. Frequency of correct identification responses for 5 pairs of Catalan ambiguous sentences.

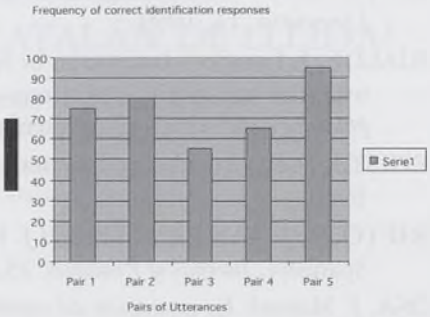


Figure 17. Frequency of correct identification responses for 5 pairs of Spanish ambiguous sentences.

4. CONCLUSION

The results obtained in the production test clearly show that the F0 peak is aligned after the accented syllable in all kinds of words (oxytones, paroxytones and proparoxytones) indicating that there is no strict anchoring of H at the end of the word. Thus, the idea to analyse Catalan and Spanish prenuclear rises with a final H word-edge tone is ruled out by the evidence of the data. Alternatively, the data showed a clear effect of within-word position on the alignment of H, since the F0 peak is more retracted in oytones than in paroxytones and proparoxytones. The results of the perception tests show that listeners do seem to rely on H alignment to distinguish sentences. The close location of H after the word boundary in oxytones causes some difficulties in identifying single sentences. However, some improvement on the identification rates is observed when sentences appear in pairs. This indicates that H alignment differences due to word position are a strong cue for word-boundary identification.

BIBLIOGRAPHY

BECKMAN, M.; M. DÍAZ-CAMPOS; J. T. MCGORY and T. A: MORGAN: «Intonation accross Spanish, in the Tones and Break Indeces Framework», *Probus,* 14, 2002, pp. 9-36.

ESTEBAS-VILAPLANA, Eva: *The use and realisation of accentual focus in Central Catalan with a comparison to English,* University College London, PhD dissertation, 2000.

FACE, Timothy: «A Phonological analysis of rising pitch in Castilian Spanish», *Hispanic Linguistics,* 11, 1999.

HUALDE, J. Ignacio: «Intonation in Spanish and the other Ibero-Romance languages: overview and status quaestinois», in C. Wiltshire and J. Camps (eds.): *Romance phonology and variation,* Amsterdam, J. Benjamins, 2002, pp. 101-115.

LEHISTE, Ilse: «An acoustic-phonetic study of internal open juncture», *Phonetica,* 5, 1960.

PRIETO, P.; J. VAN SANTEN and J. HIRSCHBERGH: «Tonal alignment patterns in Spanish», *Journal of Phonetics,* 23, 1995, pp. 429-451.

SOSA, J. Manuel: *La entonación del español: su estructura fónica, variabilidad y dialectología,* Madrid, Cátedra, 1999.

TURK, A. E. and L. WHITE: «Structural influences on accentual lengthening in English», *Journal of Phonetics,* 27, 1999, pp. 171-206.

UNA CARACTERIZACIÓN DE LA PROSODIA DE HABLA DE LABORATORIO DEL CATALÁN DE LLEIDA[1]

Ana María Fernández Planas, Josefina Carrera Sabaté
Laboratori de Fonetica. Universitat de Barcelona
Domingo Román Montes de Oca
Laboratori de Fonetica, UB: y Pontificia Universidad Católica de Chile

1. INTRODUCCIÓN

Esta comunicación se inscribe en el marco del proyecto general AMPER (Atlas Multimedia de Prosodia del Espacio Románico) y busca caracterizar patrones lingüísticos geoprosódicos a partir de los parámetros de duración, intensidad y entonación. En particular, se sitúa en AMPERCAT, la aplicación de la idea y la metodología AMPER a las principales variedades lingüísticas de Cataluña.

El grupo AMPERCAT ya ha dado fruto en diversas publicaciones con el objetivo de estudiar diferentes variedades geoprosódicas del catalán, lo cual representa un novedoso ámbito de estudio para la fonética catalana. Salvo excepciones como Mascaró (1986), Prieto (1998) y Prieto y Pradilla (2001), que caracterizan los patrones entonativos de interrogativas absolutas de diferentes variedades geolingüísticas del catalán, la mayoría de estudios muestran, fundamentalmente, interés por la entonación del área oriental, sobre todo de su zona central (Barcelona y alrededores).

Este trabajo busca caracterizar las oraciones con expansión tanto en el sujeto como en el objeto del subdialecto de Lleida que forman parte del llamado corpus fijo del proyecto AMPER. Para ello describe sus rasgos prosódicos.

1 Este trabajo ha contado con una ayuda de la SGPI del MCYT, ref. BFF2003-08487.

2. PINCELADAS METODOLÓGICAS

La informante analizada cumple con los requisitos establecidos por todos los grupos AMPER en España para la primera fase de desarrollo del proyecto: es una mujer de media edad, sin estudios superiores y residente en zona urbana. Su habla es considerada representativa del subdialecto en cuestión, *lleidata*.

El corpus está formado por 486 frases de cuatro acentos tonales trisílabos cada una, 54 frases en cada una de las modalidades estudiadas: enunciativa, interrogativa "sin que" e interrogativa "con que"[2], cada una de las cuales fue repetida tres veces. El acento tonal correspondiente al verbo es siempre llano, pero los acentos tonales correspondientes al sujeto, al objeto directo y a sus respectivos complementos son agudos, llanos o esdrújulos en todas las combinaciones posibles. Las frases estudiadas poseen complementos o bien en el sujeto o bien en el objeto, pero no en ambos a la vez.

El análisis ha seguido estrictamente el método establecido en las normas internacionales AMPER a partir del entorno Matlab. Posteriormente, los datos obtenidos a partir de las medias entre las diferentes repeticiones de cada frase se han sometido a tratamiento estadístico mediante el paquete SPSS con el objeto de obtener distintos valores descriptivos.

3. RESULTADOS

Los resultados presentados a continuación se organizan a partir de la distinción de modalidad. En cada una de ellas se tratan los tres parámetros prosódicos de interés: entonación, duración e intensidad. Para cada parámetro se presenta formalmente una figura en la que se incluyen curvas melódicas o gráficos de barras (en función de si el estudio se centra en la entonación, en la duración o en la intensidad) de la frase completa sin expansión (S+V+O); de la frase con expansión de sujeto (S+CS+V+O); y de la frase con expansión de objeto (S+V+O+CO). A modo de ejemplo, en la primera fila se utilizan las frases formadas por acentos tonales llanos, el tipo de palabra más habitual en catalán. En una segunda fila aparecen los sujetos con expansión en detalle (excepto en las interrogativas "con que" en los que aparecen los objetos con expansión) y una última fila en la que aparecen los últimos acentos tonales de las frases, es decir, los objetos con expansión en las enunciativas y en las interrogativas "sin que" y los sujetos en las interrogativas "con que". Los apartados de duración e intensidad presentan solamente dos filas de resultados gráficos. Los números del eje de las x indican las posiciones pretónica, tónica y postónica de los dos acentos tonales estudiados. En los gráficos de entonación, además, se tiene en cuenta el valor inicial de la frase o el final, según la posición del constituyente analizado, si es el caso.

---

2 Llamamos interrogativas "sin que" a las oraciones interrogativas absolutas del tipo *el capità no porta la caputxa petita?*; mientras que llamamos interrogativas "con que" a las oraciones interrogativas absolutas encabezadas por la conjunción "que" que presentan una dislocación del sujeto al final de la frase, por ejemplo: *Que no porta la caputxa petita, el capità?*

3.1. Enunciativas

3.1.1. La entonación

Considérense los resultados obtenidos en este apartado en la figura 1.

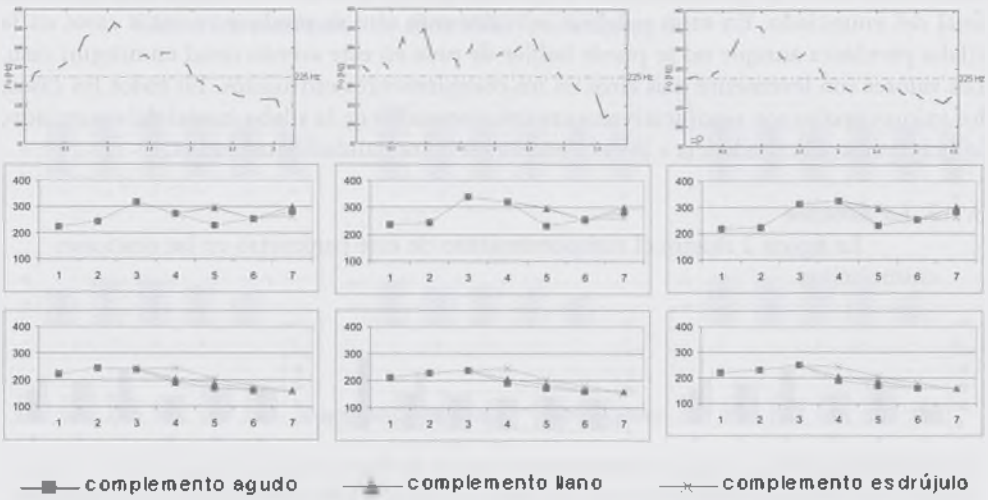


Figura 1. En la fila superior, contorno melódico de: El copista no porta la caputxa / El copista intrepid no porta la caputxa / El copista no porta la caputxa petita (de izquierda a derecha). En la fila central, de izquierda a derecha: El capità protestant/intrèpid/Sòcrates; el copista protestant/intrèpid/Sòcrates; la crítica protestant/tanoca/pràctica. En la fila inferior, de izquierda a derecha: el passaport pintoresc/propici/amb pràctica; la caputxa llampegant/petita/amb pràctica; la càtedra contrincant/completa/típica. Los valores aparecen en Hz.

En términos generales se observa que las oraciones con y sin expansión presentan para cada acento tonal un pico. También se puede apreciar una tendencia a la desacentuación relativa del último acento tonal, que coincide con la declinación final del enunciado.

Una observación más específica muestra que en el sujeto se produce un ascenso significativo entre la sílaba pretónica y la tónica, por lo que el pico se produce en la tónica (en los sujetos agudos y llanos) o en la postónica (en los sujetos esdrújulos). Los complementos agudos del sujeto tienen el pico en la sílaba postónica y los valores son crecientes a partir de la posición pretónica. En el caso de los complementos esdrújulos, el pico se ubica en la sílaba pretónica; en los complementos llanos, el valor de la sílaba postónica es levemente superior al de la pretónica. Curiosamente, los complementos llanos y esdrújulos presentan un descenso de los valores en la tónica.

Los objetos presentan un ascenso de los valores desde la pretónica a la postónica, lo que hace que el pico se tienda a producir en la sílaba postónica, salvo en el caso del

objeto agudo que tiene el pico en la sílaba tónica, aunque la diferencia de valores entre pretónica y tónica no es significativa. Se trata en todos estos casos de un ascenso menos pronunciado que el que sucede, como se ha mencionado, entre la pretónica y la tónica del sujeto. En el caso de los complementos del objeto, todos los valores se ordenan en forma decreciente desde la pretónica hasta la postónica, lo que es coherente con la declinación final del enunciado. En otras palabras, el valor más alto se produce en estos casos en la sílaba pretónica aunque no se puede hablar de pico en este acento tonal en ningún caso. Los valores son levemente más altos en los complementos esdrújulos. En todos los casos, los valores finales son significativamente inferiores a los de la sílaba inicial del enunciado, lo que muestra la tendencia a la declinación en la modalidad enunciativa.

3.1.2. La duración

La figura 2 ilustra el comportamiento de este parámetro en las oraciones enunciativas.

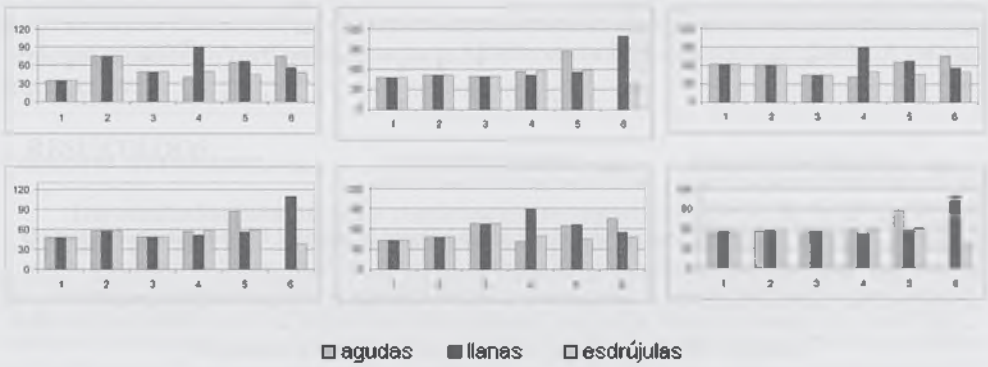


Figura 2. En la fila superior, duración en ms de: El capità protestant/intrèpid/Sòcrates; el copista protestant/intrèpid/Sòcrates; la crítica protestant/tanoca/pràctica. En la fila inferior, de izquierda a derecha: el passaport pintoresc/propici/amb pràctica; la caputxa llampegant/petita/amb pràctica; la càtedra contrincant/completa/típica.

Los sujetos agudos muestran una mayor duración en la sílaba tónica; en los sujetos llanos, la duración mayor se produce en la sílaba postónica; los esdrújulos muestran duraciones similares para las posiciones pretónica y tónica, mayores que la de la postónica. Los complementos del sujeto agudos muestran una duración creciente de las vocales pretónica, tónica y postónica. En los llanos esa relación es inversa. En los esdrújulos, las duraciones son similares en todas las posiciones.

En los objetos agudos, llanos y esdrújulos la mayor duración se concentra en la sílaba tónica, pero las diferencias son leves, especialmente entre los llanos y los esdrújulos. En los complementos del objeto agudos y llanos se observa un alargamiento de la vocal final, lo que es coherente con lo que se sabe acerca de las duraciones de las vocales finales de los enunciados. En el caso de los complementos esdrújulos, se observa un acortamien-

to de la vocal postónica, pero se debe tener en cuenta que esta vocal no es la última del enunciado (por razones metodológicas, la duración de la vocal final no se tomó en cuenta para este trabajo).

3.1.3. La intensidad

Véanse los resultados de intensidad en la figura 3.

Figura 3. En la fila superior, intensidad en dB de: El capità protestant/intrèpid/Sòcrates; el copista protestant/intrèpid/Sòcrates; la crítica protestant/tanoca/pràctica. En la fila inferior, de izquierda a derecha: el passaport pintoresc/propici/amb pràctica; la caputxa llampegant/petita/amb pràctica; la càtedra contrincant/completa/típica.

Se observa que en los sujetos llanos y esdrújulos los valores de intensidad decrecen desde la sílaba pretónica a la postónica; en cambio, en los sujetos agudos el orden es creciente. En los complementos del sujeto llanos y esdrújulos hay un orden decreciente desde la sílaba pretónica a la tónica; el complemento agudo presenta una forma levemente diferente, aunque las diferencias son mínimas.

En los objetos hay una cierta heterogeneidad de los valores: mientras que en los agudos los valores más altos se concentran en las sílabas tónica y postónica, en los llanos -en estas mismas posiciones- se dan los valores más bajos; por su parte, los objetos esdrújulos concentran los valores más altos en las posiciones pretónica y tónica. En los complementos de los objetos, los valores son en todos los casos, decrecientes, significativamente.

Resulta especialmente llamativo que, en algunos casos, coincidan los tres parámetros (frecuencia fundamental, duración e intensidad) como marcas de relieve silábico. Esto sucede en el complemento agudo del sujeto, en que se marca la sílaba postónica; en el complemento esdrújulo del sujeto, en el que se marca la pretónica y en el objeto agudo, en el que la sílaba marcada es la tónica.

3.2. Interrogativas "sin que"

3.2.1. La entonación

Véanse en la figura 4 los resultados obtenidos en la entonación de las interrogativas absolutas "sin que".

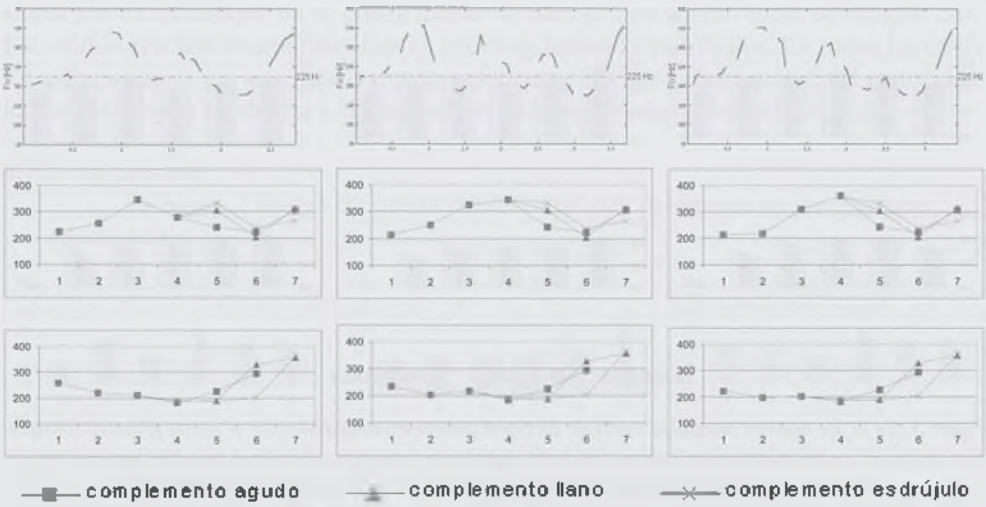


Figura 4. En la fila superior, contorno melódico de: El copista no porta la caputxa? / El copista intrèpid no porta la caputxa? / El copista no porta la caputxa petita? (de izquierda a derecha). En la fila central, de izquierda a derecha: El capità protestant/intrèpid/Sòcrates; el copista protestant/intrèpid/Sòcrates; la crítica protestant/tanoca/pràctica. En la fila inferior, de izquierda a derecha: el passaport pintoresc/propici/amb practica: la caputxa llampegant/petita/amb practica: la catedra contrincant/completa/típica.

Se observa en los contornos globales de las frases que las oraciones con y sin expansión no difieren demasiado en el sentido de que se mantiene la máxima apuntada en (Martínez Celdrán y otros, 2005; van Oosterzee y otros, en prensa; Carrera y otros, 2004) para las frases con tres acentos tonales: cada uno presenta un pico destacado, a diferencia de lo que sucede en otros subdialectos del catalán como el tortosí (véase Martínez Celdrán y otros, en prensa), en los que cada constituyente sintáctico presentra un pico, esté constituido por uno o por dos acentos tonales. En los ejemplos propuestos ese pico se realiza sistemáticamente en las sílabas postónicas de forma que se alinea con el final de la palabra en cuestión.

Acudiendo al detalle vemos que en los sujetos con expansión los núcleos presentran el pico en la postónica si son llanos o esdrújulos pero en la tónica si son agudos y, respecto al acento tonal que actúa como su complemento, si está formado por palabra llana o esdrújula se comporta de forma muy similar con un valle en la tónica y un ascenso

en la postónica. Si está formado por palabra aguda, a grandes rasgos la melodía es similar pero el descenso desde el pico del acento tonal anterior hasta la tónica del complemento no es tan gradual como sucede en los anteriores tipos de palabra sino que es brusco en la pretónica del complemento.

Los objetos directos sin distinción por el tipo de palabra que lo componen se insertan en una caída progresiva de la melodía y el último acento tonal, el del complemento, en un ascenso muy pronunciado, como era de esperar. En el caso de que ese último acento tonal esté constituido por palabras agudas el ascenso se inicia a partir de la pretónica; si está constituido por palabras llanas o esdrújulas se inicia a partir de la tónica. En todos los casos el ascenso llega hasta la última vocal de la frase y en el caso de tratarse de una palabra aguda alcanza valores frecuenciales no tan altos como en el resto de casos. Las palabras llanas y esdrújulas alcanzan al final valores muy similares, sin embargo, el ascenso más pronunciado en el caso de las palabras llanas se produce entre tónica y postónica, mientras que en las esdrújulas parece que se produce de forma brusca entre la postónica y el final de la frase, hay que recordar que falta el valor central de una vocal, la pospostónica, que no se ha tenido en cuenta en estos análisis y que, probablemente, suavizaría el ascenso.

3.2.2. La duración

Obsérvense los resultados obtenidos en este apartado en la figura 5.

Figura 5. En la fila superior, duración en ms de: El capità protestant/intrèpid/Sòcrates; el copista protestant/intrèpid/Sòcrates: la crítica protestant/tanoca/practica. En la fila inferior, de izquierda a derecha: el passaport pintoresc/propici/amb pràctica: la caputxa llampegant/petita/amb pràctica: la càtedra contrincant/completa/típica.

Se constata que en los sujetos con complemento, el segundo acento tonal suele tener mayor duración que el primero en todos los tipos de palabra. Dentro de él, las agudas presentan más duración en las tónicas, las llanas en las pretónicas y las esdrújulas no ofrecen diferencias relevantes en función de la posición dentro del acento tonal. En los objetos directos con expansión destaca la mayor duración de las últimas vocales de

las palabras acabadas en aguda y en llana. Las frases acabadas en palabra esdrújula no presentan esa mayor duración pero hay que tener en cuenta que los gráficos reflejan las posiciones pretónica, tónica y postónica, es decir, que la última vocal de la frase no ha quedado registrada y cabe pensar que también presentaría un aumento notorio como efectivamente sucede en los demás tipos de palabra que, aproximadamente, doblan la duración de la vocal anterior.

3.2.3. La intensidad

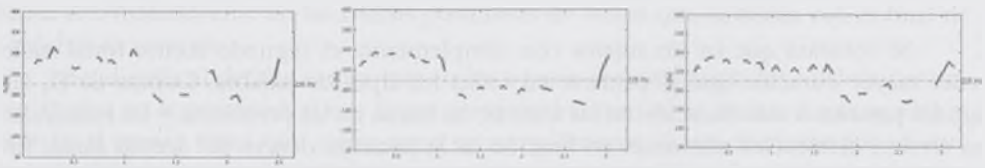
En la figura 6 se muestran los resultados de intensidad obtenidos en este apartado.

Figura 6. En la fila superior, intensidad en dB de: El capità protestant/intrèpid/Sòcrates; el copista protestant/intrèpid/Sòcrates; la crítica protestant/tanoca/pràctica. En la fila inferior, de izquierda a derecha: el passaport pintoresc/propici/amb pràctica; la caputxa llampegant/petita/amb pràctica; la càtedra contrincant/completa/típica.

Lo más destacable del análisis de la intensidad reside, de acuerdo a lo esperado, en la disminución de energía en todos los tipos de palabra al final de la frase, menos acusado en las agudas y más en las esdrújulas. El descenso se nota especialmente cuando el objeto directo está constituido por una palabra esdrújula. Al pincipio de la frase no se observan diferencias importantes entre los sujetos y sus complementos.

3.3. Interrogativas "con que"

La figura 7 ilustra el apartado entonativo en esta modalidad de habla.



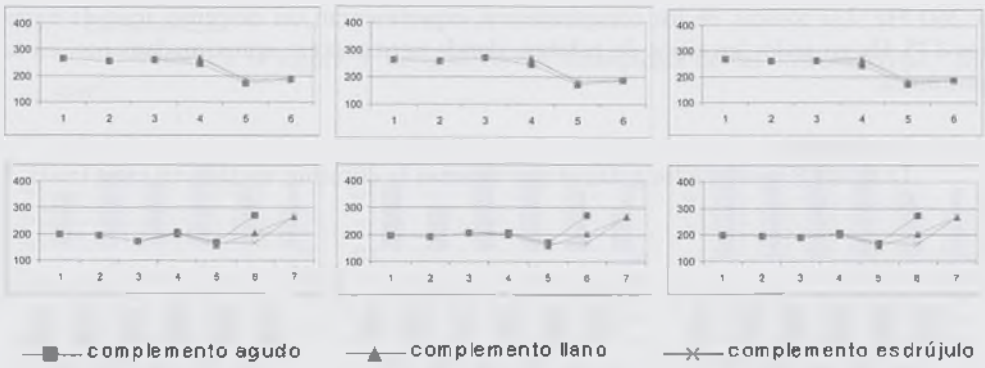


Figura 7. En la fila superior, contorno melódico de: Que no porta la caputxa, el copista?/ Que no porta la caputxa, el copista intrèpid? / Que no porta la caputxa petita, el copista? (de izquierda a derecha). En la fila central, de izquierda a derecha: el passaport pintoresc/propici/amb pràctica; la caputxa llampegant/petita/amb pràctica; la càtedra contrincant/completa/típica. En la fila inferior, de izquierda a derecha:. El capità protestant/intrèpid/Socrates; el copista protestant/intrèpid/Socrates; la crítica protestant/tanoca/pràctica.

En los contornos globales de las oraciones observamos un patrón melódico similar en todas ellas, tanto si no presentan expansión como si la presentan en el sujeto o en el objeto. Efectivamente, se constata una primera parte cerca de los 300 Hz, que corresponde al verbo de las oraciones. La sílaba tónica del objeto o de su complemento experimenta un descenso de unos 100 Hz, donde se mantiene aproximadamente hasta la última sílaba de la frase, cuyo valor final es, sin embargo, ascendente y alcanza los 250-300 Hz.

En las frases cuyo sujeto presenta expansión, las vocales del núcleo no presentan picos destacados en ninguna posición, pues todos los valores oscilan sobre los 200 Hz. Detalladamente, observamos, sin embargo, que en las frases con sujeto agudo, el valor de la postónica en el sujeto es ligeramente inferior al de las otras posiciones, tendencia que también se detecta, aunque en menor grado, en los esdrújulos. En los sujetos llanos, el valor de la postónica es ligeramente más elevado que el de los otros valores.

Por lo que respecta a los complementos de los sujetos, la vocal tónica presenta un ligero descenso frecuencial respecto de la sílaba pretónica en todos los tipos de palabra. El tono asciende cerca de unos 100 Hz en el valor final de la frase cuando acaba con palabra aguda; cuando la última palabra es llana, el valor frecuencial aumenta unos 50 Hz en la sílaba postónica y en unos 50 Hz más en la sílaba final; en las palabras esdrújulas, el valor frecuencial de la sílaba tónica se mantiene en la postónica y asciende, como en las agudas, unos 100 Hz en el valor final de la frase (recordamos que no consignamos el valor central de la sílaba pospostónica que seguramente suavizaría el ascenso tonal.

En las oraciones con expansión en el objeto directo, se observa un patrón regular en todas las posiciones del objeto como en sus complementos, con frecuencias cercanas

a 260 Hz. Las tónicas de los complementos experimentan un descenso acusada hasta los 175 Hz en todos los tipos de palabra, donde permanecen, aproximadamente, en la postónica.

3.3.2. La duración

La figura 8 contiene los gráficos que ilustran la duración vocálica en estas frases.

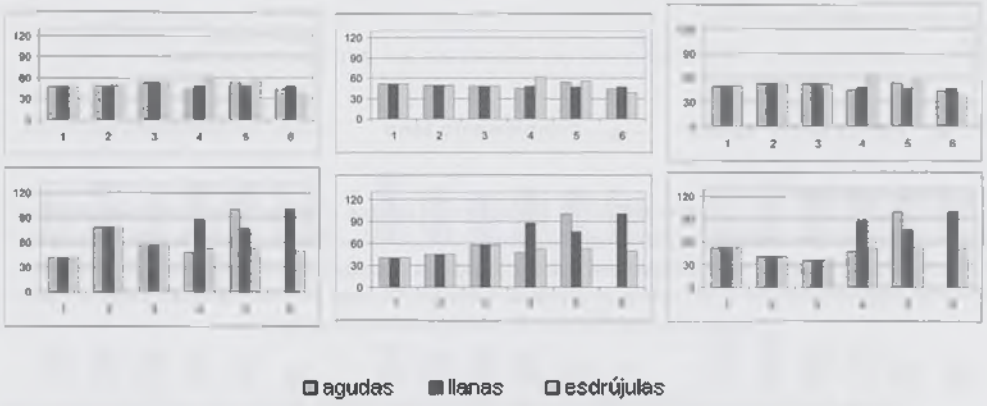


Figura 8. En la fila superior, barras de duración de: el passaport pintoresc/propici/amb pràctica: la caputxa llampegant/petita/amb pràctica; la càtedra contrincant/completa/típica. En la fila inferior, de izquierda a derecha: El capità protestant/intrèpid/Sòcrates; el copista protestant/intrèpid/Sòcrates: la crítica protestant/tanoca/pràctica.

En las oraciones con expansión en el sujeto observamos diferentes duraciones según si el núcleo es agudo, llano o esdrújulo. Así pues, en las oraciones con sujeto agudo, la sílaba tónica es la que presenta una mayor duración; cuando el sujeto es llano, la duración aumenta progresivamente desde la vocal pretónica hasta la postónica; y cuando el sujeto es esdrújulo, esta progresión es descendente desde la pretónica hasta la postónica. La duración de las vocales en los complementos del sujeto varía según el tipo de palabra: es decir, en las palabras agudas, las vocales tónicas son las que presentan más duración; en las llanas, la mayor duración se encuentra en la vocal postónica, seguida de la pretónica y la tónica; en las esdrújulas, no se observan diferencias por duración en las vocales, pero recordamos que no queda reflejada la última vocal.

En las oraciones con expansión en el objeto, la duración de las vocales del objeto es muy similar, aunque se observan pequeñas diferencias en las de los complementos: en los agudos, la duración mayor se encuentra en la tónica; en los llanos, las tres vocales tienen una duración parecida; y en los esdrújulos, la duración más acusada se encuentra en la vocal pretónica.

3.3.3. La intensidad

Véase en la figura 9 los gráficos de intensidad de las interrogativas "con que".

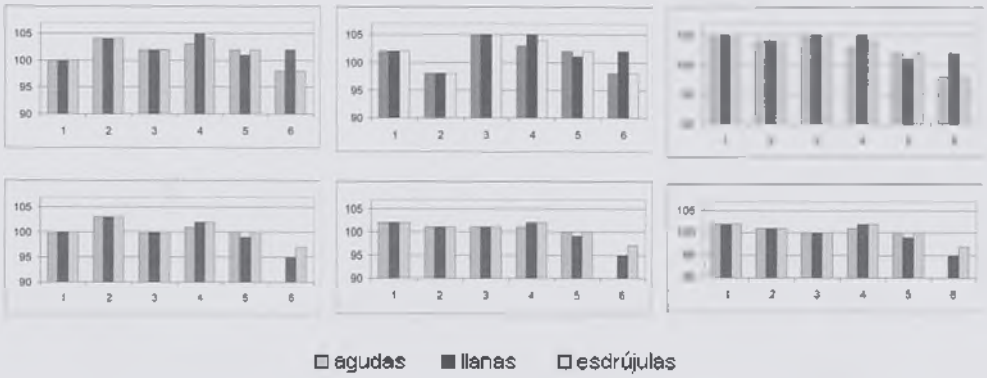


Figura 9. En la fila superior, barras de intensidad de: el passaport pintoresc/propici/amb practica: la caputxa llampegant/petita/amb practica: la catedra convincant/completa/típica. En la fila inferior, de izquierda a derecha: El capità protestant/intrèpid/sòcrates: el copista protestant/intrèpid/socrates; la crítica protestant/tanoca/pràctica.

En las oraciones con expansión en el sujeto, observamos una intensidad que presenta valores similares en todas las vocales excepto en el valor final de la frase de los complementos acabados en palabras llanas y esdrújulas, donde se registran valores inferiores. En los sujetos agudos, la mayor intensidad se registra en las vocales tónicas y en los llanos y esdrújulos, en las pretónicas.

En las oraciones con expansión en el objeto directo, la intensidad de éste varía según el tipo de palabra. Así pues, en los objetos agudos, la vocal pretónica es la que registra menor intensidad y la tónica la que presenta la mayor; en los complementos llanos, es la tónica la vocal claramente menos intensa y la postónica la más audible; en los esdrújulos, las vocales de las tres posiciones registran una intensidad similar, algo inferior en la tónica. Por lo que respecta a la extensión del complemento directo, la intensidad de las vocales sigue un patrón decreciente desde la pretónica hasta la postónica en palabras agudas y esdrújulas; en las llanas, las vocales menos intensas son las tónicas, puesto que las postónicas las superan levemente.

3.4. El caso particular del choque acentual

Entre las frases estudiadas se encuentran algunas con especial interés entonativo puesto que incluye el choque acentual o *clash* silábico de dos tónicas. Véanse a modo de ejemplo algunas curvas de F0 de estas frases en la figura 10.

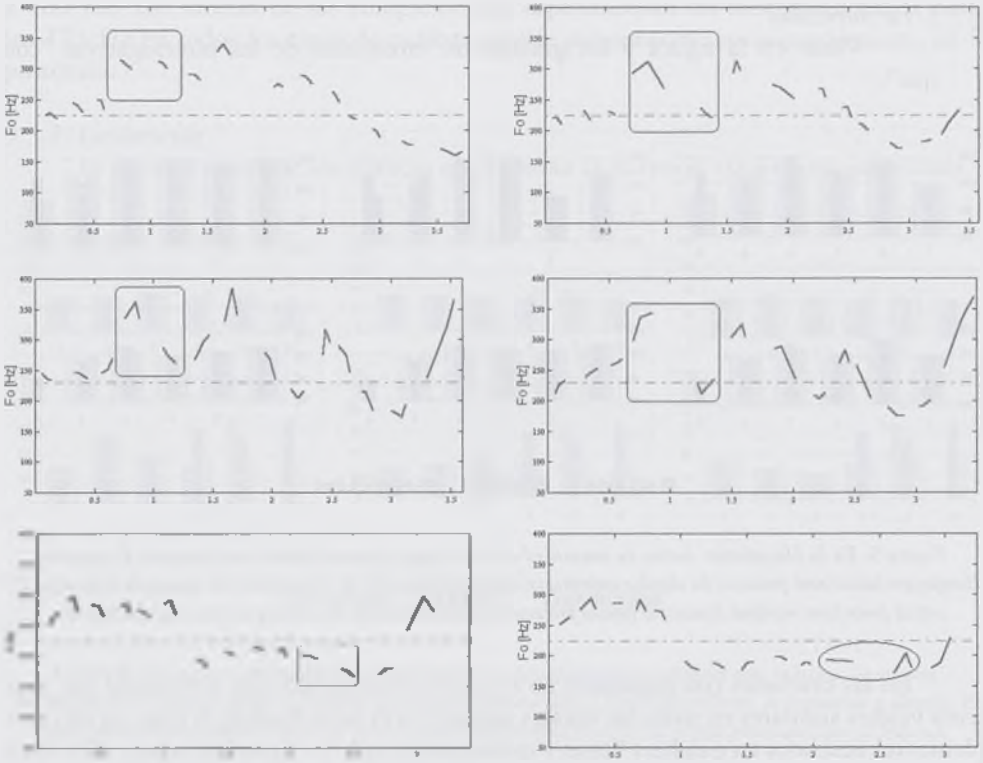


Figura 10. De arriba abajo y de izquierda a derecha: El capità Sòcrates no porta la caputxa, El capità intrèpid no ocupa la catedra, El capità Sòcrates no porta la caputxa?, El capità intrèpid no ocupa la catedra?, Que no porta la caputxa, el capità Sòcrates? y Que no ocupa la càtedra, el capità intrèpid?.

Diversos estudios indican que el choque que se suele solucionar mediante la sensación de desacentuación de una de las dos tónicas, la primera, a través de la eliminación del ascenso tonal en dicha sílaba (Prieto y otros, 2001). Nuestros resultados indican que las modalidades enunciativa e interrogativa "sin que" en *lleidatà* siguen manteniendo la tendencia general hallada en todas las frases: cada acento tonal un pico, salvo el último de la frase, en las enunciativas porque ya forma parte del tonema descendente y, en las interrogativas, porque prepara la subida acusada de la última vocal. En estos casos, el choque se soluciona mediante la situación del primer pico en la primera tónica y la posposición del pico correspondiente al segundo acento tonal implicado de tal modo que se alinea con el final de dicho acento tonal. En los ejemplos de la izquierda, incluso el segundo pico alcanza valores más altos que el primero, contrariamente a la tendencia general observada de escalonamiento descendente progresivo, como se ve en los casos de la derecha. Esta solución del mantenimiento de los dos ges-

tos tonales también es apuntada de alguna forma por Prieto y otros (2001:31) lo cual, indican, es percibido como una desacentuación parcial o mínima del primer acento, aunque hablan de un alargamiento de duración considerable en la primera sílaba que nosotros no encontramos en nuestros datos.

En las frases interrogativas "con que" no se observa ningún pico realmente destacado. Probablemente el hecho de necesitar preservar dos tesituras más o menos estables y una última vocal ascendente limita más que en las otras modalidades este recurso.

4. CONCLUSIÓN

El parámetro más destacado es la entonación frente a la duración y la intensidad en la distinción de la modalidad. En ninguna de ellas se observan diferencias significativas por el hecho de presentar expansión de sujeto o de objeto respecto a las estructuras básicas S+V+O. Por otra parte, el tipo de palabra tampoco influye decisivamente en las estructuras estudiadas en ninguno de los parámetros prosódicos.

Finalmente, el choque acentual se soluciona en las oraciones enunciativas e interrogativas "sin que" mediante el mantenimiento del primer pico en la primera tónica y la posposición del segundo hasta el final del acento tonal; las interrogativas "con que" no presentan picos realmente destacados en ninguna posición.

BIBLIOGRAFÍA

CARRERA SABATÉ, J; VAN OOSTERZEE, C.; FERNÁNDEZ PLANAS, A.M.; ROMERA BARRIOS, L.; ESPUNY, J, y MARTÍNEZ CELDRÁN, E. (2004): «Les interrogatives al tortosí i al lleidatà. un element diferenciador de subdialectes», *Estudios de Fonética Experimental*, XIII, pp. 157-179.

MARTÍNEZ CELDRÁN, E.; FERNÁNDEZ PLANAS, A.M.; CARRERA, J. y ESPUNY, J. (2005): «Approche de la carte prosodique dialectale de la langue catalane en Catalogne», *Projet AMPER*, Géolinguistique-Hors Série n.3, Grenoble: Centre de Dialectologie, Université Stendhal Grenoble 3, pp. 103-151.

MARTÍNEZ CELDRÁN, E.; VAN OOSTERZEE, C. y ESPUNY, J. (en prensa): «La prosodia del habla de laboratorio del "tortosí"», *Actas del III Congreso de Fonética Experimental*, Santiago de Compostela, 24-26 de octubre de 2005.

MASCARÓ PONS, I. (1986): «Introducció a l'entonació dialectal catalana», *Randa*, 22, pp. 5-38.

PRIETO, P. (1998): «L'entonació dialectal del català: el cas de les frases interrogatives absolutes», en A. Bover, M. R. Lloret i M. Vidal-Tibbits (eds.): *Actes del Novè Col·loqui d'Estudis Catalans a Nord-Amèrica*, Barcelona: PAM, pp. 347-377.

PRIETO, P.; OLIVA, S.; PALMADA, B.; SERRA, P.; BLECUA, B.; LLACH, S. Y OLIVA, V. (2001): «Anàlisi acústica de la resolució de xocs accentuals en català», *Estudios de Fonética Experimental*, XI, pp.11-40.

PRIETO, P. y PRADILLA, M.À. (2003): «Les inflexions finals en l'entonació del "tortosí"», en M.A. Pradilla (ed.): *Llengua i literatura a les comarques de la diocesi de Tortosa*, Benicarló: Onada Edicions, pp. 276-286.

VAN OOSTERZEE, C.; FERNÁNDEZ PLANAS, A.M.; ROMERA BARRIOS, L.; CARRERA SABATÉ, J.; ESPUNY, J. y MARTÍNEZ CELDRÁN, E. (en prensa): «Proyecto AMPER: estudio contrastivo de frases interrogativas sin expansión en *tortosí* y en *lleidatà*», *Actas del VI Congreso de Lingüística General*, Santiago de Compostela, 2004.

CARACTERIZACIÓN PROSÓDICA DAS INTERROGATIVAS DA COSTA DA MORTE E DO MORRAZO

Elisa Fernández Rei, Ana Escourido Pernas, Maruxa Caamaño Varela
Instituto da Lingua Galega

1. INTRODUCIÓN*

En traballos anteriores (Fernández Rei, Elisa et alii. 2005a, 2005b) achegabamos os nosos primeiros datos e conclusións sobre a prosodia do galego, a partir das análises realizadas a unha informante da zona de Santiago de Compostela, entoación que se pode considerar maioritaria no dominio lingüístico galego. Estes primeiros resultados foron obtidos a partir dun corpus fixo experimental co que pretendiamos observar os problemas prácticos co fin de confeccionar un corpus fixo definitivo e realizar análises fiables para contribuír ao AMPER. A partir desta experiencia, elaboramos un corpus definitivo que utilizaremos en futuras investigacións. Levamos a cabo a gravacións dunha informante de Camelle (comarca da Terra da Soneira) e outra de Bueu (comarca do Morrazo). Estas dúas zonas destacan pola peculiar entoación das interrogativas, diferente con respecto á máis común da lingua galega. As dúas informantes unha idade semellante (44 e 45 anos), carecen de estudos superiores, son naturais de Camelle e de Bueu (onde residen habitualmente) e caracterízanse por posuíren a fala da zona xeográfica á que pertencen.

2. OBXECTIVOS

Pretendemos observar as principais características da entoación das interrogativas absolutas en dous dialectos destas zonas costeiras de Galicia, sen perder de vista os datos achegados os citados estudos sobre a fala de Santiago de Compostela. Trataremos de

* Este traballo realizouse dentro do proxecto *Descripción y estudio de los modelos entonativos en el gallego actual*, dirixido por Manuel González González e financiado polo Ministerio de Educación y Ciencia (HUM2005-08282-C02-01).

describir o comportamento dos principais parámetros de estudo –frecuencia fundamental, duración e enerxía- e observar en que medida inflúen nas diferenzas prosódicas.

3. RESULTADOS DA ANÁLISE

3.1. Corpus analizado

Seleccionamos para este estudo só aquelas oracións do corpus fixo que presentan a estrutura *Suxeito + Verbo + Complemento*, sen ningún tipo de extensión ou modificador nin no suxeito nin no complemento e unicamente na súa modalidade interrogativa.

No cadro 1 amosamos a totalidade das oracións tratadas neste traballo. Contabamos con tres realizacións de cada informante a partir das que se calcularon as medias que recollemos e comentamos nos gráficos presentados.

O ferrolán falaba co capitán?
O ferrolán falaba co cabalo?
O ferrolán falaba co cómico?
O soldado falaba co capitán?
O soldado falaba co cabalo?
O soldado falaba co cómico?
O médico falaba co capitán?
O médico falaba co cabalo?
O médico falaba co cómico?

Cadro 1: Interrogativas con suxeito e complemento sen extensión.

3.2. A frecuencia fundamental

Nas variacións da frecuencia fundamental podemos observar os principais trazos que caracterizan e individualizan as interrogativas dos dous dialectos tratados neste estudo. Nos seguintes gráficos mostramos as medias de cada suxeito e de cada complemento segundo a súa acentuación -oxítona, paroxítona e proparoxítona-, realizadas pola informante de Bueu (figura 1) e pola informante de Camelle (figura 2).

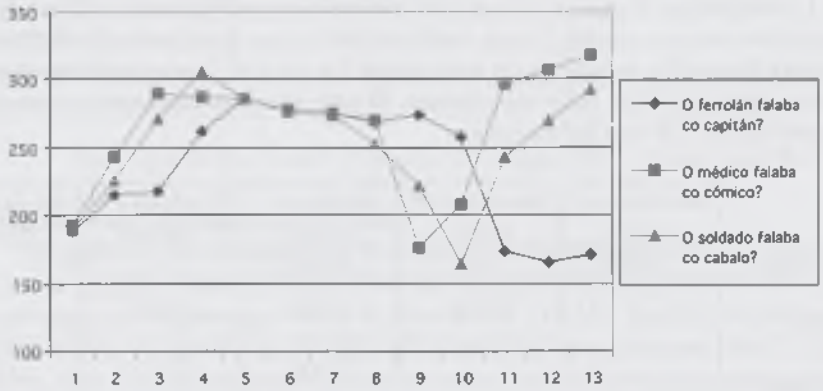


Figura 1: Comparación dos tres suxeitos na modalidade interrogativa (Bueu) [1]

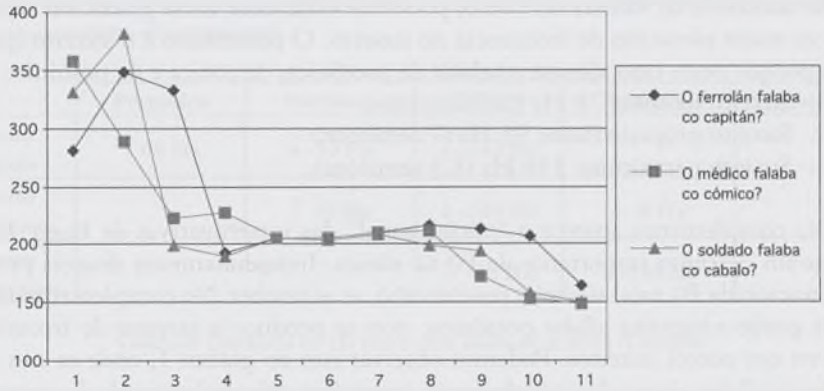


Figura 2: Comparación dos tres suxeitos na modalidade interrogativa (Camelle)

3.2.1. Descrición interrogativas de Bueu

As interrogativas de Bueu (gráfico 1) comezan no suxeito nun ton inferior aos 200 Hz, baixo con respecto á frecuencia media da informante (230 Hz) e á tonalidade en que se desenvolven as súas interrogativas. Na segunda sílaba, sexa tónica ou átona, comeza a ascensión tonal que continúa nas sílabas seguintes ata alcanzar unha frecuencia elevada, próxima aos 300 Hz.

1 As sílabas 11, 12 e 13 correspóndense cos tres valores de F0 da última sílaba.

O modo en que se produce esta subida depende da configuración silábica do suxeito. Como observamos no cadro 2, onde medimos a diferenza absoluta en Hz de cada sílaba con respecto á anterior, prodúcese un salto maior nas tónicas, fronte ás sílabas pretónicas e postónicas que ascenden máis suavemente. O máximo tonal do suxeito alcánzase nas sílabas postónicas, e se non hai na tónica.

	Subida ata a 1ª pretón.	Subida ata a 2ª pretón.	Subida ata a tónica	Subida ata a 1ª postón.	Subida ata a 2ª postón.
Suxeito oxítono	31 Hz	3 Hz	44 Hz		
Suxeito paroxítono		36 Hz	45 Hz	35 Hz	
Suxeito proparoxítono			51 Hz	47 Hz	-3 Hz

Cadro 2: Distancia en Hz entre cada sílaba do suxeito (Bueu)

Se sumamos os valores do cadro, podemos establecer unha gradación segundo a menor ou maior elevación de frecuencia no suxeito. O paroxítono é o suxeito que máis ascende porque neste caso súmase a subida da pretónica, da tónica e da postónica:

1. Suxeito oxítono: 78 Hz (5,7 semitóns)
2. Suxeito proparoxítono: 95 Hz (7 semitóns)
3. Suxeito paroxítono: 116 Hz (8,3 semitóns)

No complemento aparece o "núcleo tonal" das interrogativas de Bueu. En xeral rexístrase un descenso importante de F0 na tónica. Inmediatamente despois prodúcese unha elevación da F0 na(s) sílaba(s) postónica(s), se as houber. No complemento oxítono, que non posúe ningunha sílaba postónica, non se produce o ascenso de frecuencia ao final coma nos outros suxeitos. Podemos observar isto no gráfico 1, onde os tres puntos medidos na sílaba tónica do complemento manteñen valores baixos de frecuencia. No entanto, nos ficheiros de son real percibimos como se produce un incremento tonal ao final da realización dos exemplos oxítonos. A nosa percepción do ton pode estar asociada a unha elevación de frecuencia non na vogal analizada, senón na nasal final trabante do complemento oxítono (*capitán*).

No cadro 3 amosamos en Hz os datos do movemento de caída e subida final da interrogativa de Bueu, segundo a configuración acentual do complemento. Nos datos da columna de descenso, que están tomados entre a pretónica (2ª pretónica) e a tónica, destaca a caída menos marcada do complemento paroxítono. Nos datos da columna de ascenso, que están tomados entre a tónica e a última sílaba da oración, observamos como o ascenso final é bastante maior que o descenso da tónica.

	Descenso	Ascenso
Complemento oxítono	-92 Hz (-7,6 st)	-
Complemento paroxítono	-57 Hz (-5,2 st)	100 Hz (8,2 st)
Complemento proparoxítono	-93 Hz (-7,3 st)	131 Hz (9,6 st)

Cadro 3: Datos do descenso e ascenso no complemento das interrogativas de Bueu.

3.2.2. Descrición das interrogativas de Camelle

No gráfico 2, onde mostramos as medias das interrogativas de Camelle, o inicio da oración se produce, nos tres casos, nun ton bastante elevado (próximo aos 300 Hz), cando a frecuencia fundamental media da falante é de 210 Hz. A sílaba ou sílabas pretónicas comezan nun ton moi elevado para seguidamente caer de maneira brusca na tónica. O máximo tonal da oración prodúcese sempre sobre unha das primeiras sílabas átonas do suxeito: a sílaba anterior á tónica en suxeitos proparoxítono (1ª sílaba) e paroxítono (2ª sílaba), e a pretónica inicial no oxítono (2ª sílaba). Isto provoca que o contorno melódico de oxítonas e paroxítonas comece cun ascenso cara o primeiro pico tonal. O suxeito proparoxítono segue o mesmo esquema pero comeza xa no punto máis alto, polo que só vemos un contorno descendente.

	Pretónica	Pretónica	Tónica	Postónica	Postónica
Suxeito oxítono	↑ 68 Hz	↓ -16 Hz	↓ -142 Hz		
Suxeito paroxítono		↑ 50 Hz	↓ -184 Hz	↓ -8 Hz	
Suxeito proparoxítono			↓ -70 Hz	↓ -136 Hz	↑ 5 Hz

Cadro 4: Distancia en Hz entre cada sílaba do suxeito (Camelle).

O comportamento das postónicas do suxeito varía segundo a acentuación. No suxeito paroxítono, a postónica limítase a manter a altura alcanzada tras o descenso brusco da tónica. No suxeito proparoxítono, a postónica inicial prosegue a caída da tónica e a postónica final mantén a altura alcanzada.

O resto da oración mantense sen variacións tonais relevantes ata o complemento. A acentuación parece influír tamén no final da oración interrogativa, aínda que non tan marcadamente coma no suxeito. Hai unha tendencia clara a que as maiores caídas de F0 se produzan na sílaba tónica. As sílabas postónicas, se as houber, tenden a manter o ton baixo final e presentan valores lixeiramente máis graves que a tónica precedente, aínda que isto pode explicarse pola tendencia xeral á declinación. No cadro 5 podemos observar todo isto cos datos (en Hz) do complemento camellán segundo a súa configuración acentual.

	Pretónica	Pretónica	Tónica	Postónica	Postónica
Complemento oxítono	-3 Hz	-6 Hz	-43 Hz		
Complemento paroxítono		-4 Hz	-36 Hz	-7 Hz	
Complemento proparoxítono			-39 Hz	-19 Hz	-5 Hz

Cadro 5: Distancia en Hz entre cada sílaba do complemento (Camelle)

3.3. Duración

Cando nos primeiros traballos analizabamos a duración dos enunciados, con e sen extensión, observabamos nos dous casos que no tonema as sílabas tónicas eran as que presentaban os valores máis elevados; ao igual que as sílabas tónicas das extensións, agás cando a extensión do suxeito era oxítona (de Marín). Compre aclarar a este respecto que a explicación ofrecida a este comportamento era que a vogal baixa era a de carácter tónico, en tanto que a súa pretónica era a vogal de maior duración, por ser a de abertura máxima. No entanto, canto á duración dos núcleos do suxeito do corpus de proba, comprobamos que as sílabas que rexistraron os valores máis elevados foron, polo xeral, as terceiras, independentemente da posición do acento.

Na análise da duración deste novo corpus reparamos en que en cada un dos grupos acentuais a sílabas que máis duran son as tónicas. Esta característica dáse nas dúas falas, se ben na Costa da Morte á máis salientable. (gráficos 3 e 4)

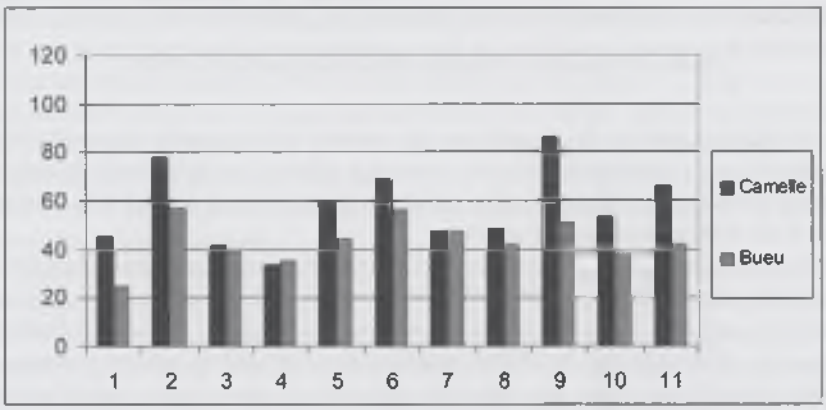


Figura 3, *"O médico falaba co cómico?"*: Duración dos enunciados proparoxítonos dos dous dialectos.

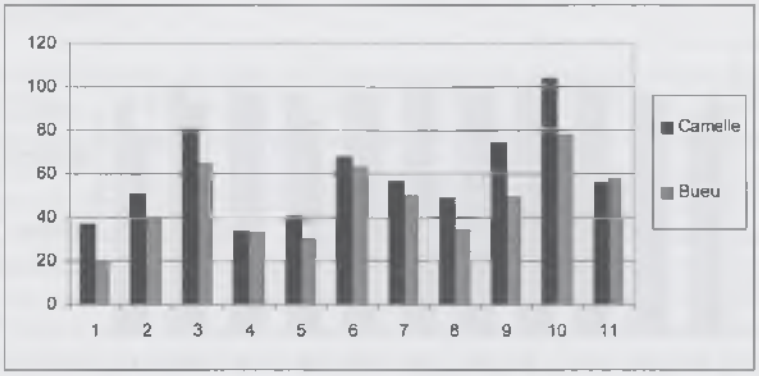


Figura 4 "*O soldado falaba co cabalo?*": Duración dos enunciados paroxítonos dos dous dialectos.

As diferenzas máis significativas entre as análises anteriores (Fernández Rei, Elisa et alii 2005a, 2006b) e as que nos atopamos agora, danse no comportamento do primeiro grupo acentual. Nos estudos anteriores observabamos que a informante acostumaba a realizar as terceiras sílabas máis longas cás do seu grupo acentual. En busca dunha posible explicación que nos sirva para comprender este feito, ocórresenos que, ou ben efectúa verdadeiramente estás sílabas cunha maior duración, ou ben esta prolongación vén dada pola escolla da natureza acentual das vogais das palabras que conforman o propio corpus. No novo córpora, ao contrario do que acontecía no anterior, as sílabas tónicas do suxeito son sempre abertas (ferrolán, soldado, médico).

3.4. Enerxía

No que respecta á enerxía, os gráficos que seguen amósannos o comportamento que este trazo presenta na modalidade interrogativa dos dous dialectos escollidos.

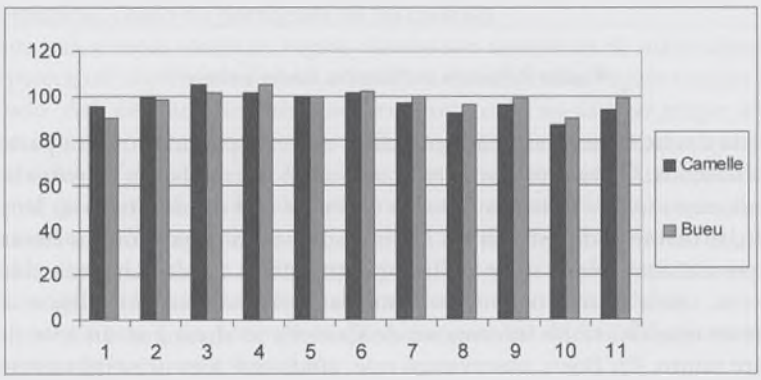


Figura 5: Enerxía en "*O ferrolán falaba co capitán?*"

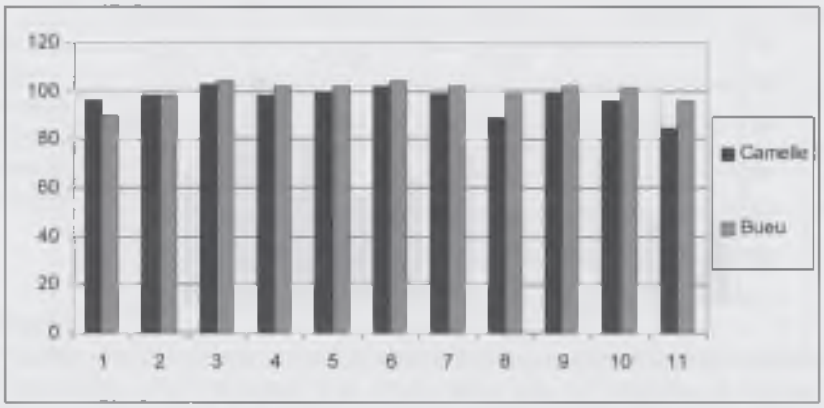


Figura 6: Enerxía en "*O soldado falaba co cabalo?*".

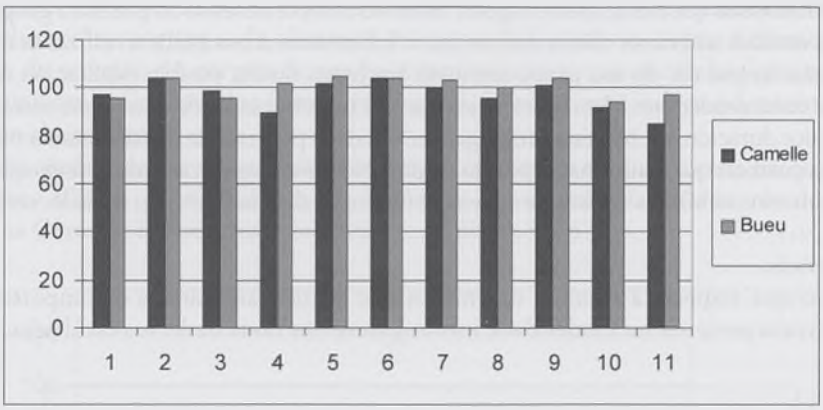


Figura 7: Enerxía en "*O médico falaba co cómico?*".

Á vista destas figuras non podemos dicir que este parámetro sirva para establecer diferenzas dialectais. Os valores que se reflicten son moi semellantes en todos os casos. Os gráficos amósannos unha tendencia lineal; a enerxía é alta e continuada ao longo de todo o enunciado, se ben se pode apreciar un lixeiro ascenso das vogais tónicas fronte ás átonas do seu grupo acentual. Nesta nova análise non se aprecia un descenso tan claro da enerxía no tonema, como viamos nas análises anteriores (Fernández Rei, Elisa et alii 2005a, 2006b). Nesta ocasión, só na informante de Camelle se distingue un leve descenso da enerxía neste punto. En Bueu, observamos que, aínda que nun principio semella que vai ter lugar este descenso da enerxía, esta tende a recuperarse na sílaba final do enunciado.

4. CONCLUSIÓNS

As interrogativas das dúas variantes lingüísticas do galego pertencentes ás comarcas do Morrazo e de Bergantiños presentan dous contornos melódicos distantes entre si e diferentes dos exemplos analizados noutros estudos sobre a fala de Santiago de Compostela (Fernández Rei et alii. 2005a, 2005b). Os puntos máis importantes nos que se poden resumir as diferenzas entre as variedades lingüísticas son os que seguen:

a) O inicio da interrogativa. O comezo da interrogativa de Bueu prodúcese nun ton baixo tendo en conta a frecuencia fundamental media da informante, mentres que a interrogativa de Camelle iníciase xa nunha altura considerable.
b) A altura tonal na que se desenvolve. En Camelle establécese unha diferenza significativa entre as sílabas pretónicas do suxeito, que se realizan nun ton moi alto con respecto á frecuencia fundamental media da informante, e o resto da oración (a partir da tónica do suxeito), que se realiza en frecuencias graves na súa tendencia á declinación final. A modalidade de Bueu comeza dun xeito moi semellante ao que describiamos para as interrogativas de Santiago (Fernández Rei et alii. 2005a, 2005b): prodúcese no suxeito unha subida progresiva, máis marcada na sílaba tónica, alcanzando nas postónicas do suxeito un ton alto, no que se manterá ata a tónica do complemento.
c) O máximo tonal. A colocación do pico tonal máis importante dentro do suxeito, que destaca sobre a relativa inmobilidade do resto da oración, é unha característica diferenciadora da interrogativa de Camelle. En Bueu, podemos falar de dous picos tonais relevantes un no suxeito e outro no complemento (na ascensión do tonema).
d) A realización do tonema. O final ascendente de Bueu nas interrogativas do tipo absoluto ou total destaca dentro das realizacións da maioría das interrogativas no galego, que se caracterizan por un final descendente. Este tonema aproxímase ás realizacións descritas polo proxecto AMPER noutras variedades románicas, como no portugués ou no castelán.
e) Nos tres acentos tonais as vogais tónicas son sempre as de maior duración.
f) Aparece un movemento tonal asociado á duración das vogais tónicas do enunciado. Así, en Camelle, tanto no primeiro como no último grupo acentual, o ton descende sempre nas vogais tónicas, que precisamente son as que presentan unha maior duración. En Bueu, no primeiro grupo acentual aparece un pico tonal que comeza na vogal tónica, que é a que máis dura, e que acada o seu máximo na(s) postónica(s). No entanto, no último grupo acentual, o ton cae sobre a vogal tónica, a máis longa, para despois ascender bruscamente na última vogal, independentemente da localización do acento léxico.
g) A enerxía non é un parámetro que sirva para establecer diferenzas dialectais, como tampouco o é a duración.

BIBLIOGRAFÍA

CARRIL, Ramón B., "Miscelánea. Notas de entonación gallega", *Revista de Filología Española*, t. LVI, cadernos 1º e 2º, 1973, 95-101.

CONTINI, M. / J.P. Lai / A. Romano / S. Roullet, "Vers un Atlas prosodique parlant des variétés romanes", in Mélanges offerts à Xavier Ravier, Sempre los camps auràn segadas resurgantas, Bouvier J-C, Gourc J. et Pic F. (eds), Toulouse, Université de Toulouse-Le Mirail FRAMESPA, Collection Méridiennes, 2002, 73-85.

CONTINI, M., "Vers une géoprosodie", in *Nazioarteko Dialektologia Biltzarra Agiriak*, Bilbao, Publ. Real Academia de la Lengua Vasca, 1992, 83-109.

CONTINI, M., Cl. Franchon Cabrera, A. Rhardisse, "Analyse comparée de l'intonation en castillan et en aragonais", in *Actes du Congrès International de Linguistique et Philologie Romane*, 1995, 143-157.

CRUTTENDEN, Alan, *Intonation*, Cambridge, Cambridge University Press, 1986.

FERNÁNDEZ PLANAS, Ana Mª., Eugenio Martínez Celdrán, Carlos van Oosterzee, Valeria Salcioli Guidi, Joan Castellví Vives e Dorota Szmidt Sierykow: "Proyecto AMPER: estudio contrastivo de frases interrogativas sin expansión del barceloní y del tarragoní", presentado no *VI Congreso de Lingüística General*, celebrado en maio de 2004, Santiago de Compostela (actas en preparación).

FERNÁNDEZ REI, Elisa, *Aproximación á entoación dos enunciados interrogativos en galego. Estudio Acústico*, Memoria de Licenciatura, Universidade de Santiago, 1995 (inédita).

FERNÁNDEZ REI, Elisa, "Contribución ó estudio da entoación das cláusulas interrogativas (totais e parciais) en galego", in Fernández Salgado, Benigno (ed.), *Actas do IV Congreso da Asociación Internacional de Estudios Galegos*, vol. I, 1997, 241-253.

FERNÁNDEZ REI, Elisa, "Situación actual das investigacións prosódicas sobre o galego", *Bolletino dell'Atlante Linguistico Italiano* (no prelo).

FERNÁNDEZ REI, Elisa, Manuel González González, Luís Xuncal Pereira e Maruxa Caamaño Varela, "Achega á entoación dunha fala do centro de Galicia. Contribución para o AMPER", *Géolinguistique*, Hors de Série, 3, 2005a, 87-102.

FERNÁNDEZ REI, Elisa, Ana Escourido Pernas, Maruxa Caamaño Varela e Luís Xuncal Pereira, "A entoación dunha fala de Santiago: fronteira prosódica entre suxeito e predicado", *Estudios de Fonética Experimental*, vol. XIV, 2005b, 141-165.

LAI, J.P., "Le sarde du Nuoro au sein du nouvel *Atlas multimédia prosodique de l'espace roman (AMPER)* ", *Géolinguistique* 9, 2004, 145-187.

LAI, J.P., A. Romano e S. Roullet, "Analisi dei sistemi prosodici di alcune varietà parlate in Italia: problemi metodologici e teorici", *Bollettino dell'Atlante Linguistico Italiano* 21, 1997, 23-70.

MOUTINHO, L., R. L. Coimbra, S. Secca Ruivo e U. Pereira Bendiha, "Atlas Prosódico Multimédia: Curvas De Uma Trajectória", póster presentado no *XVI Encontro Nacional da Associação Portuguesa de Linguística*, Coimbra, 2000.

MOUTINHO, L., R. L. Coimbra, S. Secca Ruivo e U. Pereira Bendiha, "Project d'Atlas Prosodique Multimédia des Variétés Romanes", *Travaux de L'Institut de Phonetique de Strasbourg* (TIPS), 31, 2001, 61-70.

PORTO DAPENA, José Antonio: *El gallego hablado en la comarca ferrolana*, Santiago de Compostela, *Verba*, anexo 9, 1977.

ROMANO, A., "Accent et intonation des parlers du Salento: une approche théorique et instrumentale", *Géolinguistique* 7, 1997, 93-132.

ROMANO, A., *Analyse des structures prosodiques des dialectes et de l'italien régional parlés dans le Salento (Italie): Approche linguistique et instrumentale*, thèse de Doctorat Nouveau Régime, Université Stendhal – Grenoble III, 1999.

SOBRINO PÉREZ, María dos Anxos, "A entoación do galego do Baixo Miño. Aproximación descritiva", *Cadernos de Lingua*, 20, 1999, 97-125.

SOBRINO PÉREZ, María dos Anxos, *A entoación do galego do Baixo Miño. Aproximación descritiva*, Memoria de Licenciatura, Universidade de Santiago, 1997 (inédita).

PATRONES ENTONATIVOS DEL CATALÁN EN HABLA ESPONTÁNEA

Dolors Font Rotchés
Laboratori de Fonètica Aplicada - Universitat de Barcelona

1. INTRODUCCIÓN

En esta comunicación, vamos a presentar los patrones entonativos del catalán en habla espontánea y sus principales características melódicas y fonológicas, obtenidos en una investigación de carácter experimental que hemos efectuado en el Laboratori de Fonètica Aplicada de la Universitat de Barcelona.[1]

Para llevar a cabo la investigación, nos hemos basado en el marco teórico y el método *Análisis melódico del habla* de Cantero (2002) porque creemos que responden mejor a nuestros objetivos básicamente por dos razones: ofrece un modelo completo que prevé la descripción fonética y fonológica de la entonación; y define un método de análisis del habla con medios instrumentales, objetivo y riguroso, que nos permite obtener valores exactos de la entonación, datos que serán útiles para poder desarrollar múltiples aplicaciones en la enseñanza de la lengua oral en los distintos niveles educativos, en el diagnóstico de patologías y reeducación de la voz a personas con trastornos auditivos o locutivos y en el desarrollo de las tecnologías del habla: síntesis de voz, reconocimiento del habla y sistemas de diálogo hombre-máquina.

Los estudios de la entonación del catalán más relevantes, que se han llevado a cabo desde los años 70 del siglo XX hasta nuestros días, han seguido de forma más o menos comprometida algunas propuestas lingüísticas que han aparecido a lo largo del siglo pasado -Navarro Tomás (1944), escuelas británica y norteamericana, modelo fonológico métrico y autosegmental iniciado por Pierrehumbert (1980), modelo de Aix-en-Proven-

---

1 Esta investigación se inscribe en el proyecto *Fonética aplicada a la educación* con el número BSO2002-03479 y financiado por la Comisión Interministerial de Ciencia y Tecnología (CICYT). Un desarrollo más amplio se encuentra en Font Rotchés (2005).

ce elaborado por Hirst, Di Cristo y Espesser (1998, 2000) o proyecto AMPER impulsado por Contini (Contini *et al.* 1998, 2002).

Estas propuestas de enfoques teóricos y métodos de análisis, así como también la escuela holandesa, que no hemos citado por no haberse aplicado en catalán, no desarrollan un modelo completo de descripción de la entonación de una lengua: algunas se preocupan esencialmente por explicar el fenómeno desde un punto de vista lingüístico y funcional, mientras que otras se preocupan por el análisis fonético, sin plantearse el rendimiento lingüístico; sin embargo, algunos de los modelos actuales están trabajando con la finalidad de ofrecer un modelo estándar completo que incluya la descripción fonética y fonológica.

En este sentido, estudios como el trabajo de Virgili y Blanquet (1971), los de fonética descriptiva de Bonet (1984, 1986) y Mascaró (1986, 1987), basados en los subdialectos barcelonés y menorquín, respectivamente, el de las interrogativas del central de Salcioli (1988a, 1988b), los de Prieto (1995, 2001, 2002), Payà (2001) o los de Carrera *et al.* (2004) y Fernández *et al.* (2004) sobre las interrogativas, constituyen descripciones parciales de la entonación del catalán porque hacen una descripción exclusivamente fonética o fonológica, a menudo basada en un solo dialecto o en un tipo de modalidad oracional y con pocos informantes. Estos estudios, sin embargo, hacen aportaciones que nos han servido para reflexionar sobre el tema en la elaboración de esta investigación y, a su vez, nos han motivado a dar un paso más e intentar describir la entonación del catalán.

2. EL CORPUS DE HABLA ESPONTÁNEA

Para hacer un estudio empírico de la entonación de la lengua catalana, el corpus tenía que basarse en habla espontánea, porque es el modelo que nos proporciona el habla real y genuina, la lengua en su estado natural sin la mediatización de aprendizajes lingüísticos adquiridos en un contexto escolar; tenía que contar con un número amplio y significativo de informantes catalanohablantes genuinos, suficientes y variados cuanto a sexo, edad y procedencia sociocultural y dialectal; y tenía que ser compensado, es decir, contener una representación equilibrada de los diversos perfiles de informantes y del tipo de enunciados.[2]

Partiendo de estos criterios, nos pareció que la televisión era el medio idóneo para obtener los datos necesarios ya que presenta toda una serie de ventajas que supera otros medios: variedad de programas, multiplicidad de informantes y buena calidad acústica de la señal de voz, indispensable para poder llevar a cabo el análisis acústico.

Así, pues, el corpus que elaboramos y que nos ha servido para definir los patrones melódicos del catalán se caracteriza por ser un corpus de habla espontánea, probablemente el más amplio y significativo que se ha creado hasta el momento para hacer un estudio de la entonación de una lengua, que cuenta con 580 contornos o melodías, seleccionados de 47 horas de grabaciones de material audiovisual del medio televisivo (TV3, Canal 33

2 El corpus de la investigación se trata de forma más amplia a Font Rotchés (en prensa).

y TVE2) emitidos entre el 1996 y el 2000, obtenidos en contextos de diálogo (concursos, entrevistas, debates, reportajes, *magazins*, etc.), y producidos por 160 informantes catalanohablantes genuinos, diversos y variados en cuanto al sexo (hombres y mujeres), edad (de 18 a 70 años) y procedencia sociocultural y dialectal (Cataluña, Baleares y Comunidad Valenciana).

3. LA APLICACIÓN DEL MÉTODO. *ANÁLISIS MELÓDICO DEL HABLA*

Una vez obtenido el corpus, se procesaron los datos basándonos en el método *Análisis melódico del habla*, cuyo aplicación hemos dividido en dos fases, acústica y perceptiva.

3.1 Fase acústica

Esta parte contempla desde la selección, segmentación y reconocimiento de las unidades del habla, hasta la obtención de datos acústicos, su estandarización y, posteriormente, la clasificación formal en grupos de los contornos obtenidos.

a) Análisis acústico y reconocimiento de las unidades del habla espontánea.

Para obtener los enunciados que van a formar parte del corpus y digitalizarlos, se seleccionaron intervenciones cortas que constituyen un turno de palabra del diálogo, aunque también se escogieron enunciados largos, que se dividieron provisionalmente respetando las unidades sintácticas y la melodía.

De cada enunciado provisional, se toman los valores de la F_0 (en hertzios) de las vocales, porque consideramos que son los valores relevantes de la melodía (las consonantes sonoras y glides tienen unos valores acústicos que no son esenciales para describir las melodías y las consonantes sordas no presentan ningún valor). Los valores frecuenciales de cada vocal se obtienen a partir de calcular el promedio de todos los valores que presenta. Para el análisis nos hemos servido del programa *MultiSpeech*, modelo 3700 de Kay Elemetrics y, posteriormente, se han comprobado los resultados obtenidos con el *Praat* de P. Boersma and D. Weenink del Institute of Phonetic Sciences de la University of Amsterdam.

El análisis acústico de los enunciados nos ha permitido definir los contornos o unidades del habla a partir de un criterio melódico: la inflexión final. Como consecuencia, hemos constatado que la mayoría de enunciados se correspondían a un contorno, mientras que algunos que presentaban más de una inflexión final ha sido necesario dividirlos en dos o tres contornos, según el caso. Se obtuvieron 580 contornos, una vez eliminados los que presentaban ruidos, músicas de fondo o solapamiento de voces.

Esta manera de proceder permite segmentar el habla sin depender de las unidades sintácticas (oraciones y sintagmas) ya que lo que delimitamos son contornos con una determinada melodía, que contienen, unas veces, sintagmas, y otras, oraciones; también permite analizar habla espontánea procedente de contextos múltiples en los que se puede conseguir una buena calidad de la señal acústica; por lo tanto, ya no es necesaria la pre-

sencia del investigador para inducir a los informantes a emitir unas frases muy concretas, ni tampoco se tienen que hacer las grabaciones de los informantes en el laboratorio, prácticas que hacen perder espontaneidad al habla y dificultan la amplitud y variedad del corpus.

b) El sistema de relativización de los datos acústicos y representación gráfica

Se han procesado los valores acústicos de las vocales de un contorno expresados en Hertzios y se han convertido en valores relativos o estandarizados. Los valores relativos de cada segmento vocálico se obtienen de calcular la distancia tonal en porcentaje entre una vocal y la siguiente empezando por 100% (cuando el movimiento tonal asciende el porcentaje es positivo y cuando desciende es negativo).

Una vez obtenido el algoritmo de los porcentajes y para poder proceder a la representación gráfica se empieza con un valor arbitrario 100 al que se aplica el porcentaje de ascenso o de descenso para obtener el valor siguiente y así sucesivamente. Estos valores estandarizados se validan al comprobar en el sintetizador de voz que la melodía estandarizada es idéntica a la original, pero con una tesitura de voz distinta. Véase en los gráficos 1 y 2 la representación de los valores absolutos (línea discontinua) y valores relativos o estandarizados (línea continua).

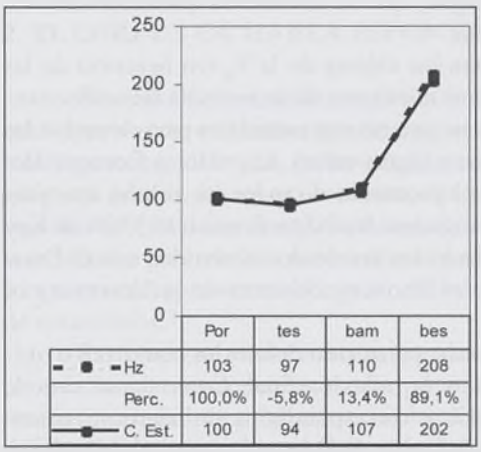


Gráfico 1. Representación gráfica de Portes bambes? '¿Llevas bambas?'

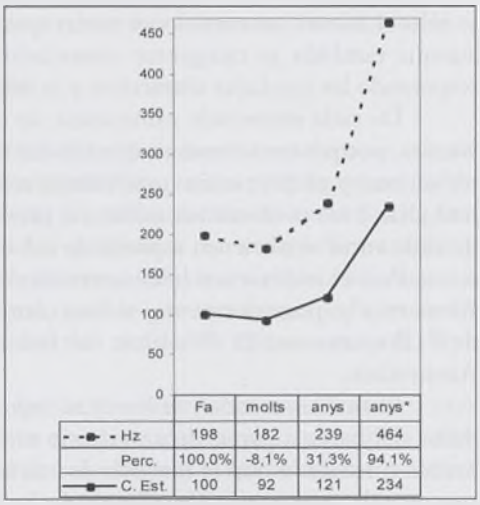


Gráfico 2. Representación gráfica de Fa molts anys? 'Hace muchos años'

El primer contorno fue producido por un hombre (voz más grave) y el segundo por una mujer (voz más aguda). Si comparamos la melodía en valores absolutos, solo se constata una tendencia ascendente en la inflexión final. Si nos fijamos en los valores relativos, vemos que ambas melodías son casi idénticas: la línea empieza en la vocal tónica en los

dos casos, sigue con un descenso (-5,8% el 1 y –8,1% el 2), después un ascenso (13,4% el 1 y 31,3% el 2) y acaba en un ascenso del 89,1% y del 94,1%, respectivamente. Si consideramos que las distancias tonales deben ser de un 10% como mínimo para ser percibidas por los informantes, constatamos que las diferencias entre ellas son prácticamente imperceptibles y, por lo tanto, melodías similares emitidas en tesituras distintas. Esta constatación se ha verificado en las respuestas de los informantes en las pruebas de percepción.

c) Clasificación formal según el criterio del rasgo melódico a la inflexión final (IF) y definición del significado fonológico de cada grupo

Si entendemos que la línea melódica de los contornos es un todo indivisible, es difícil clasificarlos en grupos por semejanzas de tipo formal, ya que, aparentemente son distintos en longitud, dirección de la línea, etc. y, a primera vista, parecen incomparables. En cambio, si dividimos en tres partes los contornos respetando la estructura acentual del grupo fónico, se nos facilita la tarea de comparación. Las tres partes son: anacrusis (segmentos tonales previos al primer segmento tónico del contorno, llamado *primer pico*), cuerpo (segmentos entre el primer pico y la última vocal) e inflexión final (núcleo del contorno y parte más significativa que va desde la última vocal tónica hasta el final).

Al comparar los contornos por la dirección de la inflexión final, hemos obtenido seis grupos provisionales (véase tabla 1) definidos por tres significados fonológicos a partir de los rasgos /±interrogativo/, /±enfático/, /±suspendido/. El rasgo fonológico /±interrogativo/ es /+I/ si los hablantes perciben su melodía como una pregunta, independientemente de que lo sea o no en el contexto; el rasgo /±enfático/ es /+E/ si la melodía se considera una exclamación; y el /±suspendido/ es /+S/, si se percibe como un enunciado inacabado. Y al revés, son /-interrogativo/, /-enfático/ y /-suspendido/ si no se perciben pregunta, ni exclamación, pero sí acabada, respectivamente. La unidad que contiene los tres rasgos que caracterizan el significado fonológico de cada grupo, se llama *tonema*, por analogía al fonema de la fonología segmental, también definido por rasgos fonológicos.

Grupo	Tonema
IF descendente	/-I-E-S/
IF ascendente (de 0% a 99%)	/-I-E+S/
IF ascendente (+100%)	/+I-E-S/
IF descendente-ascendente	/-I+E-S/
IF ascendente-descendente	/-I+E-S/
Sin inflexión final	/-I-E+S/

Tabla 1. Grupos provisionales de contornos

Aplicando esta primera fase del método, o fase acústica, se obtiene la estandarización individual de cada contorno, es decir, los algoritmos con su representación gráfica, que nos sirven para hacer comparables y clasificables los contornos, ya que se han normalizado los

valores. Como consecuencia, ja no debemos preocuparnos por la edad, el sexo o la tesitura del informante, ni tampoco por el número de informantes del corpus, ya que todos los contornos obtenidos, sean de quien sean, se han podido comparar y clasificar en seis grupos provisionales por coincidencias formales en la melodía. Ahora bien, los enunciados de los grupos tendrán que ser validados perceptivamente por hablantes, es decir, ser reconocidos como pertenecientes a un grupo. Pero todo esto ya forma parte de la segunda fase del método o fase perceptiva.

3.2 Fase perceptiva

Los contornos clasificados en grupos según el rasgo melódico de la inflexión final no sólo tienen que tener en común este rasgo, sino que también tienen que compartir significados fonológicos del tipo /±interrogativo/, /±enfático/, /±suspendido/, los cuales se comprobarán con pruebas de percepción aplicadas a informantes.

Estas pruebas, en que el informante debe responder si la melodía de cada enunciado descontextualizado que oye le sugiere una melodía de pregunta o no, enfática o no, e inacabada o no, permiten comprobar la validez de la clasificación formal de los contornos, basada en la dirección de la inflexión final, perfilar los rasgos fonéticos descriptivos típicos que definen cada grupo y hacer una interpretación fonológica.

Para comprobar la validez de los grupos provisionales, establecer los límites y conocer si existen más grupos, se procedió a preparar 127 enunciados del corpus para pasar 5 pruebas perceptivas a 153 informantes muy diversos. Para afirmar la existencia de un grupo con una inflexión final determinada, los informantes debían percibir un mismo significado fonológico.

Las preguntas eran del tipo: ¿Es un enunciado acabado? ¿Es una pregunta? ¿Es una exclamación? Ahora bien, la mayoría de enunciados fueron recortados para que los informantes no pudieran reconocerlos por los elementos gramaticales que tienen al principio, ya que se trataba de describir la melodía, no el tipo de oración.

De los resultados obtenidos, constatamos que se reconocen cinco grupos con sus significados fonológicos, pero con algunos cambios en los porcentajes de los límites melódicos y se crean tres nuevos:

- IF descendente, a partir de un 10% de ascenso, realización del tonema /-I-E-S/. El margen se sitúa desde un 10% de ascenso y todos los que presenten un descenso porque dentro de estos límites los contornos se consideran acabados (o rasgo /-suspendido/).
- IF ascendente (de 10% a 80%), realización del tonema /-I-E+S/. A partir de un ascenso de la inflexión final de un 10% y hasta un 80% los informantes perciben los contornos como no acabados (rasgo /+suspendido/).
- IF ascendente (≥80%), realización del tonema /+I-E-S/. El límite inferior de los contornos cuya melodía es percibida pregunta se sitúa en un 80% de ascenso en la inflexión final

El grupo IF descendente-ascendente se ha dividido en dos, según el ascenso a la inflexión final:

- IF descendente-ascendente (≥120%), realización del tonema /+I +E -S/ porque a partir de un ascenso del 120% los contornos se perciben como una pregunta y enfáticos.
- IF descendente-ascendente (de 10% a 120%), realización del tonema /-I+E+S/ porque entre un 10% y un 120% de ascenso los contornos se perciben enfáticos y suspendidos.

A parte de esta segregación, han aparecido dos grupos nuevos al poner a prueba una serie de contornos que presentaban un rasgo formal característico: alteración de la declinación del cuerpo del contorno causado por un ascenso en la línea melódica a partir de la sílaba anterior a la que contiene el acento sintagmático hasta ésta y, a continuación, un descenso hacia el siguiente y último segmento tonal, en el que la línea se dirige a un punto inferior al de partida del ascenso.

- IF con acento sintagmático elevado (≥50%), realización del tonema /+interrogativo +enfático –suspendido/, cuyos contornos se perciben interrogativos y enfáticos.
- IF con acento sintagmático elevado (de 10% a 50%), realización del tonema /-interrogativo +enfático -suspendido/, cuyos contornos se perciben enfáticos y acabados.

Grupos de contornos	Tonemas	Número de enunciados
IF descendente	/-I - E - S/	252
IF ascendente (de 10% a 80%)	/-I - E +S/	162
IF ascendente (≥80%)	/+I -E - S/	36
IF con acento sintagmático elevado (de 10% a 50%)	/-I +E - S/	42
IF ascendente-descendente	/-I +E - S/	26
IF descendente-ascendente (de 10% a 120%)	/-I +E +S/	21
IF con acento sintagmático elevado (≥50%)	/+I +E - S/	14
IF descendente-ascendente (≥120%)	/+I +E - S/	3

Tabla 2. Los grupos entonativos del catalán.

El grupo IF ascendente-descendente no se ha modificado. Por el contrario, el grupo compuesto por contornos sin inflexión final, entendemos que no se puede sostener como tal porque está formado por contornos truncados pertenecientes a grupos /±interrogativo//±enfático/. En realidad, no se pueden clasificar porqué les falta el rasgo melódico básico: la dirección de la inflexión final. Entendemos, pues, que constituyen una variante /+suspendida/, que aquí no trataremos.

4. PATRONES ENTONATIVOS DEL CATALÁN

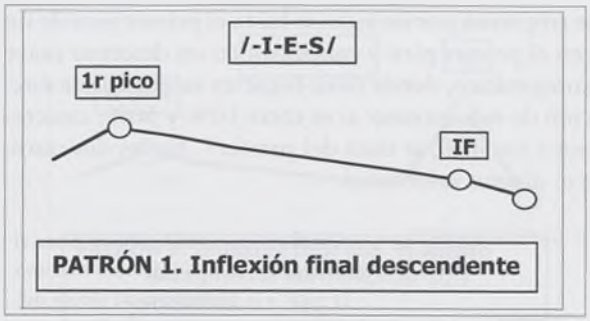
Resultado de la aplicación del método *Análisis melódico del habla*, hemos obtenido ocho grupos de contornos, definidos fonética y fonológicamente, que nos permitirán establecer los ocho patrones melódicos del catalán.

Cada uno de estos patrones se caracteriza por un conjunto de rasgos melódicos (al primer pico, al cuerpo y a la inflexión final) definidos por unos límites, superior e inferior, establecidos a partir de las diversas realizaciones de cada rasgo melódico en los contornos de un mismo grupo. Vemos pues que el patrón ya no se concibe como un modelo cerrado único y concreto, sino que es abierto y con unos límites amplios dentro de los que tienen lugar las infinitas realizaciones de las melodías de la lengua que nos sirven para transmitir emociones, intenciones y otros contenidos expresivos, así como para evidenciar diferencias socioculturales o dialectales. La zona entre los límites de cada rasgo constituye el margen de dispersión de un rasgo -concepto equivalente al de campo de dispersión de las vocales.[3]

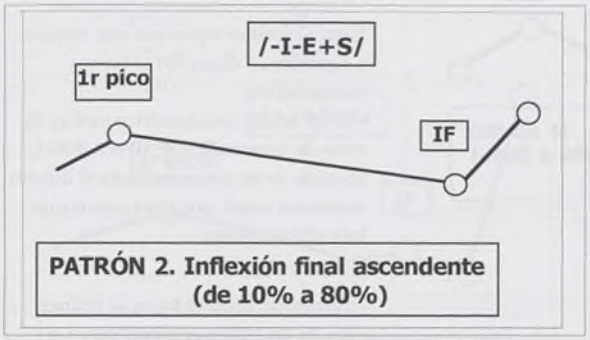
En los gráficos de los patrones que presentamos, aunque dibujamos una línea que expresa la tendencia del patrón, lo que de verdad lo define es la leyenda que lo acompaña, que constituye su caracterización fonética, así como su definición fonológica en la parte superior.

Los contornos del Patrón 1. IF descendente /-I-E-S/ o *patrón neutro* no se perciben pregunta ni enfáticos -por eso le llamamos *neutro*- (gráfico del patrón 1) y, como hemos visto en la tabla 2, son el grupo mayor con 252. Los contornos del Patrón 2. IF ascendente (de 10% a 80%) /-I-E+S/ o *patrón suspendido* se perciben inacabados (gráfico del patrón 2) y son 162. Los patrones 1 y 2 son los que presentan mayor número de contornos, 414, un 71% del total que tenemos. Los contornos del Patrón 3. IF ascendente (≥80%) /+I-E-S/ o *patrón interrogativo* (gráfico del patrón 3) se perciben como pregunta y tenemos un total de 36. Los tres patrones tienen en común que sus contornos empiezan por un ascenso hasta el primer pico de un 40% como máximo o directamente en el primer pico y continúan en un descenso suave hasta la última sílaba tónica; y se distinguen en la caracterización de la inflexión final: los contornos del Patrón 1 pueden presentar un ascenso de hasta un 10% o un descenso, los del Patrón 2 un ascenso situado entre un 10% y un 80% y los del Patrón 3 un ascenso igual o superior al 80%.

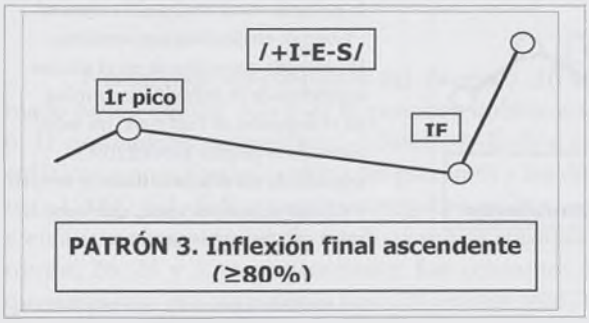
3 Para profundizar en el concepto de patrón melódico, véase en este mismo volumen el artículo de F.J. Cantero.



- Anacrusis (opcional): ascenso hasta el primer pico de un 40% como máximo.
- 1r pico: a la primera vocal tónica del contorno, que se encuentra en el punto más alto.
- Cuerpo: en declinación suave y constante.
- IF: ascendente hasta un 10% o descendente hasta un 40%.



-Anacrusis (opcional): ascenso hasta el primer pico de un 40% como máximo.
- 1r pico: a la primera vocal tónica del contorno.
- Cuerpo: en declinación suave y constante.
- IF: punto más bajo del contorno desde donde se inicia un ascenso de un 10% a un 80%.

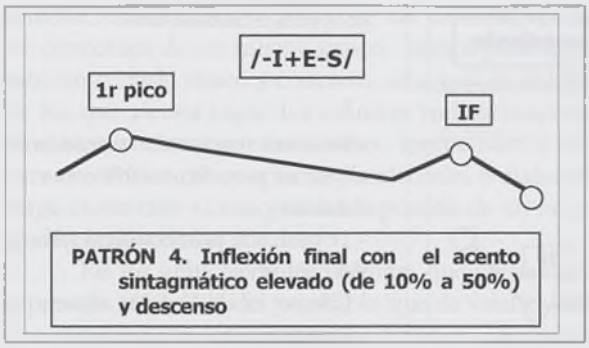


- Anacrusis (opcional): ascenso hasta el primer pico de un 40% como máximo.
- 1r pico: a la primera vocal tónica del contorno.
- Cuerpo: en declinación suave y constante.
- IF: punto más bajo del contorno desde donde se inicia un ascenso igual o superior a un 80%.

Los patrones restantes se caracterizan con el rasgo /+enfático/, aunque los tratamos en dos grupos por las coincidencias que presentan en el primer pico y cuerpo.

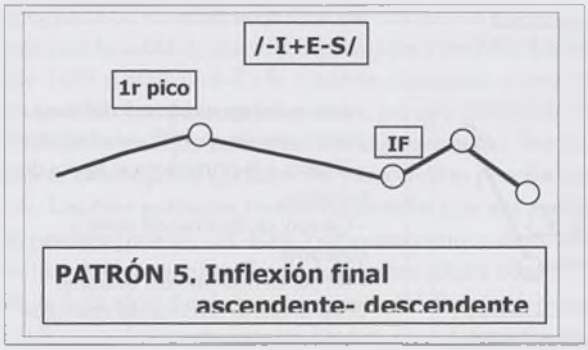
Los contornos del Patrón 4. IF con acento sintagmático elevado (de 10% a 50%) o *patrón enfático* /-I+E-S/ se perciben enfáticos (gráfico del patrón 4); y los del Patrón 7. IF con acento sintagmático elevado (≥50%) /+I+E-S/ o *patrón interrogativo enfático* se consideran interrogativos y enfáticos (gráfico del patrón 7). Los patrones 4 y 7, como se constata en la tabla 2, presentan pocos contornos, 42 y 14, respectivamente. Los contornos de

los dos patrones tienen en común que empiezan por un ascenso hasta el primer pico de un 40% como máximo o directamente en el primer pico y continúan en un descenso suave hasta la sílaba precedente al acento sintagmático, donde tiene lugar un ascenso hasta éste. Su diferencia se concreta en la elevación de este ascenso: si es entre 10% y 50%, caracteriza el patrón 4, y si es igual o superior a un 50%, se trata del patrón 7. En los dos casos, el patrón acaba en un descenso hasta el último valor tonal.



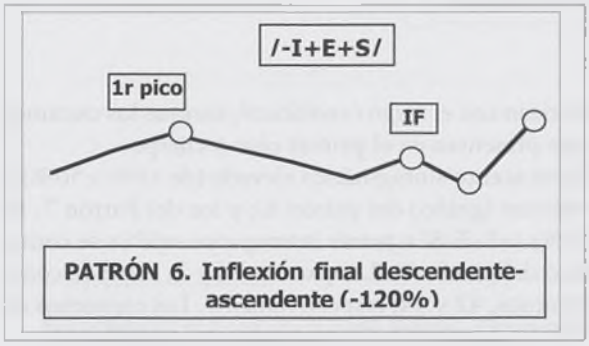
PATRÓN 4. Inflexión final con el acento sintagmático elevado (de 10% a 50%) y descenso

- Anacrusis (opcional): ascenso hasta el primer pico de un 40% como máximo.
- 1r pico: a la primera vocal tónica del contorno.
- Cuerpo: en declinación constante, hasta el último segmento que empieza a ascender en dirección al acento sintagmático.
- IF: el acento sintagmático está en la cima de un ascenso (de 10% a 50%), seguido de un descenso hasta el último segmento tonal, que es el punto más bajo del contorno.



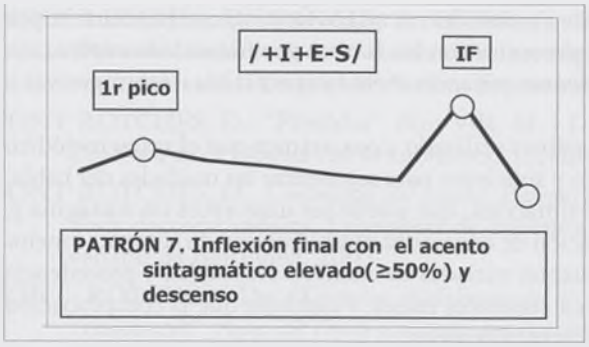
PATRÓN 5. Inflexión final ascendente- descendente

- Anacrusis: ascenso hasta el primer pico de un 50% como máximo.
- 1r pico: en una vocal átona posterior a la primera vocal tónica del contorno.
- Cuerpo: en declinación constante.
IF: ascenso que empieza en el primer segmento de la inflexión y culmina en el segundo, el cual tiene un valor parecido al primer pico (±20%), seguido de un descenso hasta el tercero y último segmento tonal, que tiene el valor más bajo del contorno.

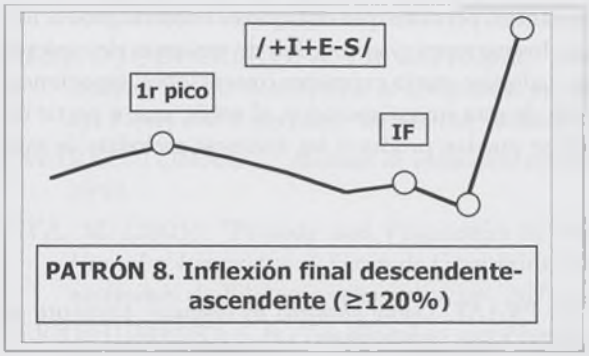


PATRÓN 6. Inflexión final descendente-ascendente (-120%)

- Anacrusis (opcional): ascenso hasta el primer pico de un 50% como máximo.
- 1r pico: en una vocal átona posterior a la primera vocal tónica del contorno.
- Cuerpo: en declinación constante.
- IF: descenso de un 10% a un 40% entre el primero y el segundo segmento, seguido de un ascenso inferior a un 120% hasta el último segmento del contorno.



- Anacrusis (opcional): ascenso hasta el primer pico de un 40% como máximo.
- 1r pico: a la primera vocal tónica del contorno.
- Cuerpo: en declinación constante, hasta el último segmento que empieza a ascender en dirección al acento sintagmático.
- IF: el acento sintagmático se encuentra en la cima de un ascenso (≥50%), seguido de un descenso hasta el último segmento tonal, que es el punto más bajo del contorno.



- Anacrusis: ascenso hasta el primer pico de un 50% como máximo.
- 1r pico: en una vocal átona posterior a la primera vocal tónica del contorno.
- Cuerpo: en declinación constante.
- IF: descenso de un 10% a un 40% entre el primero y segundo segmento, seguido de un ascenso inferior a un 120% hasta el último segmento del contorno.

Finalmente, los contornos del Patrón 5. IF ascendente-descendente, también llamado *patrón enfático,* /-I+E-S/ se perciben enfáticos (gráfico del patrón 5), los del Patrón 6. IF descendente-ascendente (-120%) /-I+E+S/ o *patrón enfático suspendido* se consideran enfáticos y no acabados (gráfico del patrón 6) y los del Patrón 8 IF descendente-ascendente (+120%) /+I+E-S/ o *patrón interrogativo enfático*, como el 7, se perciben como pregunta y enfáticos. Los patrones 5, 6 y 8, como se constata en la tabla 2, presentan pocos contornos, 26, 21 y 3, respectivamente. Los contornos de los tres patrones tienen en común que empiezan por un ascenso hasta el primer pico, situado en una vocal átona posterior a la primera vocal tónica y continúan en un descenso suave hasta la última sílaba tónica, donde empieza una inflexión tonal circunfleja. Su diferencia se concreta en las direcciones de la inflexión: ascendente-descendente para el patrón 5 y descendente-ascendente para el 6 y 8, los cuales se distinguen por el ascenso, inferior a un 120% el patrón 6 e igual o superior a un 120%, el patrón 8.

5. CONCLUSIONES

Resultado de esta investigación de carácter experimental, concluimos que el catalán tiene 8 patrones melódicos bien definidos fonéticamente a partir del rasgo melódico

de la inflexión final, y fonológicamente a partir de tres rasgos fonológicos que constituyen los tonemas. Los patrones que aquí ofrecemos son las formas genéricas de la entonación y cabe decir que cada uno puede presentar variantes melódicas enfáticas, interrogativas o suspensas que aquí no describimos.

Con referencia al método de análisis utilizado, constatamos que el rasgo melódico de la inflexión final nos ha sido válido y suficiente para segmentar las unidades del habla, las cuales coinciden con una unidad sintáctica, que puede ser unas veces un sintagma y, otras, una oración; que la estandarización de contornos nos ha permitido comparar melodías procedentes de múltiples informantes variados en cuanto a sexo, edad y procedencia sociocultural y dialectal producidas en contextos reales, y también que la comprobación perceptiva es la que da validez a los datos obtenidos.

Esta concepción de los patrones, caracterizados por unos márgenes amplios obtenidos a partir del análisis de 580 enunciados, producidos por 160 informantes y comprobados perceptivamente por 153 informantes, permite que cualquier contorno producido en catalán se pueda clasificar en uno de los patrones, ya procedan de variantes de carácter dialectal o sociocultural o signifiquen cualquier matiz expresivo (intenciones, emociones, órdenes, etc.), tal como se ha deducido de esta investigación y, al revés, que a partir de los datos precisos que hemos obtenido se puedan producir las distintas melodías de esta lengua mediante síntesis de voz.

BIBLIOGRAFÍA

BOERSMA, Paul & David WEENIK: *PRAAT. Doing phonetics by computer.* Institute of Phonetic Sciences, Univ. of Amsterdamn, 1992-2005. http://www.praat.org.

BONET, E.: *Aproximació a l'entonació del català*, Tesis de licenciatura inédita. Universitat Autònoma de Barcelona, 1984.

BONET, E.: "L'entonació de les formes interrogatives en barceloní". *Els Marges*, 33, 1986, pg. 103-117.

CANTERO SERENA, F. J.: *Teoría y análisis de la entonación.* Barcelona: Ed. de la Universitat de Barcelona, 2002.

CARRERA SABATÉ, J; VAN OOSTERZEE, C.; FERNÁNDEZ, A.M.; ROMERA, L.; ESPUNY, J. y E. MARTÍNEZ: "Les interrogatives al tortosí i al lleidatà. Un element diferenciador de subdialectes". *Estudios de Fonética Experimental.* XIII. Laboratori de Fonètica, UB, 2004, pp. 156-179.

FERNÁNDEZ, A.M.; MARTÍNEZ, E.; CARRERA, J; VAN OOSTERZEE, C.; SALCIOLI, V.; CASTELLVÍ, J. y D. SZMIDT: "Interrogatives absolutes al barceloní i al tarragoní (estudi contrastiu)". *Estudios de Fonética Experimental,* XIII, 2004, Laboratori de Fonètica, UB, pp. 130-155.

FONT ROTCHÉS, D.: *L'entonació del català.* Departament de Didàctica de la Llengua i la Literatura, Universitat de Barcelona, manuscrito inédito, 2000.

FONT ROTCHÉS, D.; CANALS, A.; ESTER, G.; HERMOSO, A. y F.J. CANTERO: "Patrones melódicos de la entonación interrogativa del catalán en habla espontánea". *Actas del II Congreso de Fonética Experimental,* Sevilla, 2002, pp. 192-197.

FONT ROTCHÉS, D.: "Prosòdia", dins Vilà, M. i D. Font (eds.) *Veu i locució. Articles de Didàctica de la Llengua i de la Literatura.* 32, enero-febrero-marzo 2004, pp. 25-38.

FONT ROTCHÉS, D.: *L'entonació del català. Patrons melòdics, tonemes i marges de dispersió.* Tesis doctoral inédita. Laboratori de Fonètica Aplicada (Campus Mundet). Universitat de Barcelona, 2005.

FONT ROTCHÉS, D.: "Criteris d'elaboració d'un corpus oral per a l'estudi de l'entonació". *Actes del Tretzè Col·loqui International de Llengua i Literatura Catalanes.* Septiembre del 2003, Universitat de Girona (en prensa).

HIRST, D.J. i A. Di CRISTO (eds.): *Intonation Systems. A Survey of Twenty Languages.* Cambridge: Cambridge University Press, 1998.

HIRST, D.J.; Di CRISTO, A. i R. ESPESSER: "Levels of representation and levels of analysis for the description of intonation systems", dins M. HORNE (ed.): *Prosody: Theory and Experiment.* Dordrecht: Kluver Academic Press, 2000, pp. 51-87.

NAVARRO TOMÁS, T.: *Manual de entonación española.* New York, Hispanic Institute, 1944.

PAYÀ, M. (2003): "Prosody and Pragmatics in Parenthetical Insertions in Catalan". *Journal of Linguistics, 2.* Grup de Gramàtica Teòrica de la UAB y l'Institut Interuniversitari de Filologia valenciana, pp. 207-227.

PIERREHUMBERT, J. B.: *The Phonology and Phonetics of English Intonation.* Tesi doctoral. MIT, 1980.

PIERREHUMBERT, J. B.: *The Phonology and Phonetics of English Intonation.* Bloomington: Indiana University Linguistics Club, 1987.

PRIETO i VIVES, P.: "Aproximació als contorns tonals del català central" *Caplletra*, 19, 1985, pp. 161-186.

PRIETO i VIVES, P.: "Patrons d'associació de l'estructura tonal en català". *Catalan Working Papers in Linguistics.* Bellaterra, 7, 1999, pp. 207-218.

PRIETO i VIVES, P.: "L'entonació dialectal del català: el cas de les frases interrogatives absolutes", dins Bover *et al.* (ed.): *Actes del Novè Col·loqui d'Estudis Catalans a Nord-Amèrica.* Barcelona, Publicacions de l'Abadia de Montserrat, 2001, pp.347-377.

PRIETO i VIVES, P.: "Entonació", dins Joan Solà (dir.): *Gramàtica del Català Contemporani,* vol. 1. Barcelona: Empúries, 2002, pp. 393-462.

SALCIOLI, V.: *La entonación: estudio fonético-experimental de la entonación interrogativa catalana.* Tesis doctoral. Barcelona, Universitat de Barcelona, Departament de Filologia Romànica, 1988.

SALCIOLI, V.: "Estudio fonético-experimental de la entonación interrogativa catalana". *Estudios de Fonética Experimental*, III, 1988, pp. 37-69.

't HART, J., R. COLLIER & A. COHEN: A Perceptual Study of Intonation. An Experimental-Phonetic Approach to speech melody. Cambridge: Cambridge University Press.

EL NIVEL DE BILINGÜISMO Y LA DISCRIMINACIÓN PERCEPTIVA DE VOCALES INGLESAS

FRANCISCO GALLARDO DEL PUERTO
Universidad del País Vasco

Los estudios que investigan el papel que juega el nivel de bilingüismo en la adquisición de una tercera lengua parecen señalar que un mayor nivel de competencia bilingüe favorece el aprendizaje del tercer idioma. Distintas variables asociadas al nivel de bilingüismo como la lengua nativa, la lengua de instrucción escolar, el nivel de uso de cada idioma y el grado de instrucción formal o de alfabetización en cada lengua parecen ser factores importantes a la hora de determinar el nivel de competencia lingüística alcanzado en la tercera lengua. Sin embargo, muy pocos han sido los estudios que han investigado la relación entre el nivel de bilingüismo y la competencia fonética en la tercera lengua. Éstos, además, se han centrado principalmente en la producción fonética. El presente trabajo pretende abordar esta cuestión en el terreno de la percepción fonética. Para ello, dos grupos de hablantes bilingües que diferían con respecto al nivel de uso de cada una de sus dos lenguas (castellano-euskera) realizaron una prueba de discriminación de pares mínimos de vocales inglesas. Los análisis estadísticos señalaron la ausencia de diferencias significativas entre los dos grupos de aprendices bilingües en la percepción de las vocales. Varios factores podrían explicar la divergencia entre los resultados de nuestra investigación y los de los estudios que descubren diferencias en función del nivel de bilingüismo: (1) el aspecto específico en estudio (el componente vocálico en nuestro caso) y las diferencias entre las lenguas con respecto a este aspecto, (2) la distinción entre percepción y producción fonética y (3) el tipo de escolarización, que puede mezclar o no en las mismas aulas a los distintos tipos de aprendices bilingües.

1. INTRODUCCIÓN

Son varios los estudios que asocian un mayor nivel de bilingüismo con una mayor habilidad para el aprendizaje de una tercera lengua. Existen una serie de factores

relacionados con el nivel de bilingüismo que parecen condicionar el mayor o menor éxito en el aprendizaje de terceras lenguas. Entre estos factores comenzaremos mencionando el nivel de alfabetización en cada una de las dos lenguas de un hablante bilingüe. Swain et al. (1990) llevaron a cabo un estudio en Toronto (Canadá) en el que descubrieron que el nivel de competencia lingüística general (lectura, escritura, escucha y habla) logrado en la tercera lengua (francés) por los inmigrantes que estaban alfabetizados tanto en sus primeras lenguas como en su segunda lengua (inglés) era mejor que el de los inmigrantes que sólo estaban alfabetizados en inglés. Con anterioridad, Thomas (1985, 1988) había dividido a sus aprendices bilingües (español/inglés) en función de si habían recibido instrucción formal en su primera lengua (español) o no, descubriéndose que aquellos estudiantes que sí lo habían hecho presentaban ventajas en la adquisición de la tercera lengua (francés) sobre aquellos que sólo habían aprendido español de una manera informal en casa.

En España, Bernaus (1996) examinó la competencia lingüística en español, catalán e inglés de escolares cuya lengua de instrucción era el catalán, concluyendo que una mejor competencia en la lengua minoritaria de la comunidad (catalán) contribuía a unos mejores resultados en la tercera lengua (inglés). De forma similar, Muñoz (2000) demostró con posterioridad que los escolares cuya competencia tanto en catalán como en castellano es mayor habían desarrollado una mayor habilidad para el aprendizaje de la lengua inglesa. Lasagabaster (1998) llevó a cabo un experimento con escolares bilingües (euskera/español) de enseñanza primaria cuyos resultados indicaron que variables tales como la lengua de instrucción escolar, la lengua nativa o la lengua de mayor uso son factores que ejercían un efecto significativo sobre el nivel general de inglés oral y escrito, Cuando esta lengua era el euskera (lengua minoritaria de la comunidad), los aprendices bilingües lograban mejores resultados en inglés. La variable de mayor impacto resultó ser la lengua de instrucción escolar. El inglés de los estudiantes escolarizados en el modelo D (instrucción en euskera) era mejor que el de los escolarizados en el modelo B (instrucción en euskera y español), el cual, a su vez, resultó mejor que el de los alumnos escolarizados en el modelo A (instrucción en castellano). En la misma línea, Sagasta (2001) descubrió sus estudiantes bilingües, todos ellos escolarizados en el modelo D, mostraban distinto nivel de competencia escrita en inglés en función del grado de uso del euskera y el español. Su estudió reveló que los estudiantes pertenecientes a ambientes vascoparlantes obtenían mejores resultados que los alumnos cuyo uso del castellano era mayor.

Todos estos estudios parecen indicar que aquellas variables indicadoras de un mayor nivel de bilingüismo repercuten en una mejor adquisición de la tercera lengua. Sin embargo, muy pocos han sido los experimentos que han investigado la relación entre el nivel del bilingüismo y la competencia fonológica en la tercera lengua, si bien podemos mencionar alguno. Cohen et al. (1967), por ejemplo, analizaron la habilidad para repetir una serie de grupos consonánticos en posición inicial en tres grupos distintos de bilingües (nativos en inglés, nativos en francés, nativos en ambos idiomas). Los grupos consonánticos podían estar presentes en inglés y francés, sólo en inglés, sólo en francés, o ser totalmente nuevos. Los resultados indicaron que los bilingües nativos en francés

siempre eran menos exactos en sus repeticiones, sobre todo en la de los grupos presentes únicamente en lengua inglesa. Los autores no dejaron pasar el hecho de que este grupo de bilingües mostraba un menor nivel de bilingüismo debido probablemente a que comenzaron la exposición a la segunda lengua (inglés) a una edad más tardía que los otros dos grupos. Mas recientemente, González Ardeo (2001) realizó una investigación sobre la adquisición de la pronunciación del inglés como tercera lengua por parte de estudiantes universitarios bilingües (español/euskera). Este autor dividió a sus sujetos en función de dos variables indicadoras del nivel de bilingüismo la lengua nativa y la lengua de instrucción escolar. La primera de estas dos variables resultó reveladora al comprobarse que los estudiantes bilingües nativos de euskera, lengua minoritaria, pronunciaban mejor una serie de palabras en inglés que los bilingües nativos de español, lengua mayoritaria.

El presente trabajo pretende ahondar en la relación entre el nivel de bilingüismo y la competencia fonológica en la tercera lengua a través de un experimento sobre la percepción de vocales en inglés como tercera lengua por parte de escolares bilingües en el País Vasco.

2. METODOLOGÍA

Los participantes de nuestro estudio son estudiantes de primaria y secundaria, todos ellos escolarizados en un mismo centro escolar en el País Vasco. En dicho centro, la lengua de instrucción es el euskera. Todos nuestros sujetos son hablantes bilingües (euskera-castellano), por lo que el inglés constituye la tercera lengua. Recordemos, no obstante, que en el País Vasco el euskera y el castellano no tienen el mismo estatus, siendo lengua minoritaria aquélla y mayoritaria ésta. Este hecho es importante ya que el mayor equilibrio bilingüe se da conforme existe un mayor uso de la lengua minoritaria (Gabiña et al., 1986; Lasagabaster, 1998; Sagasta, 2001; Sierra y Olaziregi, 1989, 1990, 1991, 1992). Los estudiantes de este centro rellenaron un cuestionario sobre el uso del euskera y el castellano en su vida diaria. Para esta investigación seleccionamos, por un lado, aquellos alumnos que mostraban un muy alto grado de utilización del euskera en detrimento del uso del castellano y, por otro lado, aquellos estudiantes cuyo uso del euskera era menor en favor de la utilización del castellano. De esta manera, contábamos,

	Edad de Inicio	*Media de Edad*	*Uso del Euskera (max=10)*	*Exposición al inglés*
+E4 (N = 10)	4	10	9.4	6 años
-E4 (N = 10)	4	10	3.1	6 años
+E8 (N = 10)	8	14	9.7	6 años
-E8 (N = 10)	8	14	3.7	6 años
+E11 (N = 10)	11	17	9.7	6 años
-E11 (N = 10)	11	17	3.1	6 años

Tabla 1: Distribución de la muestra

por lado, con un número de hablantes bilingües que mostraban un mayor equilibro entre sus dos lenguas (+E) y, por otro lado, con una serie de hablantes cuyo equilibrio bilingüe era menor (-E). Dentro de cada unos de estos dos tipos de bilingües teníamos alumnos que comenzaron la exposición al inglés a los 4 (+E4, -E4)), a los 8 (+E8, -E8) y a los 11 años (+E11, -E11). En la tabla 1 se presentan las características de nuestra muestra:

Se diseño una prueba de percepción de vocales inglesas que consistía en la presentación de 22 pares mínimos de palabras monosílabas que diferían única y exclusivamente en la vocal. Estos pares incluían todas las vocales del inglés R.P. a excepción de la vocal débil *schwa*, ya que ésta no puede ocurrir en monosílabos acentuados. Cada vocal inglesa se oponía a otras dos vocales inglesas, aquéllas que en un experimento anterior con hablantes de castellano y euskera (García Lecumberri y Cenoz, 1997) habían resultado ser las que mayor confusión creaban en cada caso.

3. ANÁLISIS

Se calcularon las medias de discriminación correcta y las desviaciones típicas para cada uno de los seis grupos. Además, se llevaron a cabo una serie de pruebas T para comprobar si las diferencias entre los dos tipos de aprendices bilingües eran significativas ($S \leq 0.05$).

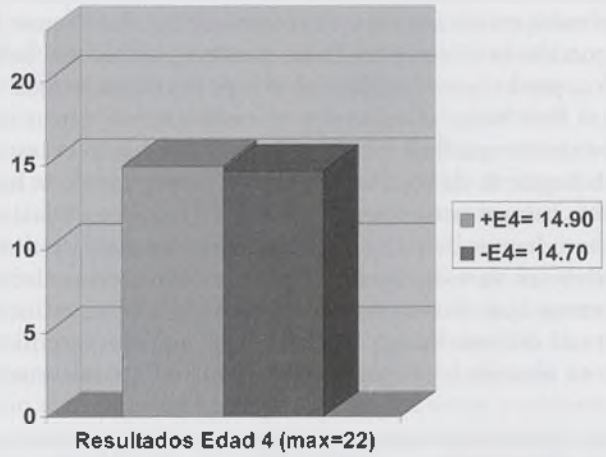


Figura 1: Medias en los dos grupos bilingües de menor edad

La figura 1 presenta las medias de percepción correcta de los dos grupos bilingües que comenzaron su exposición al inglés a la edad de 4 años. Como se observa, ambos grupos obtuvieron medias muy cercanas, no descubriéndose diferencias significativas en su comparación (T= 0.17; S= 0.86). Además, las desviaciones típicas resultaron ser muy pequeñas y también muy parecidas (MB1 = 2.47; LB1 = 2.63), lo que indicaba que los dos grupos de bilingües muestran un comportamiento igualmente homogéneo.

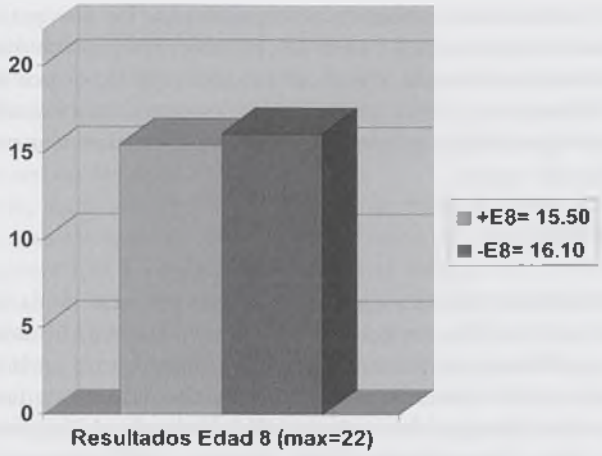


Figura 2: Medias en los dos grupos bilingües de edad intermedia

La figura 2 ofrece las medias de discriminación correcta de los dos grupos bilingües que empezaron la instrucción en inglés a los 8 años. De nuevo, las medias de los dos tipos de bilingües son tan parecidas que la prueba T no arrojó diferencias significativas (T= -0.51; S= 0.61). Las desviaciones típicas de ambos grupos son nuevamente bajas y muy parecidas entre sí (MB2= 2.51; LB2= 2.73), lo que se traduce en una discriminación vocálica igualmente homogénea en ambos tipos de aprendiz bilingüe.

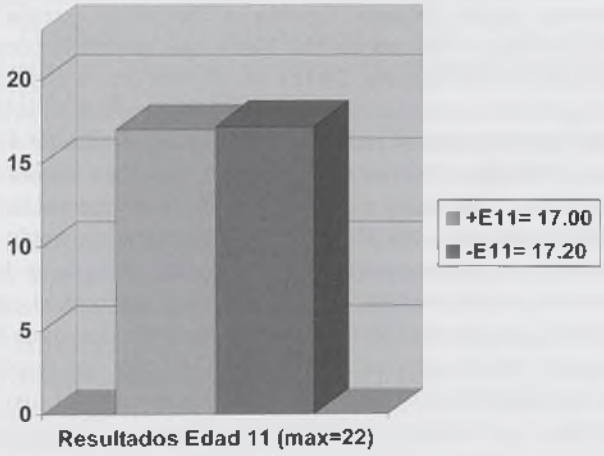


Figura 3: Medias en los dos grupos bilingües de mayor edad

La figura 3 presenta las medias de percepción para los dos grupos bilingües que empezaron las clases de inglés a los 11 años. Dichas medias fueron de nuevo tan parecidas que no se descubrieron diferencias significativas entre los dos tipos de bilingües (T= -0.21; S= 0.83). Además, a la vista de las desviaciones típicas obtenidas (MB3= 2.00; LB3= 2.15), vemos que ambos grupos muestran una consistencia muy similar en la percepción vocálica del inglés.

4. CONCLUSIONES

Nuestros resultados indican que, independientemente de la edad a la que se comenzó la exposición al inglés, los hablantes bilingües más equilibrados y los hablantes bilingües menos equilibrados muestran una habilidad perceptiva en la lengua extranjera igual desde el punto de vista estadístico. Estos resultados, por consiguiente, no apoyan la idea procedente de los hallazgos de varios estudios (Bernaus, 1996; Cohen et al., 1967; González Ardeo, 2001; Lasagabaster, 1998; Muñoz, 2000; Sagasta, 2001; Swain et al.,1990; Thomas, 1985, 1988) según la cual un mejor nivel de bilingüismo tiene efectos positivos sobre el aprendizaje de lenguas adicionales. Sin embargo, Cenoz (2003) asegura que las ventajas asociadas al bilingüismo en la adquisición de terceras lenguas son mucho más consistentes cuando se analiza la competencia lingüística general que cuando se miden aspectos lingüísticos mucho más específicos, lo que explicaría que en el caso de nuestro experimento sobre la percepción vocálica no se descubrieran tales ventajas.

Además, no podemos dejar pasar por alto el hecho de que el euskera y el español comparten el mismo sistema vocálico y presentan una fonología segmental muy parecida, mientras que en otras áreas lingüísticas, como la sintaxis o la morfología, estos dos idiomas muestran una mayor distancia tipológica. No es de extrañar, por tanto, que las investigaciones llevadas a cabo en el País Vasco que miden la competencia general en inglés (Lasagabaster, 1998; Sagasta, 2001), ya sea ésta oral o escrita, señalan que un mejor nivel de bilingüismo se traduce en una mejor competencia en la lengua extranjera, mientras que en el caso de nuestro estudio sobre la adquisición de vocales es como si nuestros estudiantes bilingües fueran en realidad aprendices monolingües ya que el hecho de ser bilingüe no añade nada a este aspecto de la competencia. Sería interesante realizar con nuestra muestra un estudio similar sobre la adquisición del componente consonántico o de aspectos suprasegamentales del inglés, campos en los que el español y el euskera difieren en mayor medida. Tal vez entonces, como en el caso del estudio de González Ardeo (2001), se descubrirían diferencias entre los dos tipos de aprendiz.

Otro factor que podría explicar la divergencia entre los resultados de nuestro estudio y los de otros experimentos fonéticos sobre los efectos del nivel de bilingüismo sería distinción entre la percepción y la producción fonéticas. Tanto el trabajo de Cohen et al (1967) como el de González Ardeo (2001), en los que un mayor equilibrio bilingüe se asocia con una mejor pronunciación de la tercera lengua, versan sobre la producción fonética, mientras que nuestro experimento analiza la percepción auditiva de la lengua extranjera. De hecho, existen varios estudios de percepción en los que se comparan

hablantes monolingües y bilingües adquiriendo una lengua adicional (Davine et al., 1971; Lambert & MacNamara, 1969; Werker, 1986) en los que no se encuentran diferencias significativas entre ambos tipos de aprendiz, lo que vendría a apoyar nuestros resultados de percepción. Sería muy interesante obtener datos sobre la producción fonética en inglés de los participantes de nuestro estudio para comprobar si lo descubierto en la producción corrobora o no nuestros resultados de percepción.

Por último, mencionaremos un aspecto de nuestra investigación que también podría explicar la ausencia de los efectos beneficiosos del equilibrio bilingüe. Nuestros dos grupos de aprendices bilingües pertenecían al mismo centro escolar y compartían las mismas aulas, mientras que en la mayoría de las investigaciones (e.g.: Bernaus, 1996; Lasagabaster, 1998; Muñoz, 2000; Sagasta, 2001) los distintos tipos de hablantes bilingües o bien pertenecían a distintos centros escolares o, si se encontraban escolarizados en el mismo centro, asistían a clases distintas. Este hecho ha podido llevar a una mayor homogenización de del aprendizaje de la tercera lengua en nuestro caso. Sería muy interesante medir a nuestros participantes en otros aspectos lingüísticos del inglés (vocabulario, gramática, etc.) para descubrir cuál de las dos variables, el aspecto lingüístico en estudio o el tipo de escolarización, ejerce un mayor peso.

REFERENCIAS BIBLIOGRÁFICAS

BERNAUS, M.: *Introduction of a Third Language from Kindergarten. How It Can Affect L1 and L2 Acquisition.* Barcelona: APAC of News (1996).

CENOZ, J.: "The additive effect of bilingualism on third language acquisition". *The International Journal of Bilingualism* 7 (2003), pp.71-88.

COHEN, S.P., TUCKER, G.R. y LAMBERT, W.E.: "The comparative skills of monolinguals and bilinguals in perceiving phoneme sequences". *Language and Speech* 10 (1967), pp. 159-168.

DAVINE, M., TUCKER, R. y LAMBERT, W.E.: "The perception of phoneme sequences by monolingual and bilingual elementary school children". *Canadian Journal of Behavioural Sciences* 3 (1971), pp.72-76.

GABIÑA, J.J., GOROSTIDI, R., IRURETAGOIENA, R., OLAZIREGI, I. y SIERRA, J.: *EIFE-1 Euskararen Irakaskuntza: Faktoreeen Eragina.* Vitoria-Gasteiz: Central Publications of the Basque Goverment (1986).

GARCÍA LECUMBERRI, M.L. y CENOZ, J.: "Identification by L2 learners of English vowels in different phonetic contexts". En J. Leather y A. James (eds.). *New Sounds 97: Proceedings of the III International Symposium on the Acquisition of a Second Language Speech.* Klagenfurt: University of Klagenfurt (1997), pp. 196-205.

GONZALEZ ARDEO, J. M.: "Engineering students and ESP in the Basque Country: SLA versus TLA". En J. Cenoz, B. Hufeisen y U. Jessner (eds.). *Looking Beyond Second Languag Acquisition: Studies in Tri- and Multilingualism.* Tübingen: Stauffenburg (2001), pp. 75-95.

AMPER-CUBA: PRIMEROS RESULTADOS EN UN PROYECTO CONJUNTO

RAQUEL GARCÍA RIVERÓN
Centro de Investigaciones Sociales, La Habana
FRANCISCO FERNÁNDEZ PÉREZ-TERÁN
CUJAE, La Habana

Es conocido que el método investigador geolingüístico contempla dos variantes o modos de hacer, que se diferencian de manera esencial desde el punto de vista epistemológico, aunque ontológicamente se ocupen del mismo objeto de estudio: la lengua en su diversidad geográfica. Así, el método geolingüístico se plantea la elaboración de *atlas de pequeños dominios y atlas de grandes dominios.* Por el método geolingüístico, el proyecto AMPER[1]. (Atlas Multimedia de la Prosodia del Espacio Románico) se inscribe dentro de los *atlas lingüísticos de grandes dominios.* Como atlas de gran dominio que es, se traza el objetivo inicial de describir comparativamente la diversidad geográfica de pocas variables (el acento y la entonación) en dos tipos de oraciones, declarativas e interrogativas, de las lenguas del mundo romance. Encarar esta empresa es algo que pocos especialistas en geografía lingüística se han puesto como meta.

Según el tratamiento de la entonación, el AMPER se inscribe en la tradición europea del siglo XX, tradición estructuralista que tiende al estudio de los sistemas reduciendo el número de variables del nivel estructural estudiado[2]. Por esta razón y teniendo en cuenta que es un Atlas de Prosodia (en general), se encara al unísono la indagación de varios de los mencionados niveles con el fin de tratar de aprehender el objeto. La importancia del AMPER está dada porque no contamos en nuestros países con trabajos *tipológicos* sistemáticos sobre la Prosodia en amplios espacios territoriales.

La propia condición de atlas de gran dominio que se dedica a la indagación de un objeto tan poco estudiado permitirá allegar datos importantes para el mundo romance en su conjunto. No obstante, esta visión *extensiva* del AMPER comporta la necesidad de incluir algunas variables de escasa o nula trascendencia regional en aras de obtener resul-

1 Contini, 1992; Fernández Planas, 2005; Dorta y Hernández, 2004.
2 Sobre el estado de la Entonología dentro del Hispanismo, véase García Riverón 2005.

tados generales válidos. Por ello, el AMPER, como cualquier proyecto de esta característica, demostrará su productividad y su capacidad descriptiva y explicativa en algunos aspectos particulares solamente si tenemos en cuenta el estado de la cuestión en aquellas áreas de nuestra geografía en las cuales se hayan realizado investigaciones sobre la entonación, y así, según se actúa en la labor científica, podremos lograr resultados parciales, regionales, más ricos en profundidad, complejidad del sistema indagado y magnitud de los datos mismos. Es necesario llegar a establecer los caminos seguros que permitan definir de modo certero y definitivo la relación dialéctica entre *lo general* y *lo particular* en un objeto tan *complejo* como es la entonación.

Entonces, en nuestro criterio es imprescindible validar, conjugar y, en su caso, acotar lo métodos y técnicas del AMPER en su relación con los proyectos en ejecución en las diferentes regiones. Conjugar y acotar algunos métodos y técnicas es el objetivo de este trabajo inicial que hoy presentamos.

El AMPER-Cuba se realiza en coordinación directa con el ámbito lingüístico español que para la Península e Iberoamérica dirige el profesor, Dr. Eugenio Martínez Celdrán.

EL AMPER CUBA Y SUS ANTECEDENTES: EL SISTEMA DE LA ENTONACIÓN DEL ESPAÑOL DE CUBA Y LA ENTONACIÓN EN EL ATLAS LINGÜÍSTICO DE CUBA (ALCU)

Cuando se comenzó el estudio del *sistema de la entonación* del español de Cuba, no podíamos actuar sin definirlo como objeto de investigación. Desde el punto de vista ontológico, la lengua es el sistema esencial de comunicación humana, sistema de innegable fundamento psico-social y cultural. Esta es la esencia ontológica de uno de sus microsistemas: el sistema de la entonación. Las vertientes psíquica y social de la lengua permiten determinar el sistema de la entonación como un *sistema dinámico complejo no lineal* en el cual por definición los niveles estructurales inferiores no influyen para *establecer el sistema en sí*.

Entonces, el antecedente de la visión que hoy se ofrece, antecedente que permitirá nuestra integración con el proyecto AMPER, es la investigación del sistema de la entonación de español de Cuba[3], definido como sistema complejo y estudiado acorde con las pautas epistemológicas del *análisis comunicativo*. Este análisis se fundamenta en el estudio de las posibles *interacciones* de los diferentes medios de expresión de la lengua en textos orales.

Según el enfoque desarrollado, se ha logrado segmentar el discurso prosódico a partir de un grupo de *oposiciones* establecidas. En consecuencia, en una primera etapa se comenzó con la investigación *cualitativa* de los fenómenos: la discretización semántico-pragmática de los actos de habla en los cuales interactuaba la entonación en el discurso.

3 Por razones de espacio, salvo que se haga alguna aclaración, cualquier referencia al sistema de la entonación se encontrará en García Riverón, 1996-1998.

El análisis preliminar permitió obtener un enfoque abarcador y ext*ensivo* en lo que a las unidades de la entonación se refiere

Esta forma de discernir permitió diferenciar dieciocho modos de entonar (distintivos y significativos) del español de La Habana. ANEXO I (García Riverón 1991a). Los modos de entonar responden a tipos de interacciones del *tono*, la *intensidad* y el *tiempo*, medidos en el laboratorio de fonética experimental. ANEXO II (García Riverón 1996).

Una vez definido el sistema de la entonación, se procedió a incluir su estudio en el *Atlas Lingüístico de Cuba* (ALCu)[4]. Para ello se preparó un cuestionario elaborado *ad hoc* y se formó a los miembros del equipo investigador tanto teórica como prácticamente con el fin de que pudiesen aprehender el objeto de estudio en cada punto poblado según el método de observación participante (García Riverón et al 1998). ANEXO III (García Riverón 1991a).

Objetivos del presente trabajo

1. Describir acústicamente la entonación de dos tipos de oraciones (declarativas neutras con sentido completo e interrogativas absolutas sin extensión) en habla de laboratorio de una informante cubana de Ciudad de La Habana, siguiendo el método al uso en el proyecto AMPER.[5]
2. Comprobar o refutar la validez de los datos, métodos y técnicas utilizados en estudios anteriores del español de Cuba a la luz de los métodos y técnicas del AMPER.

Premisas para el trabajo de la muestra del AMPER

Puesto que en publicaciones anteriores quedó demostrado que el tipo de oraciones estudiadas es similar en toda la Isla, se proyectó la indagación en un solo punto poblado: Ciudad de La Habana. Se grabó el Corpus 1 (experimental dirigido) por una informante de sexo femenino ente 20 y 50 años sin estudios superiores.

Análisis acústico.

Una vez grabadas las oraciones, usando el programa Goldwave se digitalizó el *corpus* con el fin de evitar cualquier tipo de ruido. Asimismo, se convirtió la señal en ficheros wav para proceder al análisis con el programa MatLab. Los datos se ampliaron con el programa Praat. Se segmentó el continuum sonoro en el oscilograma y se procedió a obtener los datos de los tres indicadores acústicos fundamentales: frecuencia del fundamental, tiempo e intensidad.

---

4 García Riverón, 1991 a y b; 1989b, García Riverón, y colab., 1988.

5 Se sigue a Dorta y Hernández, 2004.

RESULTADOS DE LOS ANÁLISIS

I. Estructuras enunciativas.

El tono. {ANEXO IV y V El saxofón se toca con pánico}

Como era de esperar en la mayoría de los casos ocurre un ascenso del tono fundamental en la primera vocal acentuada (saxofón, guitarra, cítara). Aunque en investigaciones anteriores del español cubano no fueron estudiadas de modo sistemático las variaciones del cuerpo del *entonema* en tanto no era objeto de la investigación (se analizaron oraciones con pretonema llano), el ascenso en la primera vocal acentuada de las declarativas neutrales emitidas como habla de laboratorio es característico de nuestra variante de lengua. A veces en las oxítonas (ante vocablo inacentuado) también es esta la posición de frecuencia fundamental máxima (*FoMáx)* de los enunciados declarativos neutrales, al tiempo que en las paroxítonas (guitarra) y proparoxítonas (cítara) se registró un ascenso mayor en la postónica (estos aspectos del comportamiento de Fo en el pretonema serán sometidos a un riguroso análisis estadístico). Después del primer ascenso, ocurre un marcado descenso en el se de *se toca* (palabra inacentuada, no funcional o no plena semánticamente) para comenzar a ascender en la siguiente tónica (to) hasta la subsiguiente átona (ca), donde comienza el descenso final característico de la declarativas, descenso que abarca el *centro de entonación* (última sílaba tónica) y el *segmento postónico* en estructuras con los tres patrones acentuales (paciencia, pánico y obsesión). En las proparoxítonas finales el descenso es más suave y en las oxítonas puede ser más abrupto.

Estos datos del comportamiento del tonema coinciden con los obtenidos en trabajos anteriores sobre el español de Cuba.

La duración

En todos los casos, los tiempos (*T)* iniciales de los segmentos en las declarativas son relativamente menores, sobre todo si los comparamos con los datos obtenidos en el final de las oraciones y con el comportamiento de *T* en las interrogativas. [ANEXO VI El saxofón se toca con pánico]. En los vocablos inacentuados o no plenos semánticamente (se, con) *T* disminuye y aumenta en la tónica (pá) de las proparoxítonas (pánico) para alcanzar (en la mayoría de las realizaciones) su punto más alto (o al menos un punto alto) en la(s) postónica(s) subsiguiente(s). También se vio en la investigación anterior que a veces se registra un tiempo mayor en la primera postónica que en la postónica final.

En las paroxítonas (paciencia), *T* se comporta de modo similar en el pretonema. Como norma, este indicador asciende hacia el final de la estructura, especialmente en el tonema, aunque en algunas realizaciones se pueda hallar una ligera disminución. [ANEXO VII El saxofón se toca con paciencia].

En cuanto a los vocablos de terminación oxítona (*obsesión*), la mayoría de las veces ocurre un aumento del tiempo que se dilata hasta el final del tonema. La tendencia general en todas las estructuras declarativas es el aumento del indicador T hacia el final del tonema.

La intensidad

La variable intensidad (*I*) se muestra de modo semejante en las declarativas e interrogativas: en general los dos tipos de oraciones comienzan con mayor *I*, que se va reduciendo paulatinamente a lo largo del segmento. [ANEXO VIII El saxofón se toca con pánico] En la declarativa proparoxítona en las combinaciones, *El saxofón se toca con pánico* y *La guitarra se toca con pánico*, la Intensidad máxima (Imax) se sitúa generalmente en la primera vocal tónica (en cítara aumentó en la vocal postónica), al tiempo que en la interrogativa la Imax se sitúa en el *centro de entonación*. Estos datos coinciden con los obtenidos en trabajos anteriores.

II. Estructuras interrogativas

El tono

En todos los registros [ANEXO IX El saxofón se toca con pánico] las oraciones interrogativas analizadas se pronuncian con niveles del tono por encima de las enunciativas. En el *pretonema* de las interrogativas, el tono asciende en la primera sílaba acentuada, independientemente del esquema acentual de que se trate. A diferencia de las declarativas, esta no es la posición de frecuencia fundamental máxima, que para las interrogativas se sitúo en el *centro de entonación* (pánico, paciencia, obsesión). Tampoco concuerda la curva de la interrogativa con la declarativa en el comportamiento del fundamental después de la primera sílaba tónica pues en la primera desde esta posición comienza a ascender el tono (el ascenso puede incluir las palabras inacentuadas no plenas semánticamente) o puede presentar un ligero descenso. Sin embargo, en la declarativa, después de la primera acentuada, ocurre un marcado descenso en se de se toca (palabra inacentuada y no plena semánticamente). En el *centro de entonación* el tono de las interrogativas asciende significativamente. Después del ascenso del tono en el *centro de entonación*, la curva termina con final descendente, que cierra la circunflexión característica de las interrogativas del español cubano. Los datos obtenidos por la metodología de AMPER coinciden plenamente con todos los obtenidos en trabajos anteriores.

La duración

En algunos de los casos, los tiempos iniciales de los segmentos vocálicos de las interrogativas son relativamente más amplios [ANEXO X La guitarra se toca con pánico] que en las enunciativas, lo que concuerda con los datos que hemos obtenido en otros trabajos. En ambos tipos de oraciones en los vocablos inacentuados o no plenos semánticamente (se, con) el tiempo disminuye y aumenta en la tónica. En las proparoxítonas (pánico), el *Tmax* se encuentra con frecuencia en la vocal tónica del tonema (pá) aunque puede alcanzar la mayoría de las veces su punto más alto en la postónica subsiguiente (muestras 15 y 18 *La guitarra se toca con pánico* y *la cítara se toca con pánico)*. En las paroxítonas (paciencia) el tiempo puede aumentar en el *centro de entonación* y continuar ampliándose hasta el final. En las oxítonas (obsesión), se registró un aumento sistemático del tiempo de los tonemas, aumento que también fue observado en trabajos anteriores.

Como es lógico pensar, el tonema oxítono es el que presenta mayor estabilidad temporal y tonal.

CONCLUSIONES DE LA INVESTIGACIÓN Y POSTULADOS PARA EL ESTUDIO DE LA ENTONACIÓN DEL ESPAÑOL CUBANO

Los datos obtenidos con la metodología AMPER coinciden con los resultados obtenidos anteriormente sobre el español cubano en cuanto a la uniformidad de muchos de ellos y a la posible variabilidad inherente a cualquier sistema complejo.

Se pudo comprobar que normalmente ocurre un ascenso del tono fundamental en la primera vocal acentuada del cuerpo tanto en las oraciones declarativas cuanto en las interrogativas. Estas últimas se pronuncian en niveles tonales superiores a los niveles en los cuales se realizan las declarativas. Asimismo, en las declarativas el centro de entonación y el segmento postónico es descendente, al tiempo que en las interrogativas es ascendente-descendente.

En general *T* tiende a aumentar hacia el final de las estructuras. La intensidad, se comporta de modo similar en las declarativas e interrogativas: los dos tipos de oraciones comienzan con mayor intensidad, que se va reduciendo paulatinamente a lo largo del segmento.

Aunque por razones de espacio no es posible abundar en este tema, podemos decir que los datos obtenidos con la metodología del proyecto AMPER confirman los rasgos distintivos utilizados hace algunos años para definir los entonemas **E-1** (enunciativo) y **E-3** (interrogativo) en el sistema de la entonación cubana.

Las variaciones de algunos indicadores acústicos registradas en las nuestras del AMPER se encontraron anteriormente. Las fluctuaciones de Fo y, sobre todo las fluctuaciones T, son normales y pueden tener una explicación en subsistemas de otros niveles estructurales (tipo de esquema acentual, longitud del segmento), pueden estar dadas por el grado de formalidad con que el informante haya leído el texto, etc. Esto es, pueden tener un fundamento ontológico. En otros casos, pueden ser fenómenos casuales, emergentes en un sistema complejo, sin una determinación ontológica. En consecuencia, estos hechos serán avalados por un análisis cualitativo con gran cantidad de informantes y con el control de las variables necesarias. Es de suponer que las fluctuaciones han de ser mayores cuando se opere con habla espontánea, por lo que deben crearse unas bases teórico-metodológicas sólidas que permitan la colación de datos.

Todo parece indicar que para Cuba se sustenta el postulado que glosa que el papel de las variaciones de tipo de estructura prosódica de las palabras en el *centro de entonación* y el *segmento postónico* (tonema) a los efectos de la definición del sistema de la entonación, no influyen en el *tipo de unidad comunicativa* de la entonación dentro del sistema *establecido*. Este punto se pudo comprobar en trabajos anteriores en los cuales han sido aplicadas encuestas semántico-pragmáticas de percepción tanto a estructuras plenas como a logatomos. Los datos del AMPER también serán sometidos a este análisis.

Dada la complejidad del sistema y el estado de la cuestión en Cuba, es necesario establecer un paso a paso analítico con muestras deslizantes que permitan fundamentar

etapas metodológicas y diferentes abordajes. Entonces, nosotros debemos ampliar nuestra metodología para proceder a un estudio estadístico y cualitativo del nivel estructural inferior, nivel desde el cual se investiga en el AMPER, denominado por nosotros en nuestras investigaciones Nivel menos 1 (N-1).

El resto de las investigaciones que venimos realizando en estos momentos (incluidos las indagaciones comparativo- tipológicas con otras lenguas o variante de lengua) se sitúan en los niveles cero (N) y uno (N 1) del sistema prosódico.

Para establecer las conexiones de un atlas de gran dominio (como es el AMPER) con nuestros trabajos se delimitó la dinámica a indagar trazando constricciones al *sistema complejo* de la entonación, puesto que consideramos que no es pertinente investigar varios niveles estructurales al unísono, dada la poca productividad investigadora que esto implica.

Lo que se acaba de explicar se puede resumir en un esquema de investigación deslizante como el que a continuación presentamos, esquema que nos permite sistematizar el proceso indagador y la colación de los datos para describir y explicar un sistema dinámico, complejo, de la entonación. En figura recapitulativa, el trabajo se viene realizando en los siguientes niveles.

NIVEL DEL ENTORNO N1	G L O B A L	ENTORNO Todo fenómeno del medio ambiente que influye en la dinámica indagada (Condicionantes socio-psicológicos y medios extralingüísticos de expresión LA HABANA.
DINÁMICA INDAGADA El-estar-siendo N	A N I D A D O	EL SISTEMA DE LA ENTONACIÓN Entendido el sistema de la entonación como sistema complejo con intercambio de materia, energía, información y sentido LA HABANA y otros espacios geográficos y sociales en los cuales no esté definido el sistema.
LOS COMPONENTES ESTRUCTURALES EN EL NIVEL ELEMENTAL N-1 Elementos estructurales inferiores.	L O C A L	NIVELES ESTRUCTURALES INFERIORES Las entidades menores distinguibles y susceptibles de involucrarse en la dinámica indagada, AMPER-Cuba.

Para finalizar, queremos hacer constar que este trabajo debe mucho a los profesores A. Veiga Rodríguez, M. González González, E. Martínez Celdrán y A. Mª. Fernández Planas, porque de un modo u otro nos han apoyado en nuestra labor. Asimismo, hemos logrado integrar nuestro quehacer a la investigación general gracias a la ayuda de la Secretaria de Estado de Universidades del Ministerio de Educación y Ciencia SAP-2003-0200.

REFERENCIAS BIBLIOGRÁFICAS

CONTINI, M (1992): "Vers une géoprosodie". *Actes du Nazioarteko Dialectologia Biltzarra Agiriak.* Publicación Real Academia de la Lengua Vasca (Bilbao, 1991), pp. 83-109.

DORTA, J. Y B. HERNÁNDEZ (2004). "Prosodia de las oraciones SVO declarativas e interrogativas en el español de Tenerife". *Estudios de Fonética Experimental*, XIII, pp. 225-273.

FERNÁNDEZ PLANAS, A.M. (2005): "Aspectos generales acerca del proyecto "AMPER" en España". *Estudios de Fonética Experimental*, XIV, pp. 13-27.

GARCÍA RIVERÓN, R. (1991a): "La entonación en el Atlas Lingüístico de Cuba. *Español Actual*, 55, 103- 112.

GARCÍA RIVERÓN, R. (1991b): "El Atlas Lingüístico de Cuba, *Lingüística española actual*, XII, 2,199- 221.

GARCÍA RIVERÓN, R. (1996-1998): *Aspectos de la entonación hispánica.* 3 volúmenes. Cáceres: Universidad de Extremadura.

GARCÍA RIVERÓN, R. (2005) "El estudio de la entonación". *Moenia* (en prensa).

GARCÍA RIVERÓN, R. et. al. (1988) *El Atlas Lingüístico de Cuba. Cuestionario.* La Habana: Academia, Academia de Ciencias de Cuba.

ANEXO I

A partir de estas consideraciones se deslindaron en el sistema entonativo de Ciudad de La Habana 18 entonemas que pasamos a relacionar:

Entonemas:

Entonema 1	No tiene hambre. (1)
Entonema 2	¿En qué hospital? (2)
Entonema 3	¿Vas a la playa? (3)
Entonema 4	¿Y hoy? (4)
Entonema 5	Si vienen... (5)
Entonema 6	¡Ese café me hace un daño! (6)
Entonema 7	José (7)

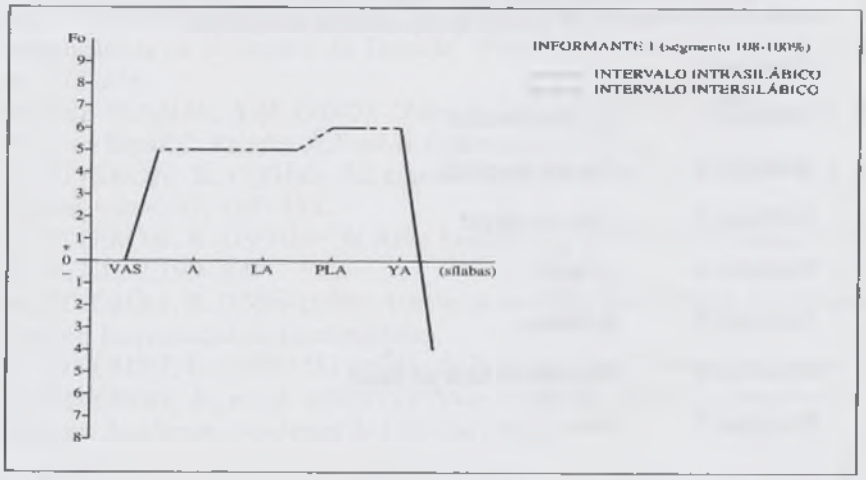
Variante de entonema:

Variante a del entonema 1	Sin engañarme (1a)
Variante b del entonema 1	Claro (1b)
Variente c del entonema 1	¿Y a qué tú le tienes miedo, Lazarito? (1c)
Variante a del entonema 2	¿Y qué pasa? (2a)
Variante a del entonema 3	¿Ahora? (3a)
Variente b del entonema 3	Te le echo hielo (3b)
Variente a del entonema 4	¿El maitre es el compañero que se ocupa...? (4a)
Variante a del entonema 5	Todo el mundo quiere los niños gordos... (5a)
Variante b del entonema 5	Como tú no fumas (5b)
Variante a del entonema 7	Aurora (7a)

Realizaciones

Realización a del entonema 6	Tengo unas ganas de tomarme un refresco frío. (6a)

ANEXO II



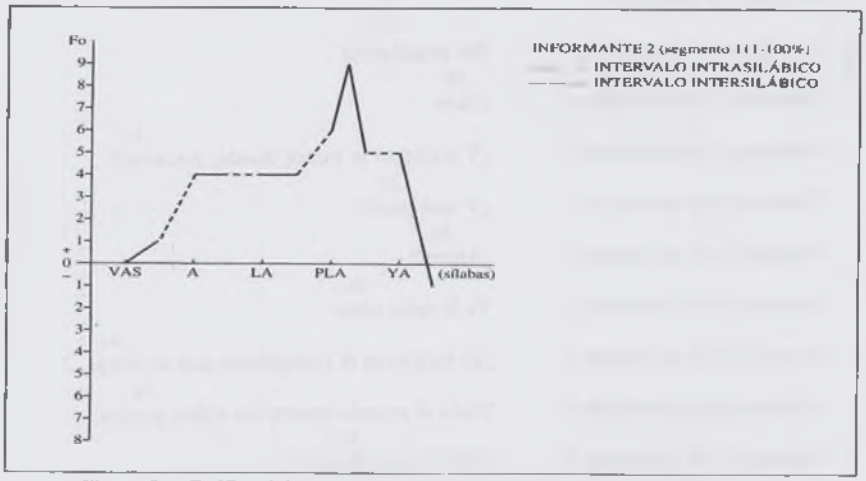


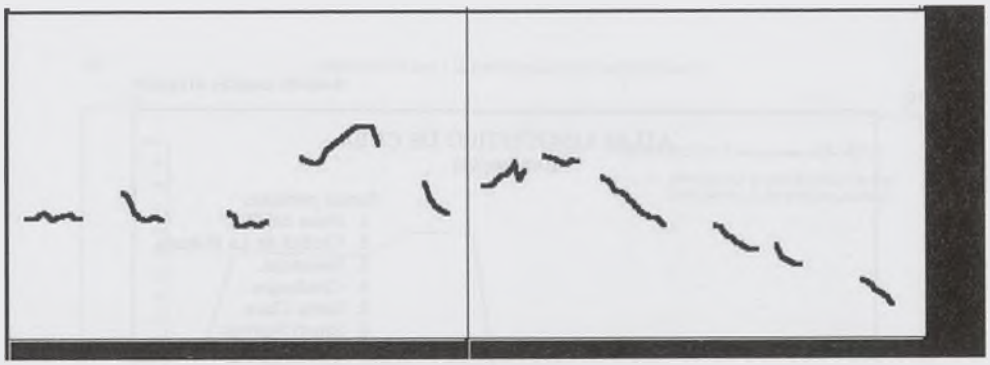
Figura 8. Gráfico del comportamiento del Fo (en semitonos). Entonema 3.

ANEXO III

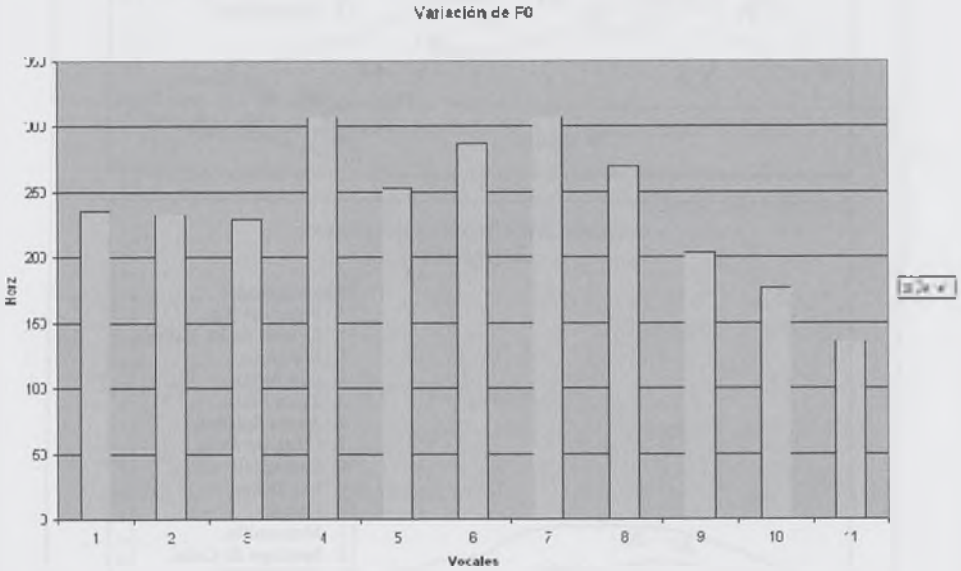

ATLAS LINGÜÍSTICO DE CUBA
Entonación
Puntos poblados:
1. Pinar del Río.
2. Ciudad de La Habana.
3. Matanzas.
4. Cienfuegos.
5. Santa Clara.
6. Sancti Spíritus.
7. Ciego de Ávila.
8. Camagüey.
9. Las Tunas.
10. Holguín.
11. Manzanillo.
12. Santiago de Cuba.
13. Guantánamo.
VE-1b


ATLAS LINGÜÍSTICO DE CUBA
Entonación
Puntos poblados:
1. Pinar del Río.
2. Ciudad de La Habana.
3. Matanzas.
4. Cienfuegos.
5. Santa Clara.
6. Sancti Spíritus.
7. Ciego de Ávila.
8. Camagüey.
9. Las Tunas.
10. Holguín.
11. Manzanillo.
12. Santiago de Cuba.
13. Guantánamo.
VE-1c

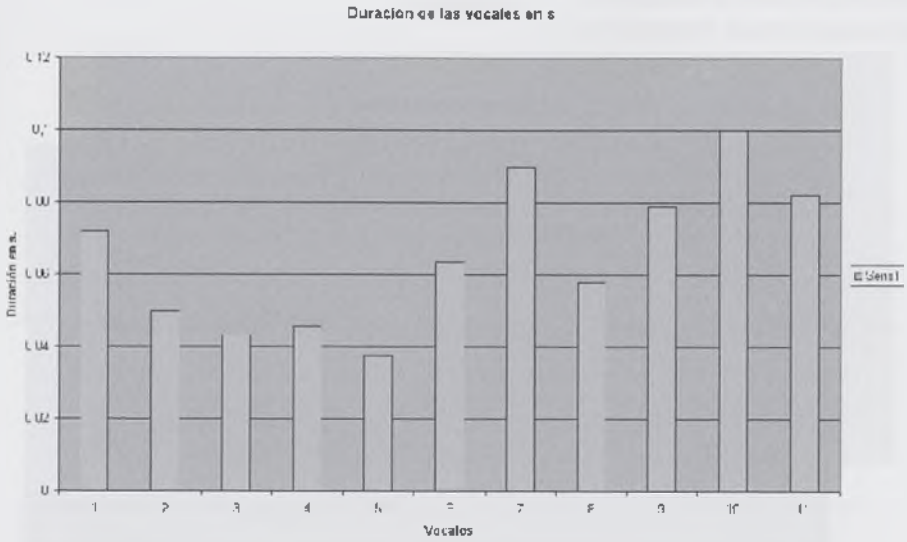
ANEXO IV



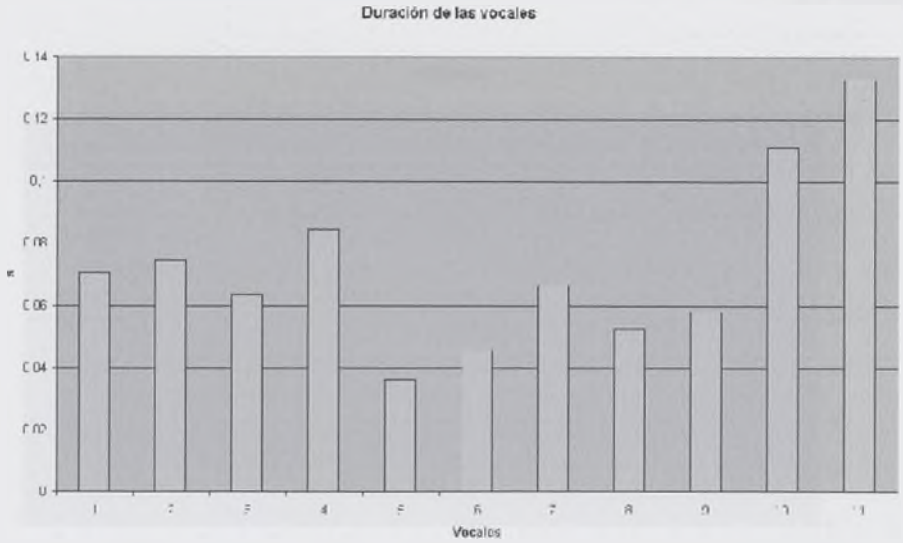
ANEXO V



ANEXO VI

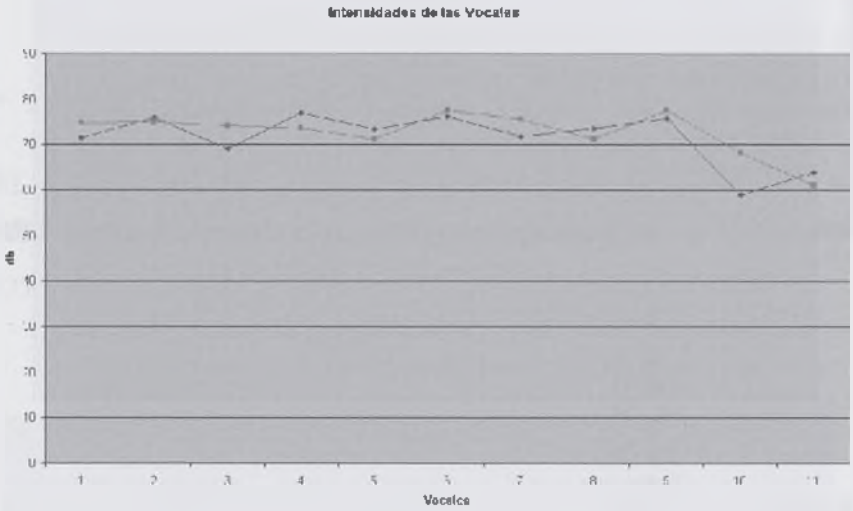


ANEXO VII

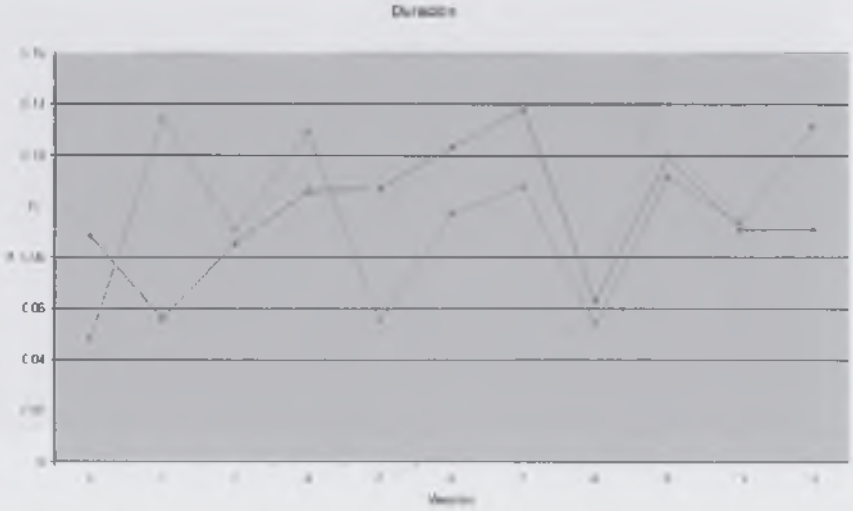


ANEXO VIIIANEXOS VIII, IX Y X

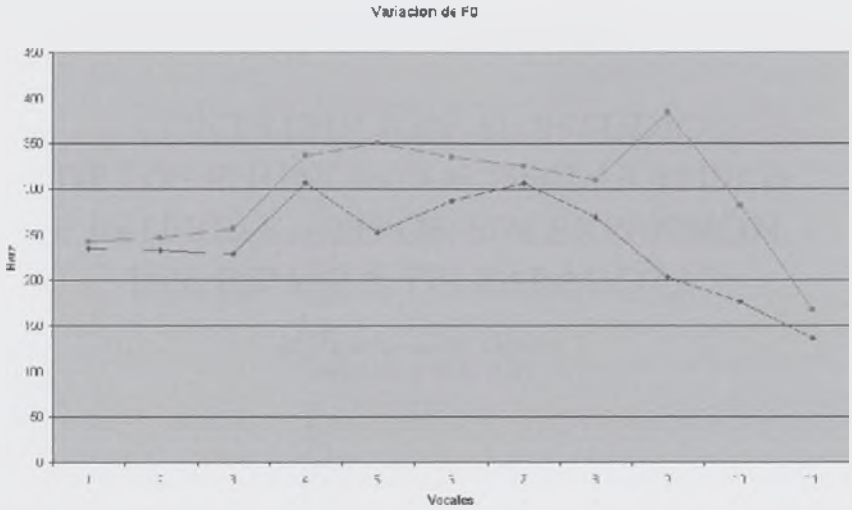
Curvas en gris claro: Interrogativas
Curvas en gris oscuro: Enunciativas.



ANEXO IX.



ANEXO X



CONTRIBUCIÓN AL ESTUDIO DE LOS ENUNCIADOS DECLARATIVOS E INTERROGATIVOS, SIN EXPANSIÓN, DEL ESPAÑOL DE ZARAGOZA*

Mª Pilar González Olivera
Estudiante de Tercer Ciclo

1. LA ENTONACIÓN DE ARAGÓN

La entonación aragonesa, tanto del español de Aragón como de las distintas hablas de la región[1], tan apenas ha sido investigada experimentalmente, a pesar de que se ha insistido de manera reiterada en sus peculiaridades, sobre todo en las referidas al español.

Ya en 1872, un periodista italiano, en su diario de viaje, al llegar a Zaragoza señaló "...la cadencia particular con la que hablaban los cocheros, los mozos y los jóvenes que se disputaban mi maleta." Incluso dio unas primeras impresiones de esta *cadencia particular*: «todo castellano sabe imitar ese acento y en ocasiones lo ridiculiza, por lo que tiene de tosco y monótono».[2]

En 1917, el fonetista tradicional por excelencia del español, Tomás Navarro Tomás, observaba que "...en el habla popular de algunas regiones españolas la vocal final es aún más larga que en castellano, destacándose principalmente en aragonés."[3]

Unos años más tarde, en 1935, daba una serie de características de la pronunciación de los aragoneses:

---

* Este estudio ha sido posible gracias al Proyecto de Investigación IBERCAJA (MODALIDAD B), referencia IBE2004B-HUM-02.

1 A lo largo de la geografía aragonesa, además del castellano o español, se hablan diversas variedades procedentes del antiguo aragonés y del catalán. La bibliografía sobre las distintas variedades que se hablan en Aragón es muy extensa y puede verse un estado de la cuestión en ENGUITA, José María (editor), *Jornadas de Filología Aragonesa*.

2 DE AMICIS, Edmundo, *España. Diario de un viaje de un turista escritor*, p. 71.

3 NAVARRO TOMAS, Tomás, "Cantidad de las vocales inacentuadas", p. 384.

> "En los pueblos de Aragón, uniformados lingüísticamente en su mayor parte desde hace siglos por la influencia del castellano, el acento con que éste se pronuncia es, probablemente, el mismo con que dichos pueblos hablaron su antiguo dialecto aragonés antes de adoptar la lengua de Castilla. El ansotano y el cheso, restos del dialecto aragonés conservados en escondidos valles del Pirineo, coinciden esencialmente por su carácter prosódico con el acento peculiar del resto de la región. El rasgo más característico de este acento consiste en el tono relativamente alto con que de ordinario terminan las frases, aunque no sean interrogativas. En circunstancias análogas, la inflexión final de una aseveración corriente termina en aragonés en una nota seis u ocho semitonos más alta que en castellano."[4]

Tomás Buesa Oliver, nos recuerda, además de las palabras anteriormente citadas, las siguientes de Navarro Tomás: "El aragonés habla en general con articulación clara, con tono relativamente lento y grave, y con reforzada intensidad que imprime cierta rudeza al franco y noble afecto de la expresión."[5] Y ofrece algunas características de la pronunciación en Aragón, a partir de su experiencia como encuestador del *Atlas Lingüístico Etnográfico de Aragón, Navarra y La Rioja (ALEANR)*.[6] Estas características son las siguientes:

a) Repugnancia hacia las palabras esdrújulas convirtiéndolas en llanas.
b) La articulación de los sonidos si no es igual, es muy parecida a la del español estándar.
c) El refuerzo del tono en la inflexión final de la curva entonativa da lugar al alargamiento de la vocal última, cosa que en el español estándar no se da.
d) Predominio de la nitidez en el timbre de las vocales.

Actualmente, en las últimas décadas, se ha estado insistiendo no sólo en las supuestas peculiaridades de la entonación en Aragón, sino también en la necesidad de abordar su estudio experimental.[7] En esta línea contamos ya con algunos intentos.

---

4 NAVARRO TOMÁS, Tomás, *El acento castellano*, p. 44.

5 BUESA OLIVER, Tomás, "Particularidades del español hablado en Aragón", pp. 119-120.

6 Dichas encuestas fueron realizados entre 1963 y 1968 por todo Aragón, Navarra y La Rioja, además de por algunos puntos colindantes de las regiones vecinas. Junto a él participaron Manuel Alvar (director del Atlas) y Antonio Llorente.

7 Pueden verse las consideraciones de LÁZARO CARRETER, Fernando, *El habla de Magallón. Notas para el estudio del aragonés* vulgar, p. 4, MARTIN ZORRAQUINO, Mª Antonia, "Elementos para una sociolingüística del habla de Zaragoza", pp. 135-136, y "Estudio sociolingüístico del habla de Zaragoza: problemas y primeros resultados", p. 19; ENGUITA, José María "Las hablas de Zaragoza", p. 204, y "Estado actual de los estudios del español de Aragón", p. 337, MARTÍN ZORRAQUINO, Mª Antonia y José María ENGUITA, *Las hablas de Aragón*, pp. 49-50, o PORROCHE, Margarita "Estudio de una elaboración humorística del español hablado en Aragón", p. 214. Más datos en CASTAÑER MARTÍN, Rosa Mª *et alii*, "Aproximación al estudio de la entonación aragonesa", en prensa.

Vilar Pacheco[8] compara la entonación del español hablado en las comarcas turolenses de la Sierra de Albarracín (zona no marcada, es decir, más parecida al español estándar) y de Calamocha (zona marcada). Se centra en enunciados aseverativos o declarativos obtenidos a través de conversaciones dirigidas en las que se ha procurado la mayor espontaneidad posible.

Los resultados que se reflejan son los siguientes:

a) En cuanto al tonema final, los enunciados de la Sierra de Albarracín se ajustan bastante al patrón entonativo del español estándar; mientras que en la zona de Calamocha, la inflexión final apenas es perceptible y se mantiene casi a la misma altura desde el inicio.
b) En ambas zonas, aunque fundamentalmente en la Sierra de Albarracín, encontramos un cuerpo entonativo abrupto, contrastando con lo dicho por Navarro Tomás sobre la forma homogénea del cuerpo entonativo del español. Aparece un marcado movimiento ascendente del fundamental en las sílabas tónicas.
c) El tono parece más alto en la comarca de Calamocha (125 -150 Hz.), sin embargo en la Sierra de Albarracín los valores de *Fo* se sitúan en la media del español estándar (100 Hz.).

A modo de conclusión, Vilar dice que la entonación de Calamocha parece que se corresponda con la típicamente aragonesa, con una inflexión final muy débil.[9]

La metodología AMPER es aplicada en el trabajo de Contini *et alii*, "Analyse comparée de l'intonation en castillan et en aragonais"[10], en el que se compara la entonación del español de una informante madrileña de 30 años de edad, con el aragonés de un hombre oriundo del valle de Bielsa, de 60 años de edad.

Se observa que tanto en aragonés como en español, para las frases declarativas, el esquema entonativo es ascendente en el sintagma nominal y descendente en el verbal. El reparto de la curva melódica se encuentra alrededor del nivel medio de Fo. La cumbre se sitúa al final del sintagma nominal. Asimismo, hay un alargamiento de las vocales finales.

Aparecen también algunas diferencias, entre ellas:

a) En aragonés, con las palabras proparoxítonas y paroxítonas, el esquema en la primera palabra prosódica es ascendente, con ruptura entre la sílaba pretónica y tónica; mientras que en español, en las palabras paroxítonas y oxítonas se constata una subida mayor en la vocal final.
b) Con respecto a la vocal final de frase, sube ligeramente en las palabras proparoxítonas y paroxítonas en aragonés; sin embargo, en español la subida se da en las palabras oxítonas.

---

8 VILAR PACHECO, José Mª, "Aproximación a la entonación del español en Teruel (Albarracín frente a Calamocha".

9 Conclusión y resultados que Vilar expone de manera provisional a la espera de un estudio más amplio.

10 En este artículo se hace referencia siempre a *aragonés* para referirse a la variedad específica del valle de Bielsa, que es el *belsetán* o *belsetano*. Sobre los elementos suprasegmentales de esta habla, trata la tesis doctoral de A. Rhardisse, *Accent et intonation du parler de Bielsa (Aragon –Espagne)*, defendida en la Université Stendhal de Grenoble.

c) La duración de la sílaba final es mucho mayor en español que en aragonés.

Por otro lado se observa que las curvas interrogativas se diferencian claramente de las afirmativas.

En aragonés, la diferencia con la curva declarativa se encuentra en la sílaba postónica del sintagma nominal. El Fo asciende hasta el valor máximo de la frase y al final, en la sílaba pretónica del sustantivo del sintagma verbal, se produce un descenso abrupto de sus valores.

En español, hay un descenso de Fo al final del sintagma nominal; el Fo desciende, incluso por debajo del tono medio de la curva, tras la sílaba postónica de dicho sintagma. Este descenso termina al final del sintagma verbal, en una subida fuerte que alcanza el valor máximo de la curva.

Con respecto a la duración, tanto en aragonés como en español se produce un alargamiento importante de la vocal final de la curva interrogativa frente a la declarativa.

En aragonés, la sílaba más larga es la acentuada en caso de las palabras oxítonas. Para las palabras proparoxítonas y paroxítonas la vocal más larga es la final, sea acentuada o no. En cambio en español se considera la duración como un medio para resaltar las estructuras acentuales.

En cualquier caso, se contrapone, más en aragonés que en español, una duración mayor al final del sintagma nominal en las curvas declarativas, con una duración mayor al final de las frases interrogativas.

El grupo de investigación del que formo parte presentó en Barcelona una información incipiente, con unos primeros datos de nuestra aportación al proyecto AMPER.[11] Dábamos cuenta de la bibliografía existente y analizábamos como primera muestra la entonación de una hablante de Zaragoza, aunque el análisis no se efectuó con los medios técnicos adecuados.

El corpus también era muy limitado, tan sólo tres frases sin expansión con la estructura S+V+O, siendo los acentos inicial e intermedio siempre llano o paroxítono, en las modalidades declarativa e interrogativa.

Las curvas entonativas analizadas no son tan diferentes en el español de Zaragoza con respecto al español estándar. Sin embargo, sí se encuentra una mayor elevación del tono al final del sintagma nominal sujeto. En la inflexión final de las declarativas, aparece un descenso que contradice las afirmaciones tradicionales. En las interrogativas, el descenso del cuerpo de la unidad se produce siempre hasta la penúltima sílaba de la frase, ascendiendo bruscamente en la última.

Con respecto a la duración, sorprende el alargamiento de la vocal inicial de frase tanto declarativa como interrogativa, así como el incremento de duración que experimentan las sílabas finales, tanto en el sintagma nominal como en el verbal. [12]

---

11 CASTAÑER MARTÍN, Rosa Mª *et alii*, "Aproximación al estudio de la entonación aragonesa", en prensa.

12 Los resultados de este estudio, como allí reconocemos, hay que tomarlos con cautela por los motivosseñalados, lo exiguo del corpus y la limitación técnica.

2. METODOLOGÍA

2.1 Características del informante y corpus

Respecto al informante, se ha escogido uno que cumple con los requisitos sociolingüísticos establecidos por el proyecto AMPER-ESPAÑA.

El informante es una mujer aragonesa, nacida y residente en Zaragoza capital, sin estudios superiores, de 40 años de edad.

Aunque en el proyecto AMPER encontramos distintos tipos de corpus, este estudio se ha centrado en frases controladas obtenidas mediante lectura. Son frases de modalidad declarativa e interrogativa, de 11 sílabas cada una, cuya estructura sintáctica es SVO sin expansión, coincidiendo con tres acentos tonales. El acento tonal intermedio es siempre paroxítono, mientras que los acentos inicial y final poseen diferente tipología acentual. Las frases son las siguientes:

El saxofón se toca con obsesión (*kwk*)
El saxofón se toca con pánico (*kwp*)
El saxofón se toca con paciencia (*kwt*)
La cítara se toca con obsesión (*pwk*)
La cítara se toca con pánico (*pwp*)
La cítara se toca con paciencia (*pwt*)
La guitarra se toca con obsesión (*twk*)
La guitarra se toca con pánico (*twp*)
La guitarra se toca con paciencia (*twt*)

En definitiva, el corpus analizado consta de 18 frases, 9 declarativas y otras tantas interrogativas, repetidas cada una de ellas tres veces, por lo que el total de frases analizadas es de 54.

2.2 Análisis acústico

El corpus se ha obtenido a través de una grabadora MARANTZ, modelo CP430.

Más tarde, se ha procedido a la digitalización, optimización y extracción de ruidos mediante el programa *Goldwave* versión 4.25, convirtiendo las frases en ficheros *wav* para su ulterior análisis con el programa *MatLab* versión 4.0.

A partir del oscilograma de cada frase que nos da el programa, segmentamos las vocales, tomando tres valores tonales de cada una de ellas, así como la duración y la intensidad. Se estilizan las curvas y se hace una media de los valores de las tres repeticiones de cada frase. Por último, se obtienen los gráficos correspondientes al análisis efectuado.

3. RESULTADOS

3.1 Análisis de F0

En primer lugar, presentamos las curvas obtenidas a través de la media de los valores de F0 dependiendo de la estructura prosódica de las palabras y de su posición en la oración, es decir, según sean las palabras oxítonas, paroxítonas o proparoxítonas en el sintagma nominal o en el sintagma preposicional.

En los siguientes gráficos, los números del eje X corresponden a las vocales de la oración. Se ha tomado como valor de estudio el valor final de cada vocal.

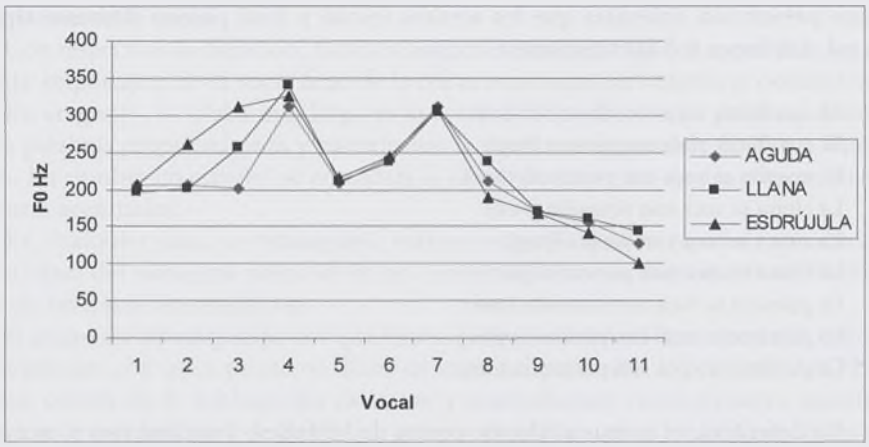


Fig. 1. *Curvas de las oraciones declarativas*

La figura 1 muestra una inflexión ascendente de la voz en el sintagma nominal desde de la vocal pretónica. En otras palabras, la curva asciende fuertemente a partir de la vocal 3 para la oxítona, de la vocal 2 en la paroxítona y de la vocal 1 en la proparoxítona. La última vocal del sintagma nominal es la más aguda de toda la frase, con valores de más de 300 Hz (311 Hz oxítona, 340 Hz paroxítona y 326 Hz proparoxítona) produciéndose así el primer pico tonal. Hay una diferencia de siete, nueve y ocho semitonos respectivamente.

El tono desciende bruscamente al principio del segundo acento tonal para volver a ascender a lo largo del verbo, teniendo así el segundo pico tonal. Este movimiento contradice la afirmación de Navarro Tomás, para quien "...durante la pronunciación del cuerpo de la unidad, la entonación española en las frases enunciativas mantiene la altura de la voz en un nivel relativamente uniforme",[13] y, sin embargo, coincide con lo que

13 NAVARRO TOMÁS, Tomás *Manual de entonación española*, p. 67.

Canellada y Madsen proponen para el núcleo informativo que "...va colocado inmediatamente después de la gran subida tonal, o punto de anticadencia."[14]

El sintagma preposicional es un continuo descenso, siendo más pronunciado en la proparoxítona. Los valores de la vocal final son los más bajos de la oración (128 Hz oxítona, 144 Hz paroxítona y 100 Hz proparoxítona). Para Navarro Tomás, el tonema de cadencia final de la frase baja unos ocho semitonos aproximadamente, cosa que ocurre en la oxítona y paroxítona de nuestros resultados, pero no así en la proparoxítona, cuyo descenso llega a once semitonos.

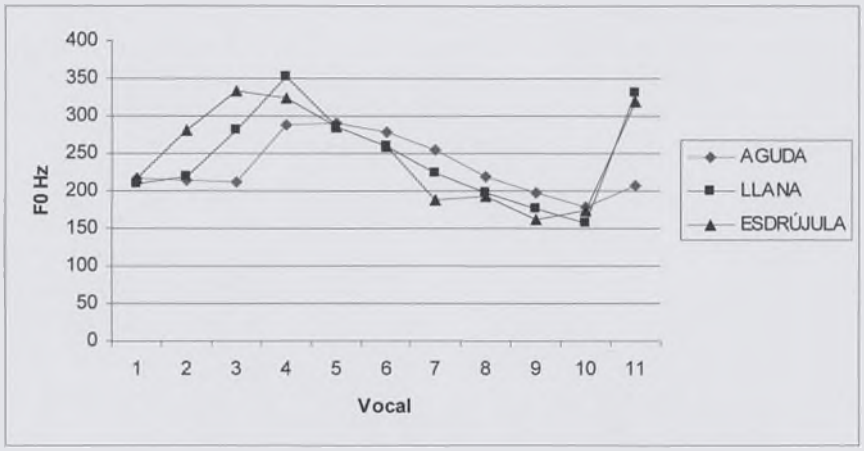


Fig. 2. *Curvas de las oraciones interrogativas*

En la figura 2 encontramos el mismo movimiento que en la figura 1, pero con valores más altos que en las curvas declarativas, a excepción de la última vocal del sintagma nominal. En la curva de la proparoxítona, ya ha comenzado el descenso en esta vocal 4. A continuación el tono va descendiendo hasta la penúltima vocal de la frase que posee los valores prácticamente más bajos de la curva, para ascender fuertemente en la vocal final. Este ascenso no es tan brusco en las oxítonas, ni siquiera llega a la frecuencia media de la hablante (220 Hz). En las proparoxítonas el ascenso empieza con la vocal tónica.

Coinciden estos resultados con lo propuesto por Navarro Tomás: "Las sílabas débiles iniciales, [...] se dicen en la pregunta con movimiento ascendente."[15] Continúa diciendo: "El cuerpo de la unidad interrogativa en la entonación española es generalmente descendente."[16] Sigue con la descripción del final de la frase interrogativa absoluta: "El descenso comprende hasta la última sílaba fuerte o hasta la que inmediatamente le prece-

14 CANELLADA, Mª Josefa y John Kuhlmann MADSEN, *Pronunciación del español*, p. 132.
15 NAVARRO TOMÁS, Tomás *Manual de entonación española*, pp. 135-136.
16 *Ibíd.*, p. 136.

de. El final de la frase interrogativa absoluta es siempre ascendente."[17] Este descenso suele ser se siete u ocho semitonos, coincidiendo sólo en la oxítona de nuestros resultados, ya que la paroxítona y proparoxítona desciende catorce y diez semitonos respectivamente.

Y coinciden también con Canellada y Madsen, para quienes "...desde las primeras sílabas, toda la frase interrogativa se encuentra en un nivel tonal más alto que la aseverativa."[18] Y "...todo el cuerpo de la frase es descendente hasta la última tónica, que es la que ocupa el nivel más bajo. Inmediatamente después, a lo largo de la última átona, se da la gran elevación final."[19]

3.2 Análisis de la duración

Consideremos el análisis de la duración vocálica en las figuras 3 y 4, de las frases declarativas e interrogativas, respectivamente. Al igual que en el análisis del Fo, los valores que presentamos son el resultado de la media de los valores de duración dependiendo de la estructura prosódica de las palabras y de su posición en la oración.

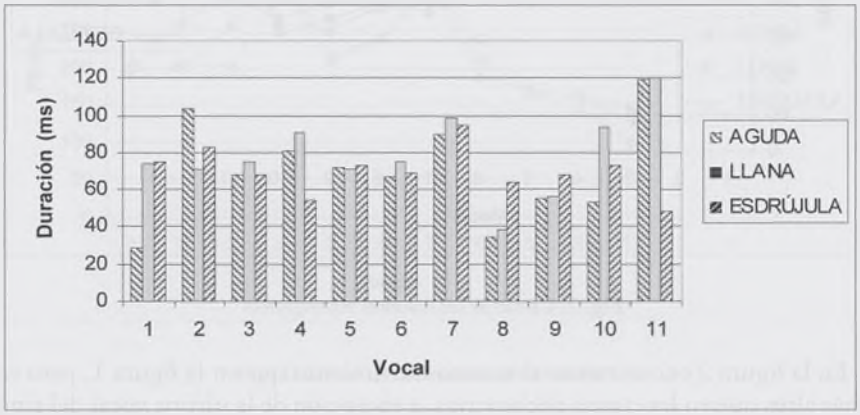


Fig.3. *Duración de las frases declarativas*

Llama la atención la diferencia de duración que encontramos en la primera vocal, la del artículo determinado, dependiendo de si es vocal inicial o vocal final; en otras palabras, la vocal del artículo *el*, de la frase oxítona, tiene una duración bastante menor que la vocal del artículo *la* de las frases paroxítonas y proparoxítonas.

En el primer acento tonal, las oscilaciones que presentan los valores de duración son distintas, según sea el tipo de palabra. En la oxítona tenemos una oscilación cóncava,

17 *Ibíd.*, p. 141.
18 CANELLADA, Mª Josefa y John Kuhlmann MADSEN, *Pronunciación del español*, pp. 148.
19 *Ibíd.*, p. 149.

siendo la vocal más larga la número 2, y la más corta la vocal pretónica. En la paroxítona la oscilación es ascendente, siendo las vocales pretónica y postónica la más corta y larga respectivamente. Sin embargo, en la palabra proparoxítona ocurre lo contrario, la oscilación es descendente. Esto no concuerda con lo dicho por Martínez Celdrán y Fernández Planas, que constatan que en el pretonema, "la duración mayor se encuentra siempre en la sílaba tónica."[20]

El verbo posee una duración similar en todas las frases.

En el último acento tonal, el movimiento es ascendente y destaca la vocal final que es la vocal más larga de la frase. Excepto en la proparoxítona, cuya vocal final es la más corta de este último acento.

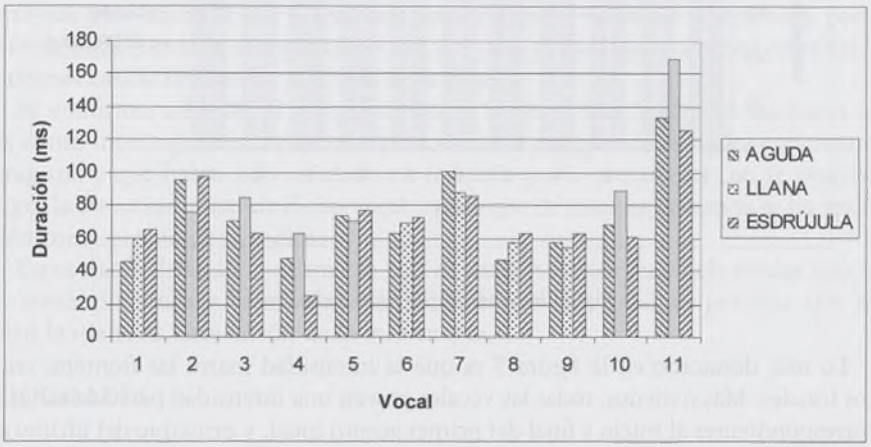


Fig. 4. *Duración de las frases interrogativas*

En la figura 4 no se destaca la duración del artículo determinado del sintagma nominal, como ocurre en las declarativas. Sin embargo el sustantivo es más homogéneo. Tanto en oxítonas como en proparoxítonas, el movimiento es descendente. En la paroxítona tenemos un movimiento convexo. Podemos decir, que aquí la mayor duración corresponde a la vocal tónica, salvo en la oxítona.

El verbo, al igual que en las frases declarativas, es similar en todas las frases.

Vuelve a destacar en la inflexión final la última vocal, la más larga de toda la frase. El movimiento de la duración del sintagma preposicional es ascendente, en la proparoxítona casi es horizontal, alargando bruscamente la última vocal.

20 MARTÍNEZ CELDRÁN, Eugenio y Ana María FERNÁNDEZ PLANAS, "El tono fundamental y la duración: dos aspectos de la taxonomía prosódica en dos modalidades de habla (enunciativa e interrogativa)", p. 188.

3.3 Análisis de la intensidad

Las figuras 5 y 6 muestran los gráficos de barras correspondientes a la intensidad. Como en los análisis anteriores, los valores que presentamos son el resultado de la media de los valores de intensidad dependiendo de la estructura prosódica de las palabras y de su posición en la oración.

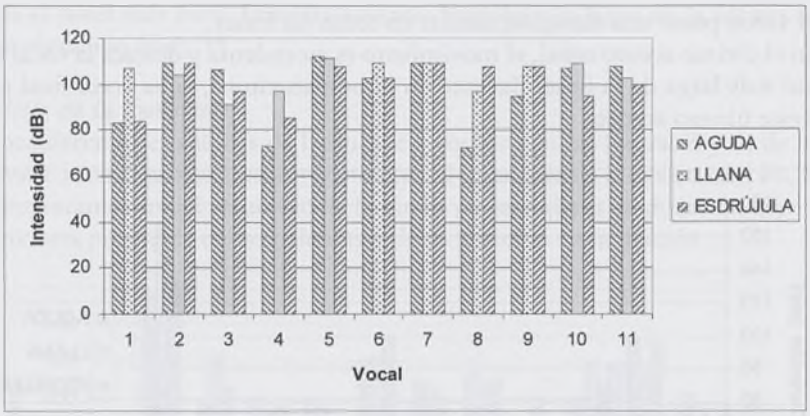


Fig. 5. *Intensidad de las frases declarativas*

Lo más destacado en la figura 5 es que la intensidad marca las fronteras entre los acentos tonales. Más o menos, todas las vocales poseen una intensidad parecida salvo la 1, 4 y 8, correspondientes al inicio y final del primer acento tonal, y principio del último acento tonal, respectivamente; sobre todo se marca en los ejemplos con acentuación oxítona.

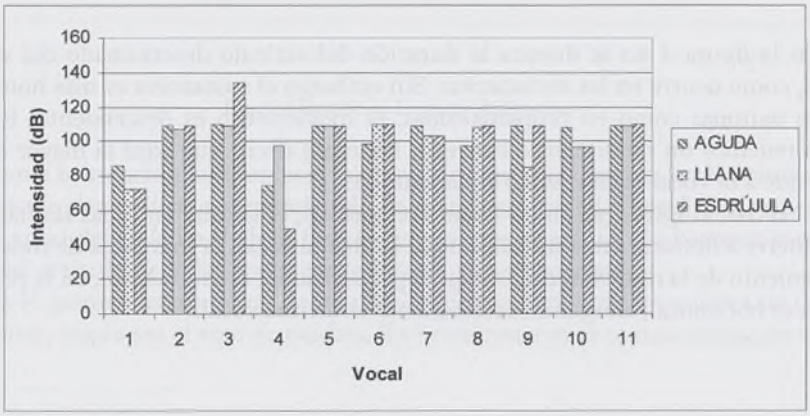


Fig. 6. *Intensidad de las frases interrogativas*

En las frases interrogativas, figura 6, se puede decir que la intensidad apenas varía. La frontera entre el segundo acento tonal y el último no se dibuja. No obstante, encontramos diferencia de valores en el sintagma nominal, a pesar de que el movimiento que hace el gráfico en la oxítona, paroxítona y proparoxítona es el mismo. Destaca la intensidad de la proparoxítona en este primer acento tonal, donde la intensidad va aumentando, sobrepasando los máximos valores con diferencia en la postónica, y cayendo bruscamente la intensidad de la vocal contigua.

4. CONCLUSIÓN

Podemos decir que la prosodia de las frases declarativas e interrogativas analizadas presenta una gran afinidad con lo señalado para el español estándar. Ahora bien, podemos destacar de estas curvas la abrupta elevación del tono al final del sintagma nominal, donde se alcanza casi la frecuencia más alta de la frase.

Se confirma, además, el alargamiento de la vocal final, tanto en las frases declarativas como interrogativas, rasgo que caracteriza al español de Aragón, concretamente de Zaragoza, y que había sido señalado en la bibliografía precedente; no se percibe, sin embargo, la elevación tonal de dicha vocal, rasgo que de manera reiterada se ha apuntado también como propio de la región.

Estos datos deberán confirmarse con el análisis de las frases obtenidas con la encuesta semiespontánea y la conversación libre en las que creemos percibir con mayor fidelidad la marcada entonación de la informante.

BIBLIOGRAFÍA

BUESA OLIVER, Tomás, "Particularidades del español hablado en Aragón", *Jornadas de Filología Aragonesa en el L Aniversario del AFA*, José Mª Enguita (ed.), Zaragoza, Institución «Fernando el Católico», 1994, pp. 113-138.

CANALLADA, Mª Josefa y John K. MADSEN, *Pronunciación del español*, Madrid: Editorial Castalia, 1987.

CASTAÑER MARTÍN, Rosa Mª *et alii*, "Aproximación al estudio de la entonación aragonesa", en prensa.

CONTINI, Michel, Claudine FRANCHON y Amina RHARDISSE, "Analyse comparée de l'intonation en castillan et en aragonais", *Atti XXI Congresso Internazionale de Lingüística e Filología Romanza*, Tubinga: Max Niemeyer Verlag, vol. V, 1998, pp. 143-157.

DE AMICIS, Edmundo, *España. Diario de un viaje de un turista escritor*, Madrid: Cátedra, 2000.

ENGUITA, José Mª, "Las hablas de Zaragoza", *Actas del I Curso de Geografía Lingüística en Aragón*, Zaragoza, Institución «Fernando el Católico», 1991, pp. 201-239.

ENGUITA, José Mª, "Estado actual de los estudios sobre el español de Aragón", *Jornadas de Filología Aragonesa en el L Aniversario del AFA*, Zaragoza: Institución «Fernando el Católico», II, 1999, pp. 319-366.

ENGUITA, José Mª (editor), *Jornadas de Filología Aragonesa en el L Aniversario del AFA*, Zaragoza: Institución «Fernando el Católico», II, 1999.

LÁZARO CARRETER, Fernando, *El habla de Magallón. Notas para el estudio del aragonés vulgar*, Zaragoza: Institución «Fernando el Católico», 1945.

MARTÍN ZORRAQUINO, Mª Antonia, "Elementos para una sociolingüística del habla de Zaragoza", *Turia*, 4-5, 1986, pp.121-139.

MARTÍN ZORRAQUINO, Mª Antonia, "Estudio sociolingüístico del habla de Zaragoza: problemas y primeros resultados", *Actas del I Congreso de Lingüistas Aragoneses*, Zaragoza: Institución «Fernando el Católico»,1991, pp. 169-200.

MARTÍN ZORRAQUINO, Mª Antonia y ENGUITA, José Mª, *Las lenguas de Aragón*, Zaragoza: CAI, 2000.

MARTÍNEZ CELDRÁN, Eugenio y Ana Mª FERNÁNDEZ PLANAS, "El tono fundamental y la duración: dos aspectos de la taxonomía prosódica en dos modalidades de habla (enunciativa e interrogativa) del español", *EFE*, XII, 2003, pp.165-200.

NAVARRO TOMÁS, Tomás, "Cantidad de las vocales inacentuadas", *RFE*, IV, 1917, pp.371-388.

NAVARRO TOMÁS, Tomás, *El acento castellano. Discurso leído por el autor en el acto de su recepción académica el día 19 de mayo de 1935*, Madrid: Tipografía de Archivos, Olózaga, I, 1935.

NAVARRO TOMÁS, Tomás, *Manual de entonación española*, Madrid, Guadarrama, 3ª edición, México, Colección Málaga, 1966.

PORROCHE, Margarita, "Estudio de una elaboración humorística del español hablado en Aragón", *Jornadas sobre la variación lingüística en Aragón a través de los textos*, Zaragoza: Institución «Fernando el Católico», 2004, pp. 205-225.

VILAR PACHECO, José Mª, "Aproximación a la entonación del español en Teruel (Albarracín frente a Calamocha)", *Xiloca*, 25, 2001, pp.177-212.

SCOTTISH ENGLISH AND FRENCH AFFECTIVE PROSODY: A CROSS-LINGUISTIC STUDY OF BILINGUAL CHILDREN

Ioulia Grichkovtsova
Speech and Hearing Sciences, QUMC, Edinburgh, UK

ABSTRACT

The main objective of this study is to investigate the production of child affective prosody cross-linguistically. Taking ethical considerations into account, only acting and light induction methods were chosen to record child affective speech. Children were playing emotions on token utterances with the help of visual materials (16 repetitions for each emotion). English and French token utterances were selected with a similar sound and prosodic structure. This methodology enabled us to record a cross-linguistically comparable corpus of bilingual Scottish-French children and their monolingual peers. Acoustic measurements of pitch range variation and peak alignment are taken. This paper presents results for two bilingual children, comparing their prosody across emotions and across languages.

1. INTRODUCTION

The area of affective speech research has received much attention over the last years. Recent cross-linguistic studies (Scherer *et al.* 2001; Chen and Gussenhoven 2003) show that there is a number of available means to express emotions in speech (pitch range, rhythm, intonation patterns, voice quality, etc), but their usage, level of importance and meaning vary in different languages. They suggest that the vocal realization of emotions is based both on universal mechanisms and social/cultural conventions. There are several studies (Tickle 2000; Abelin and Allwood 2000) that show differences in the production and perception of emotions in several languages. All this research is concerned with adults, and studies of child affective speech are very rare and mostly based on perception only (van der Meulen *et al.* 1997; Clément 1999). The main objec-

tive of this research is to investigate the production of affective speech by monolingual and bilingual children cross-linguistically. Given that bilingual children produce two languages that differ in some acoustic correlates for affective speech, it is likely that they may differ in their production of affective speech from the monolinguals in each of their two languages. This phenomenon has been widely attested in bilingual studies of other phonetic aspects (Flege *et al.* 1997), and there is some evidence that this may also be the case for production and perception of emotion (Mejvaldova 2001).

2. METHODOLOGY

The starting point for this study was to develop an appropriate methodology to record child affective speech. Though spontaneous speech is claimed to contain the most authentic emotions, its collection gives rise to numerous difficulties: speakers should be recorded without knowing it, while expressing their emotions naturally, but such a routine is ethically problematic, especially with children. Thus, this study records acted speech which does not suffer the same ethical problems. Moreover, acting of emotions can be presented as a game to children, so that they can be more naturally involved in the expression of emotions. Acting has already been effectively used in the study of child facial expressions of emotions (Field and Walden, 1983).

Visual materials were developed in a realistic manner together with a professional illustrator and based on the research of facial expressions. They represent a child expressing four emotions (happiness, sadness, anger and fear); the child's gender and age are adapted to that of the recorded children. The randomized cards have 16 repetitions for each emotion. One token utterance was selected with a similar sound and prosodic structure for English and French: "I see a banana there./Je vois une banane, moi." The subject is instructed to say the token utterance in the same way as the child at the picture, thus the child is playing emotions through association with the drawing. Visual materials played a very important part in the recordings: they served as the reference of the expressed emotions, and an affect inducing material, as the subjects associate themselves with the picture.

Neutral was not represented as a facial expression card, neutral utterances were recorded in a separate test at the very beginning of the recording session. In this test, the child was instructed to say the token utterance (I see * there./Je vois *, moi.) with a word from a pile of cards (picture naming); 16 repetitions of the original word (banana/banane) were mixed among the others. Later all the utterances with the original token wording were selected for analysis. A small number of utterances was rejected due to noise, hesitation or incorrect wording. 16 monolingual and 8 bilingual children were recorded in total. Bilingual children were recorded in two sessions (one for each language) by different experimenters in order to control for language mode and code-switching; there was a period of about two weeks between the sessions. This methodology allowed us to record a cross-linguistically comparable corpus of 6 bilingual Scottish-French children and 12 monolingual peers, aged between 7 and 10.

2.1. Participants

This paper presents data for two bilingual girls, taken from the recorded corpus. Child 1 is 10, and child 2 is 7 years old. They were both born in Edinburgh, Scotland, with a French mother and a Scottish father. They go to Edinburgh mainstream schools, and attend the same French afternoon school twice a week. However there is some difference in the children's family language background and the amount of input the children are receiving in each of the languages. The parents of child 1 chose one parent - one language strategy. Father speaks English, and mother speaks French with the child, but the parents speak English amongst themselves. Therefore, the child receives French input mainly from her mother, and her home dominant language is English. The parents of child 2 chose to speak the less dominant language at home - French. The father of child 2 is Scottish, but he is fluent in French, and he speaks French with his wife and daughter. Child 2 has French as her dominant home language.

2.2. Recording

All children were recorded with a TASCAM DA-P1 portable DAT recorder either in the recording studio or in a quite school room. The speech was digitized and stored at 44100 Hz. The analysis was performed on the speech downsampled to 22050 Hz.

2.3. Measurements

Using *wavesurfer* software, acoustic measurements of pitch range variation and peak alignment were taken. All the utterances were manually labeled at the following pitch points: IF0 (initial value of F0), H1 (the peak of the first accented word), V (valley, the lowest point between two peaks), H2 (the peak of the main accented word - main accent), FL (final low - the lowest value after the peak). In cases where there was no well defined peak, an absolute F0 max value was taken near the accented syllable. Segment duration measurements were taken at the following points: the first consonant, the first vowel and the second consonant of the main accent. F0 values were then extracted automatically to data files with the help of ESPS algorithm, and again checked manually for any pitch perturbation or voice quality errors.

2.3.1. Pitch range

Following Ladd (Ladd 1996), we describe pitch range using two partially independent measures of variation: overall level and span. Overall level refers to the "height" of a speaker's voice, span refers to the width of pitch frequencies covered by a speaker (how big the excursions are). These measures have been shown (Patterson 2000) to give better and perceptually relevant results, than the more commonly used measures, such as statistical moments (mean, standard deviation, difference between maximum and minimum F0, etc.). In our measurements, FL point was taken as the overall level, and the difference of H2 and FL, as the span. All measures are expressed in semitones.

2.3.2. Peak alignment

Recent research showed that peak alignment is an important measurement for cross-linguistic and second language acquisition studies (Mennen 2004), and we believed that it might also be relevant for the present study. We present the peak alignment ratio: the ratio of the distance of the main peak from the onset of the main accented vowel as a function of the duration of that vowel. The CV duration of the main accent syllable in ms is also shown.

3. RESULTS

This paper presents results on pitch range (overall level and span), peak alignment ratio and CV duration for two bilingual children, comparing their emotions and languages. Separate multivariate ANOVAs were carried out for each child. There were two factors: EMOTION (five levels: anger, fear, sadness, happiness, and neutral) and LANGUAGE (two levels: English and French). The dependent variables were pitch span, pitch level, duration, and peak ratio. The significance of factor LANGUAGE varied across the dependent variables. Factor EMOTION was significant across all the variables for both children, and a post hoc Bonferroni test was carried out to investigate better its significance.

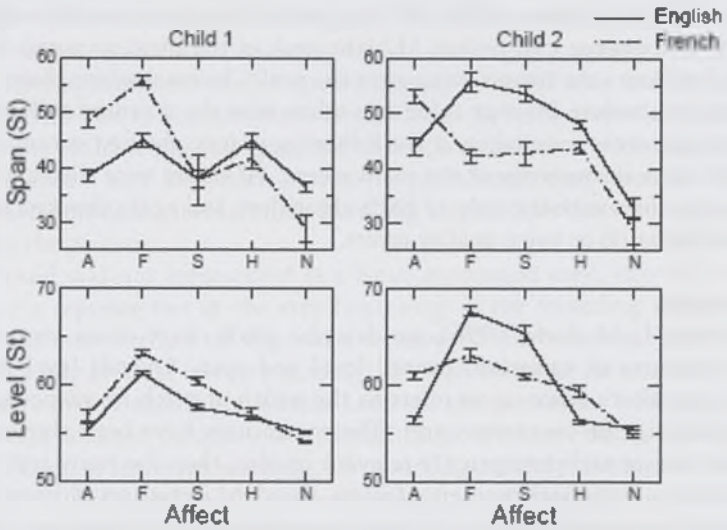


Figure 1: Pitch range for two bilingual children with standard errors. Abbreviations for Affect: A (Anger), F (Fear), S (Sadness), H (Happiness), N (Neutral).

3.1. Pitch range

The results for pitch range are shown in Fig.1. Results for Child 1, for the dependent variable span, show no significant main effect of LANGUAGE [$F(1,145)<1$). There was, however, a significant main effect of EMOTION [$F(4,145)=42.912$; $p<0.001$] (showing that the child realises the different emotions in a similar way in both languages), together with a significant interaction between the two factors [$F(4,145)=16.004$; $p<0.001$]. As can be seen in Figure 1 (top panel), this interaction is due to the fact, that in her French, the child realises anger and fear with a wider span, and neutral with a narrower span, than in her English. There is, however, no significant difference in the realisation of sadness and happiness in the two languages. Post hoc test showed that in English span was wider in fear and happiness (between which there was no significant difference), than in the other emotions (between which there also was no significant difference). In French, span was also widest in fear, but was then followed by anger and happiness (which were not significantly different from one another), and sadness and neutral (which were also not significantly different).

The results for pitch level of Child 1 show significant main effects of EMOTION [$F(4,145)=117.657$; $p<0.001$] (showing that in both languages fear has the highest pitch level, neutral has the lowest level, with anger, happiness and sadness just slightly higher than neutral), LANGUAGE [$F(1,145)=21.556$; $p<0.001$] (with a higher level in French than in English), as well as a significant interaction between the two factors [$F(4,129)=8.286$; $p<0.001$]. Figure 1 (bottom panel) exemplifies that the interaction is due to the fact that although French is higher in anger, fear and sadness, it is lower in neutral, and there is no significant difference in the realisation of happiness between the two languages. Post-hoc tests show that in English there is no significant difference between anger and sadness, happiness and sadness, and anger and neutral, whereas in French all emotions differ significantly from one another, apart from anger and sadness.

Child 2 differs from Child 1, in both pitch span and pitch level. Results for span show a significant main effect for LANGUAGE [$F(1,128)=20.902$; $p<0.001$], and EMOTION [$F(4,128)=42.05$; $p<0.001$], as well as a significant interaction [$F(4,128)=15.295$; $p<0.001$]. Child 2 has different realisations for some, but not all emotions in her two languages. Span is narrowest for neutral in both languages, as for Child 1. In French, anger has the widest span, with similar span for fear, sadness and happiness (between which there is no significant difference), positioned between anger and neutral. In English, however, fear and sadness (between which there is no significant difference) have the widest span, followed by sadness and happiness (which also do not significantly differ), anger and happiness (with similar spans) and finally neutral. Note that the general pattern for English is similar to that of Child 1, with the exception of sadness, which is realised with a much narrower pitch range by Child 1. Notice also that Child 2 does not differentiate pitch span for neutral in the two languages, but does differentiate sadness. Child 1, on the other hand, shows the opposite pattern in that it doesn't differentiate sadness, but does differentiate neutral in the two languages.

Pitch level of child 2 presents an interesting realisation pattern across emotions. The results show no significant effect of LANGUAGE [F (1,128<2], but a significant effect of EMOTION [F (4,128)=129.111; p<0.001], as well as an interaction between the factors [F (4,128)=38.52; p<0.001]. Figure 1 shows that the interaction is actually caused by the fact that anger and happiness have a higher level in English than in French, whereas fear and sadness have a lower level. Neutral has a similar level in each language. If we compare child 2 with child 1, it is clear that the former child shows a greater extent of language differentiation in the realisation of some emotions, and distinguishes the languages in a different way for other emotions (i.e. fear and sadness). Post hoc test for child 2 showed that only two emotions - fear and sadness - are significally different from each other [p<0.01], and from other emotions [p<0.001] in English level; neutral is significantly different from all the emotions [p<0.001], and fear and happiness are significantly different from each other [p<0.001] in French level.

The data shows that both children make cross-linguistic differences in the realisation of their emotions through pitch range. There are both similarities and differences in their realisation patterns across languages and across emotions.

3.2. CV duration and peak alignment ratio

The CV duration of the main accent syllable was measured in ms, and the results are shown in Fig. 2. The results for Child 1 main accent syllable duration show a significant main effect of EMOTION [F (4,145)=10.507; p<0.001], with the longest syllable duration in fear, followed by anger, sadness, neutral, and happiness. There was, however, no effect of LANGUAGE [F (1,145)<1], nor an interaction between the factors [F (4,145)<1].

Child 2 also shows no significant effect of LANGUAGE [F (1,128)<1.5]. However, there was a significant main effect of EMOTION [F (4,128)=6.661; p<0.001], as well as a significant interaction between the two factors [F (4,128)=3.1; p<0.006]. It is clear that child 2 has a different pattern than child 1. In French, anger has the longest duration which is significant. It is followed by fear, sadness, happiness and neutral (between which there is no significant difference [p<1.0]. In English, on the other hand, sadness has the longest syllable duration, followed by anger, and then by fear, happiness and neutral. There are also more significant differences between emotions than in French.

Peak ratio results are presented in Fig. 3. The results for child 1 peak ratio show a significant main effect of EMOTION [F (4,145)=6.062; p<0.001], of LANGUAGE [F (1,145)=21.373; p<0.001] and an interaction between the factors [F (4,145)=7.084; p<0.001]. The results for child 2 peak ratio show no significant main effect of LANGUAGE [F (1,128)<1], but they show a significant main effect of EMOTION [F (4,128)=6.716; p<0.001], and an interaction between the factors [F (4,128)=7.233; p<0.001]. The interaction is due to the fact that in most emotions the peaks occur later in French than in English, but in fear peaks occur earlier. Post hocs show that there are no significant differences between the timing of peaks in each emotion in English,

Figure 2: CV duration for the main accent syllable for two bilingual children with standard errors.

whereas in French peaks are aligned earlier in fear than in sadness ($p<0.01$), happiness ($p<0.001$), and neutral ($p<0.01$), between which there is no significant difference. Note also, that both children show more variability in their realization of each emotion in French, than in English, as reflected in larger standard errors.

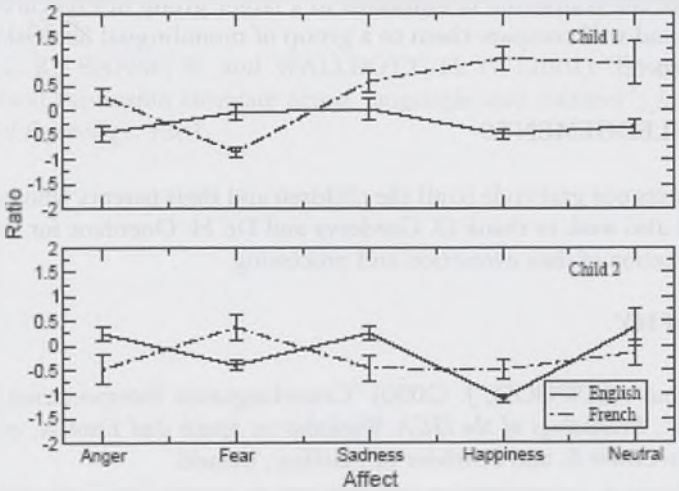


Figure 3: Peak alignment ratio for two bilingual children with standard errors.

4. DISCUSSION

Both children realise differences between emotions, and this is apparent in each of the measurements taken in this study, although it is most obvious in pitch span and level. However, the children differ in the way these emotions are realised across the two languages. Even though both children realise their emotions differently in each language, (as shown through a main effect of the factor LANGUAGE and/or through a significant interaction between the factors LANGUAGE and EMOTION), there are differences in their realization patterns across languages and across emotions. It is not possible to come to any firm conclusions about the reasons for these individual differences, until we have the results of monolingual English and French children and other Scottish-French bilingual children. However, once we obtain these data we will explore factors such as the children's age and amount of language use and exposure.

5. CONCLUSIONS

This study shows that our method for collecting a cross-linguistically comparable corpus of bilingual Scottish-French children and monolingual English and French children was appropriate. All children were able to act out emotions in both their languages. Our study also confirms that the chosen measures of pitch range and peak alignment are effective in capturing cross-linguistic differences in affective speech. It appears that bilingual children realise cross-linguistic differences, and sometimes realise emotions differently in their two languages. However, there is also some variability between speakers. The study presented here forms part of a larger study in which we will investigate the realisation of emotions in a larger group of French/English bilingual children and will compare them to a group of monolingual Scottish-English and French children.

6. ACKNOWLEDGEMENTS

We express our gratitude to all the children and their parents who participated in this study. We also wish to thank O. Gordeeva and Dr. H. Ouerdane for their assistance with automatization of data extraction and processing.

BIBLIOGRAPHY

ABELIN, A. and ALLWOOD, J. (2000) "Cross-linguistic interpretation of emotional prosody", *Proceedings of the ISCA Workshop on Speech and Emotion*, eds. Cowie R., Douglas-Cowie E. and Schröder M., Belfast, Ireland.

van der MEULEN, S., JANSSEN, P. and den OS, E. (1997) "Prosodic abilities in children with specific language impairment", *Journal of Communication Disorders*, 30, pp. 155-170.

CHEN, A. J. and GUSSENHOVEN, C. (2003) "Language-dependence in the signalling of attitude in speech", *Proceedings of workshop on the Subtle Expressivity of Emotion at CHI 2003 Conference on Human and Computer Interaction*, eds. Suzuki N. and Bartneck C.

CLEMENT, J. (1999) "Structure des représantations prosodiques. Développement normal et pathologique du traitement de la prosodie", PhD thesis, Université de Paris 5.

FIELD, T. M. and WALDEN T. A. (1983) "Production and discrimination of facial expressions by preschool children", *Child Development*, 53, pp. 1299-1300.

FLEGE, J., BOHN, O.-S. and JANG, S. (1997) "Effects of experience on non-native speakers in production and perception of English vowels", *Journal of Phonetics*, 25, pp. 437-470.

LADD, D. R. (1996) "Intonational phonology", Cambridge University Press.

PATTERSON, D. (2000) "A linguistic approach to pitch range modelling", PhD thesis, University of Edinburgh.

MEJVALDOVA, J. (2001) "Expressions prosodiques de certaines attitudes en tchèque et en français. Etude comparative", PhD thesis, Univerzita Karlova, Université de Paris 7.

MENNEN, I. (2004) "Bi-directional interference in the intonation of Dutch speakers of Greek", *Journal of Phonetics*, 32, pp. 543-563.

TICKLE, A. (2000) "English and Japanese speakers' emotion vocalization and recognition: A comparison highlighting vowel quality", *Proceedings of the ISCA Workshop on Speech and Emotion*, eds. Cowie R., Douglas-Cowie E. and Schr¨oder M., Belfast, Ireland, 2000.

SCHERER, K. R., BANSE, R. and WALLBOTT, H. G. (2001) "Emotion inferences from vocal expression correlate across languages and cultures", *Journal of Cross-Cultural Psychology*, 32, 1.

CROSS-LINGUISTIC PROSODIC VARIATION IN TWO ROMANCE LANGUAGES: PORTUGUESE AND ITALIAN

GRAZIA M. INTERLANDI
Scienze della Comunicazione - Universita di Torino - Italia
ANTONIO ROMANO
Dipartimento di Scienze del Linguaggio - Universita di Torino - Italia
LURDES DE CASTRO MOUTINHO, ROSA LÍDIA COIMBRA
Centro de Investigação de Línguas e Culturas - Universidade de Aveiro - Portugal

1. INTRODUCTION[1]

The aim of this study is to summarize some of the main differences into two quite different Romance areas: Portuguese and Italian. The work we are carrying on is part of a joint research project (AMPER, *Atlas Multimédia de la Prosodie de l'Espace Romaine* -Multimedial Prosodic Atlas of Romance Areas) which is collecting data on prosodic variations throughout these linguistic domains. The project involves nowadays several universities whose work is coordinated by the Universities of Grenoble (France), Turin (Italy), Barcelone-Tarragone (Spain), Iaşi (Romania), Aveiro (Portugal).

Since we do not pretend to deal with two homogeneous languages, but with a variety of dialects, the database commented in our presentation is constituted by recorded speech collected from different varieties of the two languages here compared. Sentence intonation and pitch patterns have been analysed according to the same methodological criteria and prototypical curves submitted to perceptive validation.

The acoustic parameters which are considered for this study are the variation of F0 in the sentence, the alignment of the intonational configurations with the syllabic tier and the vowel duration.

---

1 The final version of the paper has been written by Grazia M. Interlandi, who also analysed and discussed the topics concerning the pitch *continuum* of varieties spoken in Turin (see 3.2). The other authors contributed with the speech make-up, figures and graphics already published in previous studies and general observations about the partial outcomes of *AMPER-ITA* and *AMPER-POR*.

As we show at the end of this presentation, the results show a significant variation inter- and intra-language as it regards the yes-no question, both for F0 variation and vowel duration patterns.

2. THE CORPUS AND THE METHODOLOGICAL APPROACH

The corpus consisted of spoken speech elicited with the methods illustrated in the volume *AMPER* appeared as the *hors serie* no. 3 of the review *Géolinguistique*. The sentences that are here analysed have been recorded in two different areas for both languages, one in the North (Beira Litoral, for Portuguese, and Turin, for Italian) and one in the South (Alentejo in Portugal and the South Salento in Italy).

Yes-no questions and declarative sentences have been constructed with the same criteria, as it regards word stress, sentence accent, vowel quality and phonetic contexts, so that we can compare the prosodic strategies of the two languages.

The speech has been analysed with the *AMPER-fox* scripts, that have been implemented specifically for the Project.

We have observed F0 variations in the vocalic segments, the alignment of the intonation configurations in the syllable and the vowel duration. The pictures we selected for this comparison show stylized pitch contours of some examples of declarative sentences and yes-no questions in the two varieties of Portuguese and Italian analysed in this study. Furthermore, we show the vowel duration graphs concerning the examples of intonation curves.

3. RESULTS OVERVIEW

3.1 Portuguese

Both varieties of Portuguese show in the declarative sentence a general declination of the pitch contour.

As it regards yes-no questions, significant differences can be noted: the speaker from Beira Litoral shows more variation in the pitch contours and use a rising-falling intonation in correspondence with the sentence accent; on the contrary, the speaker of Alentejo use a similar intonation configuration in the declarative sentence and in the yes-no questions. Moreover, the timing of F0 variations is different: the prosodic variation can be observed then in the different alignment of pitch contours with the syllabic tier.

Vowel duration is also different into the two Portuguese varieties and comparing the two sentence modalities.

3.2 Italian

The two varieties of Italian do not show significant differences in the declarative sentence: we observe a declination of F0, that is typical in this modality. The speakers of Turin show however a greater pitch range on the NP; in the Figure 5 a topicalization on the subject is marked also by the length of the accented syllable.

Concerning the yes-no questions, we can note some different pitch contours, comparing both the two varieties and the two speakers of Turin.

The Italian spoken in Turin can be described as a real geographical-social *continuum* of varieties marked by different intonation configurations: we observe actually that F0 is always falling-rising on the accented syllable and falling on the last postaccented syllable, but F0 on the end of the sentence can be on a middle or on a low level; we can characterize therefore two different prosodic varieties of Turinese, the Piedmontese Regional Italian (more prototypical; Fig. 3) and the New Turinese (a new intonation model, especially used by the young people; Fig. 4). Other configurations are possible, but they are less prototypical for a yes-no question and are used on the contrary to mark a check question (Fig. 5).

The Italian variety spoken in Salento shows a rising intonation on the accented syllable and a falling pitch contour on the postaccented syllables (Fig. 6).

Also different is the pitch range used in the two varieties: the Italian spoken in Turin usually shows a larger range (here > 150 Hz) than the Italian spoken in Salento (about 80 Hz). Moreover, the top of the pitch contour is generally aligned with the first accented syllable in both Italian dialects, but always with the postaccented syllable in some varieties of Turinese.

As it regards the vowel duration in Turinese, we note significant differences in Fig. 4 and 5 compared with Fig. 3: vowel duration is greater in the accented syllables of the question than of the declarative sentence, whereas the contrary is true for the examples in Fig. 4 and 5, that seems confirming the prosody change in the more stardardized New Turinese compared with the Piedmontese Regional Italian.

4. CONCLUSIONS

By comparing the F0 contour variations in these two Romance areas we note strategically different prosodic patterns, especially concerning the time structure. As for pitch contours, more in detail we observed that:

- in both languages the declarative sentence shows a declination contour reproducing what is now known as a prosodic universal;
- the Portuguese uses a similar intonation in the yes-no questions compared to the declarative sentence, whereas the Italian varieties oppose a declination pattern leading to a global rising-falling configuration;
- the Italian speakers use a greater pitch range than the Portuguese ones and reach a higher peak of F0 on the accented syllable of the NP;
- we observe, also, a similar intonation comparing the speakers from Salento and from Beira Litoral, even if the timing of the F0 peak is different (on the accented syllable in Portuguese, on the postaccented syllable in Italian); the Portuguese variety of Alentejo on the contrary is close to the prototypical intonation of Turinese, but with a different the pitch range (see above).

As it regards timing, we confirmed the common tendency to cancel unstressed vowels in Portuguese; moreover, the vowels with length higher than 50 ms appear more frequently in Italian than in Portuguese.

In conclusion, with the choice of varieties here discussed we found the expected different prosodic strategies of the two Romance languages analysed in this study.

The possibility of generalizing these outcomes obviously undergoes a careful intra-language reflection, like the one carried over on the different areas here considered in previous research work by the four authors who contributed to this comparison.

BIBLIOGRAPHY

CONTINI M., LAI J. P., ROMANO A., ROULLET S.: «Vers un atlas prosodique parlant des variétés romanes». In: BOUVIER L.-C. et al. (orgs.), *Mélanges offerts à X. Ravier*, CNRS, Univ. de Toulouse Le Mirail, 2003, pp. 73-84.

CONTINI M., LAI J. P., ROMANO A., ROULLET S., de CASTRO MOUTINHO L., COIMBRA R.L., PEREIRA BENDIHA U., SECCA RUIVO S.: «Un projet d'atlas multimédia prosodique de l'espace roman». In: *Speech Prosody 2002. Proceedings of the 1st International Conference of Speech Prosody* (Aix-en-Provence, 11-13 April 2002), Aubenas d'Ardèche, Lienhart, 2002, pp. 227-231.

INTERLANDI G.M.: *L'intonazione delle interrogative polari nell'italiano parlato a Torino: tra varietà regionale e nuova* koiné, Tesi di Dottorato non pubblicata, Università degli Studi di Pavia, 2004.

INTERLANDI G.M.: "Il continuum della variazione pragmatico-espressiva nell'intonazione dell'italiano parlato a Torino". In: ALBANO LEONI F., CUTUGNO F., PETTORINO M., SAVY R. (a cura di), *Il parlato italiano, Atti del Convegno nazionale* (Napoli, 13-15 febbraio 2003), Napoli: M. D'Auria Editore, 2005, CD-Rom.

INTERLANDI G.M.: "L'italiano parlato a Torino: varietà regionale o nuova koiné? L'intonazione delle interrogative polari, *Géolinguistique*, 3 (hors série: Projet AMPER - Atlas Multimédia Prosodique de l'Espace Roman), 2005, pp. 217-240.

INTERLANDI G., ROMANO A.: «Quale intonazione per il torinese?». In: REGNICOLI A. (a cura di), *La fonetica acustica come strumento di analisi della variazione linguistica in Italia. Atti delle XIIe Giornate di Studio del G.F.S.*, (Macerata, 13-15 dicembre 2001), Roma: Il Calamo, 2001, pp. 117-122.

INTERLANDI G.M., ROMANO A.: "Le continuum intonatif de l'italien parlé à Turin: résultats d'un test d'identification». *Proc. Of the Workshop MIDL 2004 «Identification des langues et des variétés dialectales par les humains et par les machines»* (Paris, 29-30 nov. 2004), Paris: Ecole Nationale Supérieure des Télécommunications, 2004, pp. 157-160.

de CASTRO MOUTINHO L., COIMBRA R.L., SECCA RUIVO S., PEREIRA BENDIHA U.: «Contribuição para o estudo da variação prosódica do Português Europeu". In: SÁNCHEZ MIRET F. (org.), *Actas do XXIII CILFR* (Salamanca, España, 22-28 Set. 2001), vol. I, Tübingen: Niemeyer, 2001, pp. 245-252.

de CASTRO MOUTINHO L., COIMBRA R.L., SECCA RUIVO S., PEREIRA BENDIHA U.: "Project d'atlas prosodique multimédia des variétés romanes. Contribution Portugaise", *Travaux de l'Institut de Phonétique de Strasbourg* (TIPS), 31, 2001, pp. 61-70.

de CASTRO MOUTINHO L., COIMBRA R.L., PEREIRA BENDIHA U., ROMANO A., CONTINI M.: "Estudo comparativo da variação prosódica em duas línguas românicas: o Português e o Italiano", *Actas do XIX Encontro Nacional da APL - Associação de Linguística Portuguesa* (Lisboa, outubro 2003), Lisboa, APL, 2004, pp. 719-723.

ROMANO A.: "Un projet d' Atlas multimédia prosodique de l'espace roman (AMPER)». In: SÁNCHEZ MIRET F. (org.), *Actas do XXIII CILFR* (Salamanca, España, 22-28 Set. 2001), vol. I, Tübingen: Niemeyer, 2001, pp. 279-294.

ROMANO A.: «Applicabilité des systèmes de transcription et d'analyse de l'intonation aux cas de variabilité dialectale présentés par la situation géoprosodique italienne». In: AUBERGÉ V., LACHERET-DUJOUR A. & LŒVENBRUCK H. (orgs.), *Actes des Journées Prosodie 2001* (Grenoble, 10-11 Oct. 2001), 2003, pp. 115-118.

ROMANO A., INTERLANDI G.M.: «Variabilità geo-socio-prosodica: dati linguistici e statistici», *Géolinguistique*, 3 (hors série: Projet AMPER - Atlas Multimédia Prosodique de l'Espace Roman), 2005, pp. 259-280.

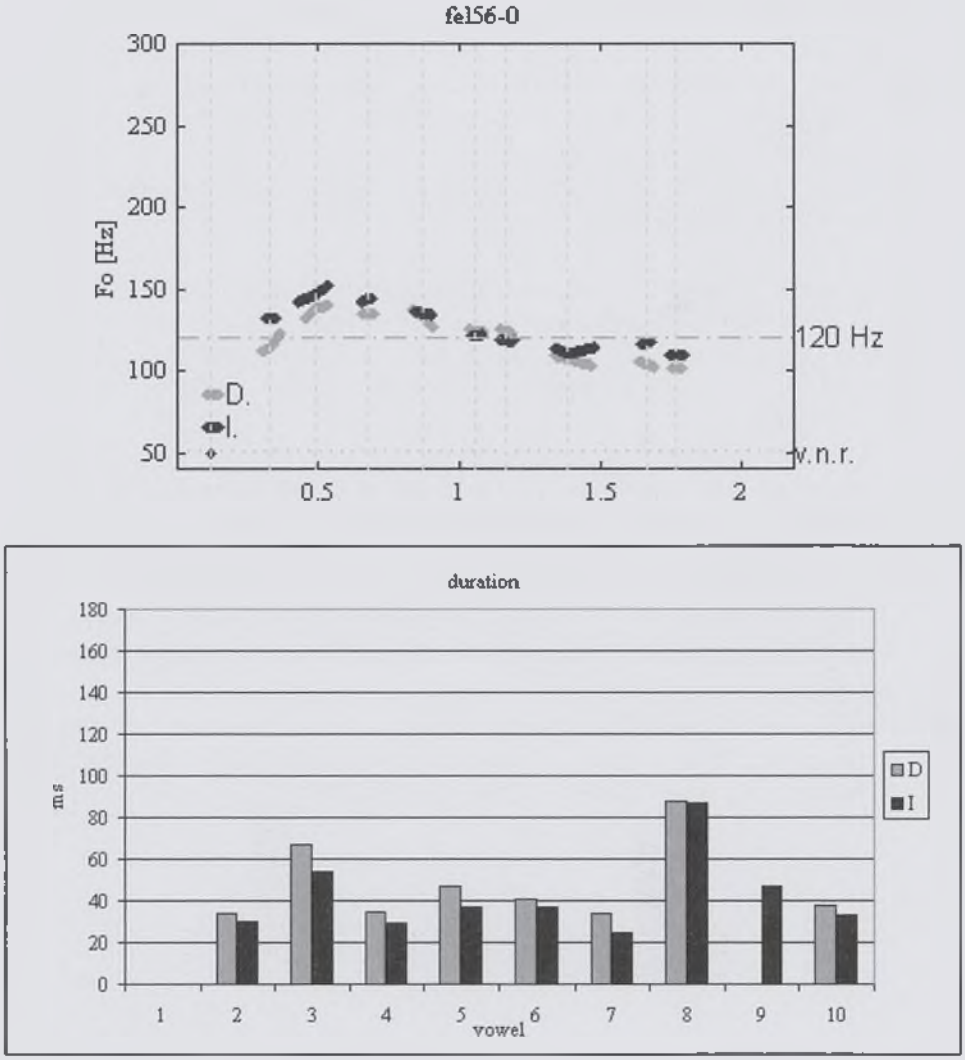
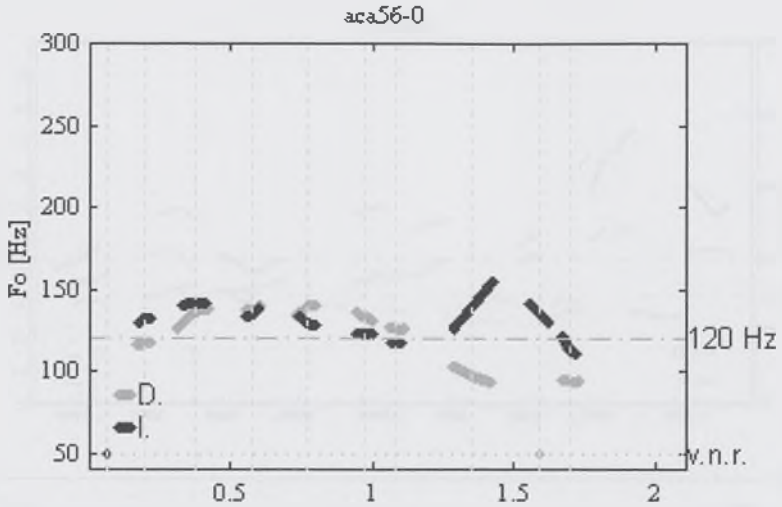


Fig. 1 - Portuguese of Beira Litoral (stylized pitch contour and vowel duration graph).
Sentence: "O Toneca toca no pássaro./?"



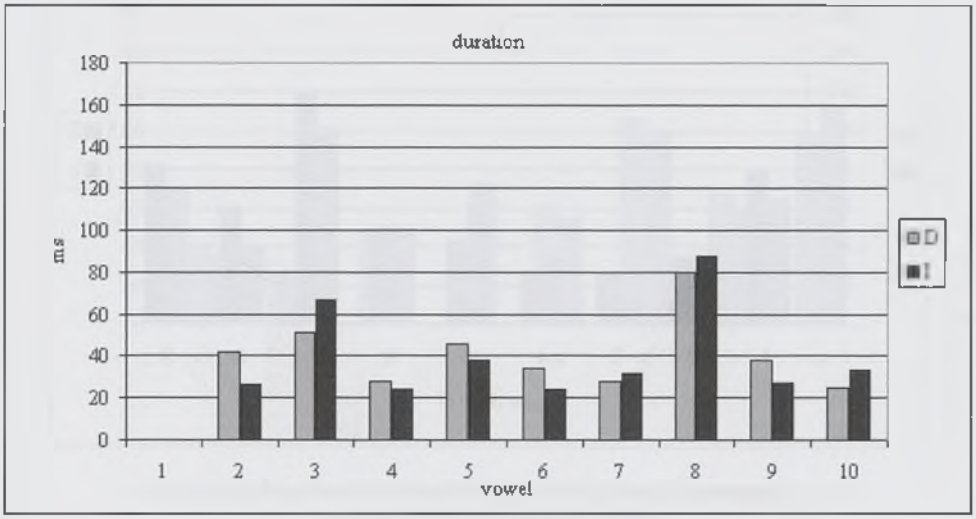
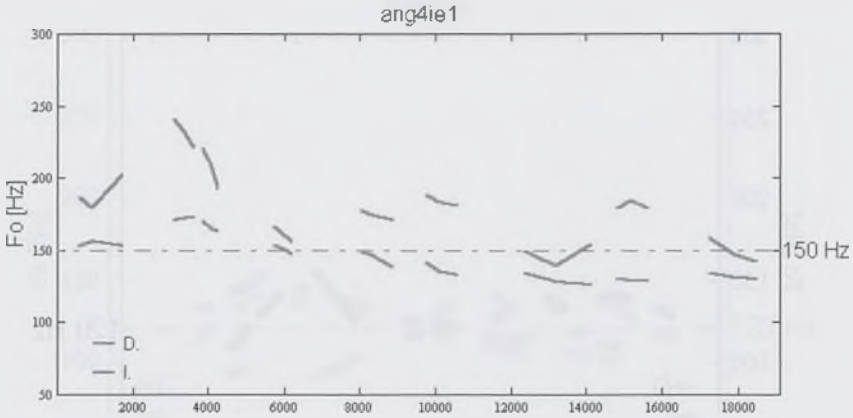


Fig. 2 - Portuguese of Alentejo (stylized pitch contour and vowel duration graph).
Sentence: "O Toneca toca no pássaro.!?"



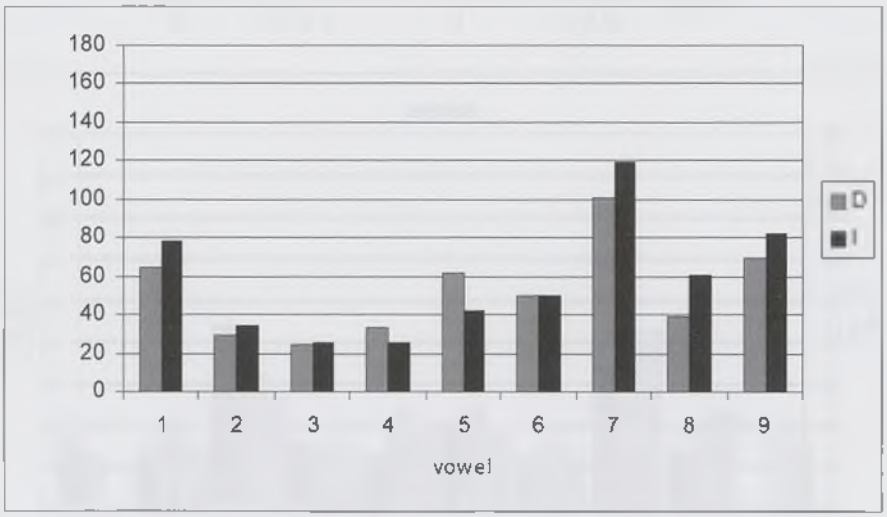


Fig. 3 - Piedmontese Regional Italian (stylized pitch contour and vowel duration graph).
Sentence: "Anna ha sentito il medico.!?"

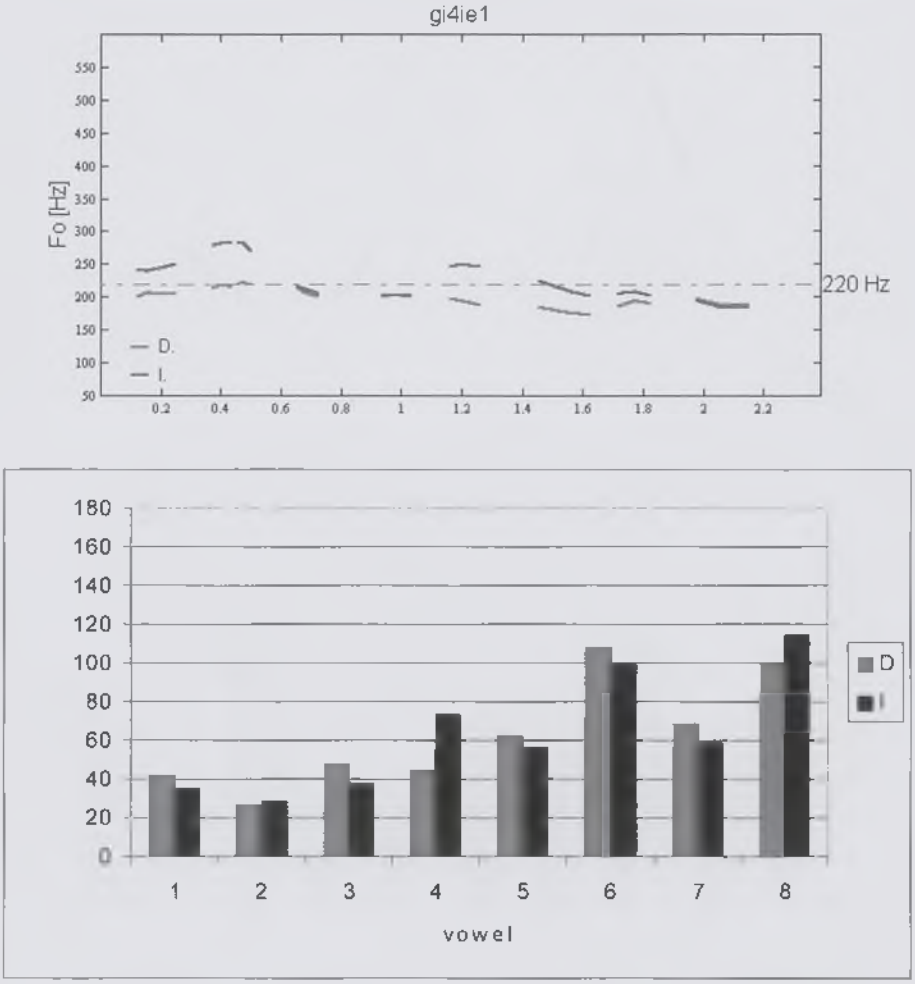


Fig. 4 - New Turinese (stylized pitch contour and vowel duration graph).
Sentence: "Anna ha sentito il medico.*I?*"

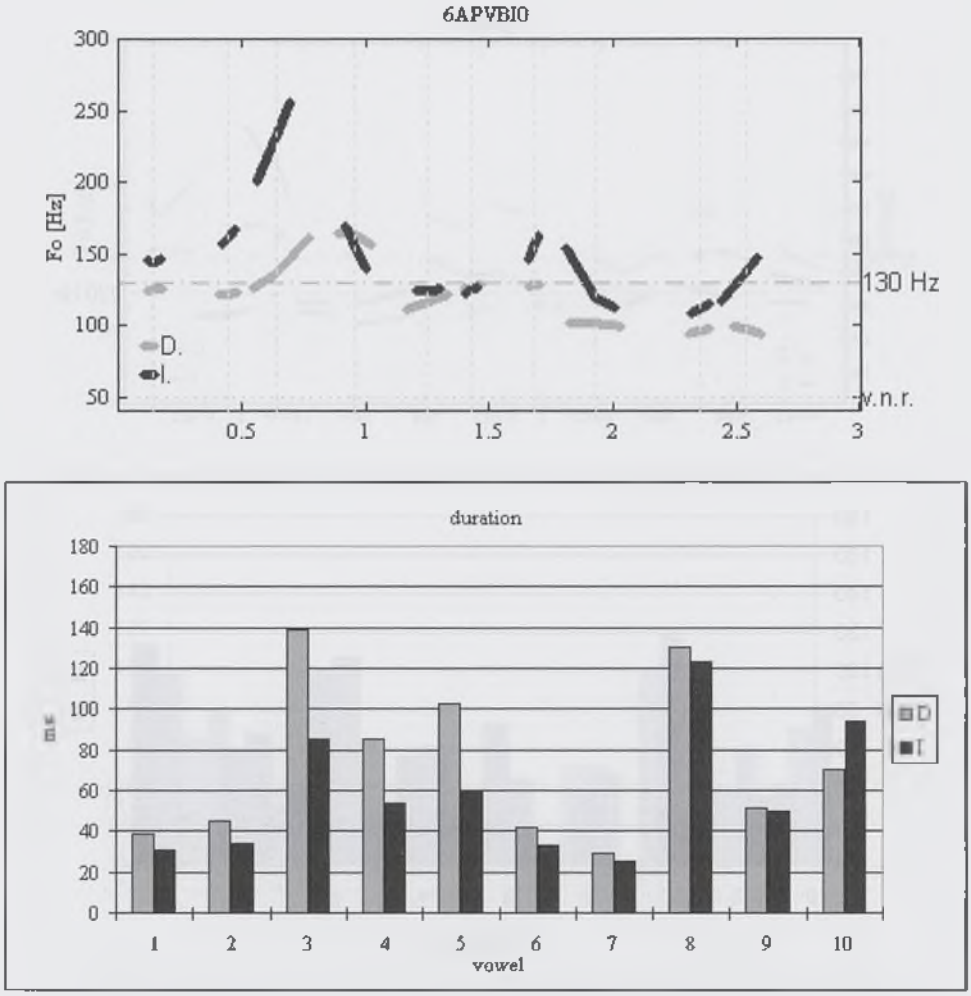


Fig. 5 - Standard Italian of Turin (stylized pitch contour and vowel duration graph).
Sentence: "Il pilota vola su Bàssora./?"

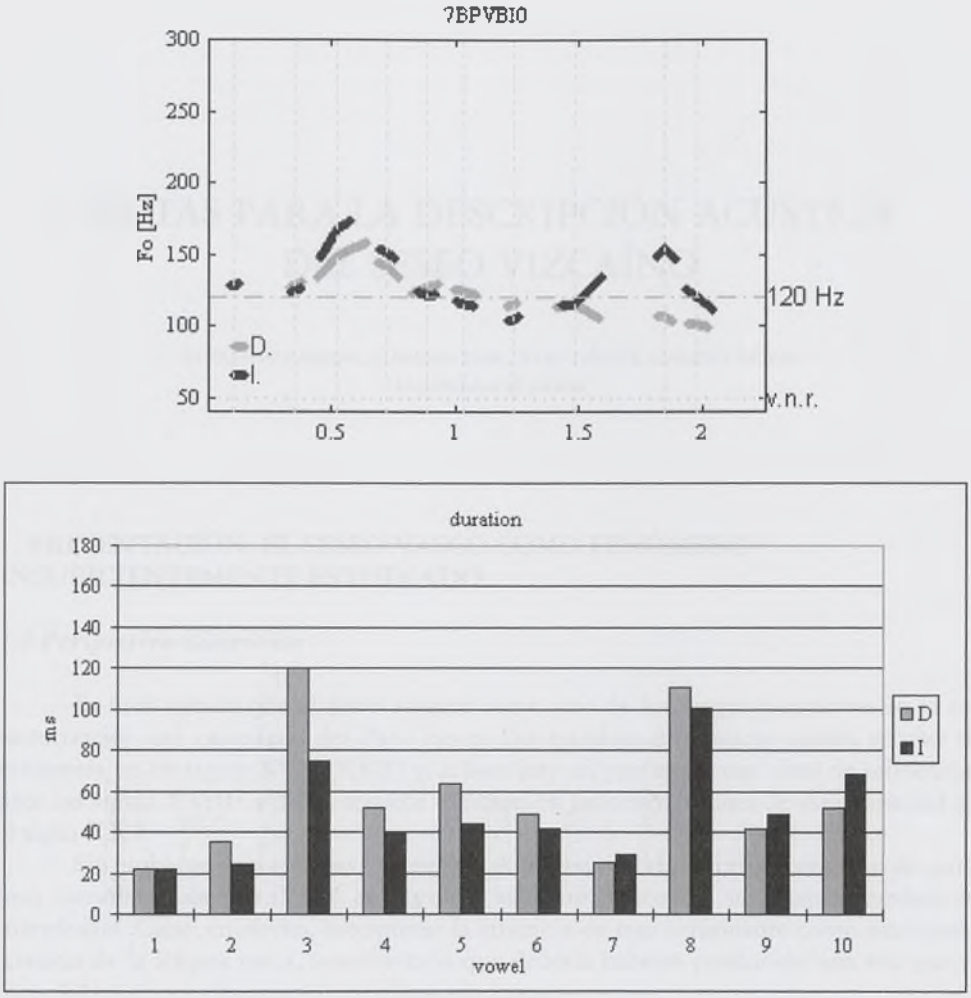


Fig. 6 - Italian of South Salento (stylized pitch contour and vowel duration graph).
Sentence: "Il pilora vola su Bassora.!?"

NOTAS PARA LA DESCRIPCIÓN ACÚSTICA DEL SESEO VIZCAÍNO

ALEXANDER IRIBAR, CARMEN ISASI, SARA GÓMEZ, CARMEN MORAL
Universidad de Deusto

1. PRESENTACIÓN: EL SESEO VASCO COMO FENÓMENO INSUFICIENTEMENTE ESTUDIADO

1.1.Perspectiva diacrónica

Es bien sabido que el seseo aparece como uno de los rasgos constantes en la caracterización del castellano del País Vasco. Los estudios diacrónicos suelen señalar su existencia en los siglos XVI y XVII[1] y, si bien hay un paréntesis casi total de referencias para los siglos XVIII[2] y XIX, aparece también en las descripciones de esta variedad en el siglo XX.[3]

Sin embargo, son muchas las preguntas no respondidas sobre el conjunto de cambios consonánticos que derivó en algunos hablantes vascos en un sistema carente de interdental. Cabe, en efecto, interpretar la ausencia de esta consonante como una transferencia de la lengua vasca, transferencia que debería haberse producido una vez que se consolidara el consonantismo castellano moderno.

Frente a esta perspectiva, no deberíamos olvidar que los testimonios hallados en los documentos tardo-medievales vizcaínos permiten identificar la confusión en una etapa anterior a la generalización de la interdental castellana, y replantear, en consecuencia, si el rechazo de [θ] por parte de los hablantes de lengua vasca es la única clave para el entendimiento de la cuestión. Como ya se propone en Isasi (*1998*: 747):

1 Recordemos, entre otras referencias, ECHENIQUE (*1997*), ISASI (*1999*).
2 Vid. ISASI (*en prensa*), GÓMEZ-RAMÍREZ (*2003*).
3 Véase, por ejemplo, TÚRREZ, I. (*1988*); OÑEDERRA, L. (*2004*).

> El origen de las confusiones debe hallarse bien en los alófonos de /ʂ/ y /ʐ/ predominantes en el castellano de la zona, bien en el propio sistema de sibilantes del euskera, donde podría darse la neutralización de las fricativas que los vasquistas describen hoy día para algunas áreas. Las interrogantes afectan por tanto a la historia de la evolución consonántica del castellano, pero se extienden también a un capítulo igualmente incompleto de la historia de la lengua vasca.

Por otra parte, al margen de cuáles fueran los motores del cambio, disponemos igualmente de un conocimiento sumamente limitado de su diacronía, tanto en su vertiente de sociolingüística histórica como en otros aspectos, v.g. la importancia de la difusión léxica en la expansión y /o en la conservación de soluciones seseantes.

Suponemos, en definitiva, que los pares medievales de apicales y predorsales se neutralizaron en un sólo elemento; un proceso que según ISASI (*1999*) se prodría considerar más propiamente como *çeçeante* para sus primeras etapas documentadas (en el caso de Vizcaya fines del XV y principios del siglo XVI), aunque la reducción desembocara en un *seseo* sobre cuya realización apenas se ha escrito tampoco.

1.2.El seseo moderno

El conocimiento de este fenómeno en lo contemporáneo es también muy parcial. Junto a la carencia de estudios sociolingüísticos suficientes sobre su distribución que vayan más allá de presuposiciones[4], sorprende que se venga transmitiendo una noción de *seseo* en la que no se ha establecido -que sepamos- cuál es el elemento resultante de la neutralización. Esta circunstancia diferencia el español del País Vasco del de otras áreas confundidoras, para las que se cuenta con datos al parecer más seguros sobre las realizaciones y la integración fonológica de resultados equivalentes[5].

En este sentido, y a falta de propuestas fundamentadas en otros datos, se pueden manejar dos considerandos: por una parte, que *la ese* del castellano en Vizcaya es predominantemente una apical[6] y, en segundo lugar, que tal vez el resultado confundidor pueda tener relación con la presencia o ausencia de elementos apicales y predorsales en el euskera de cada hablante. En cuanto a este último aspecto, hay que recordar igualmente que en Vizcaya la distinción /s/-/ʂ/ se ha perdido en gran parte del territorio.[7]

2. EL PRESENTE TRABAJO: OBJETIVOS Y METODOLOGÍA

Este trabajo pretendía, en su planteamiento inicial, aportar datos para una caracterización fonética del seseo vizcaíno desde una perspectiva experimental y eminentemente

---

4 Como atribuir el rasgo a personas de edad, de escasa cultura o de ámbito aldeano.

5 Prescindimos aquí de la revisión bibliográfica de este aspecto colateral.

6 BIZCARRONDO-ELEJABEITIA (*1992*).

7 URRUTIA-ETXEBARRIA-TÚRREZ-DUQUE (*1988*).

acústica. Los primeros datos obtenidos aconsejaron la adopción de dos objetivos secundarios: indagar en las posibles diferencias entre las realizaciones sibilantes seseantes y las procedentes de la *ese* canónica, y averiguar la relación existente entre las realizaciones sibilantes del español y las del euskara. Con todo, se debe subrayar que estas notas son sólo una aproximación inicial a un trabajo de mayor calado: no se pretende aquí tratar cuestiones como el origen del seseo, su distribución geográfica, su caracterización sociolingüística o su implicación fonológica.

Para el objetivo primero se efectuó una encuesta entre un pequeño número de informantes bilingües de diversos pueblos vizcaínos donde pueden encontrarse individuos seseantes.[8] El análisis de las primeras grabaciones aconsejó reducir el estudio a los materiales obtenidos en la localidad de Lekeitio, ampliando los datos mediante una nueva encuesta, en función de los objetivos secundarios señalados.

Así pues, la primera encuesta se realizó a dos informantes seseantes de Lekeitio.[9] De este material se han seleccionado, para cada informante, 30 apariciones de la sibilante seseante,[10] más otras 60 apariciones de [s] no seseante, para establecer la oportuna comparación. Además, como segundo grupo de control, se han seleccionado en cada informante 40 realizaciones sibilantes en euskara, 20 provenientes del fonema predorsodental /s/ (grafía <z>) y otros 20 correspondientes al fonema apical /ṣ/ (grafía <s>).[11]

Estas 280 realizaciones sibilantes han sido analizadas con el sonógrafo KAY DSP 5500 de la Kay Elemetric Corp. Las magnitudes medidas han sido las siguientes:

- Intensidad
- Frecuencia de inicio de la fricación
- En los casos oportunos, inicio de una zona inferior de fricación más débil.
- Inicio y final de las zonas interiores de mayor energía que se manifestaran en los primeros 8 KHz. del espectro. Se han distinguido hasta tres de estas zonas.[12]

Todos estos datos se han tratado con el programa estadístico SPSS. Además de las ya señaladas, el archivo incluye las siguientes variables: informante, localidad, lengua, origen de la sibilante, palabra, sílaba, posición en la sílaba (explosiva o implosiva), elemento anterior, elemento posterior y notas.

Los detalles sobre la segunda encuesta y los experimentos complementarios, se desarrollarán más abajo.

---

8 La encuesta consistió en la grabación de una conversación libre, en la que el tópico principal fue *cómo ha sido tu vida en la marina aldea*. La conversación se repitió, algo más brevemente, en euskara. El material se obtuvo con un micrófono SONY ECM-717 y se grabó en un DAT SONY TCD-D100.

9 El Informante 1 es J.U.E., de 66 años. El Informante 2 es F.G.B., de 72 años. Ambos son pescadores jubilados, lekeitianos y euskaldunes de nacimiento.

10 En realidad, el número de apariciones que superaba este número era pequeño.

11 No se ha considerado para este trabajo la sibilante palatal /ʃ/, que no guarda relación con nuestro objeto de estudio.

12 Como puede observarse, en este primer análisis no se han incluido algunas mediciones importantes, como final de la fricación, transiciones vocálicas, etc.

3. CARACTERIZACIÓN DEL SESEO VIZCAÍNO: EL INFORMANTE 1

Tomadas en su conjunto, las realizaciones de [s] en español y en euskara del informante muestran una gran semejanza. Ambas comienzan a una altura frecuencial similar y presentan zonas interiores de energía muy parecidas.[13] La siguiente tabla muestra los datos generales de las dos sibilantes del informante:[14]

	[s] del español	[s] del euskara
Inicio de la fricación	2.214 (S=257)	2.238 (S=226)
Frecuencia inferior	1.416 (30%)	1.140 (15%)
1ª zona de intensidad	2.637 – 4.003 (98%)	2.837 – 4.115 (85%)
2ª zona de intensidad	4.995 – 6.368 (64%)	4.994 – 6.030 (40%)
3ª zona de intensidad	5.193 – 6.433 (6%)	5.380 – 6.880 (5%)

Es necesario señalar que, en ocasiones, el valor medio del final de una de las zonas de mayor energía es inferior al del comienzo de la zona superior. Esto ocurre por la variación que registran estas zonas. Si manejáramos un valor medio para cada zona, los tres valores mantendrían siempre la gradación esperable.

Para caracterizar el seseo del informante, debemos examinar con más detalle las posibles diferencias entre la [s] seseante y la canónica, y comparar ambas con las realizaciones sibilantes del euskara. Los datos principales aparecen en la siguiente tabla:

	/s/ del español		/s/ del euskara	
	Origen: /θ/	Origen: /s/	Origen: /s/	Origen: /ʂ/
Inicio de la fricación	2.150	2.247	2.249	2.228
Frecuencia inferior	1.408	1.426	1.240	940
1ª zona de intensidad	2.558 – 3.338	2.679 – 4.091	2.857 – 4.135	2.814 – 4.091
2ª zona de intensidad	5.069 – 6.374	4.939 – 6.364	4.455 – 5.424	5.317 – 6.394
3ª zona de intensidad	5.240 – 6.780	5.170 – 6.260	5.120 – 6.640	5.640 – 7.120

13 Por supuesto, pueden señalarse pequeñas diferencias, como, por ejemplo, que las zonas interiores de energía aparecen más claramente en la [s] del español, o que todos los valores de ésta son ligeramente inferiores (salvo el correspondiente a la frecuencia inferior, que presenta, curiosamente, la mayor diferencia entre ambas realizaciones). En cualquier caso, estas pequeñas diferencias, y otras que podrían señalarse a lo largo del trabajo, no resultan significativas para nuestro propósito.

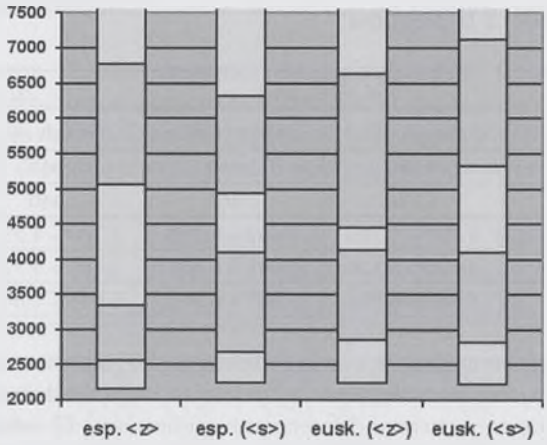
14 Los valores corresponden, en cada lengua, a todas las realizaciones sibilantes, independientemente de su origen. Todos los valores son medias, expresadas en Hertzios. Debajo del correspondiente al inicio de la fricación se coloca su desviación estándar. Debajo del resto de los valores aparece el porcentaje de casos en los que se ha medido cada dato.

Como puede apreciarse, las dos realizaciones del español son prácticamente idénticas; ambas comienzan a una altura similar, y las zonas interiores presentan, como es lógico, ligeras variaciones: algo más baja la primera y la segunda, y algo más alta la tercera en la [s] seseante. De acuerdo a estos datos, las dos *eses* son prácticamente iguales; de mostrar alguna desviación en su punto de articulación en relación al general, la realización seseante sería, en todo caso, algo más atrasada.

Debemos ahora examinar la relación de la sibilante española con la del euskara. La cuestión es doble: por una parte, ¿el informente posee una o dos realizaciones sibilantes en euskara?; por otra, ¿es diferente la [s] del español (seseante y no seseante) de la -o las- del euskara?

El Informante 1 refleja plenamente el proceso de reducción fonológica experimentado en el dialecto vizcaíno del euskara, de modo que no hay apenas diferencia entre las realizaciones sibilantes de /s/ (<z>) y /ʂ/ (<s>). Si bien, atendiendo a la frecuencia de inicio de la fricación, las realizaciones de /s/ parecen ligeramente más adelantadas (tan sólo 21 Hz.), las zonas superiores de energía son más altas en las realizaciones de /s/.

En definitiva, no puede señalarse ninguna diferencia significativa entre las realizaciones sibilantes del informante, independientemente de la lengua y del origen fonológico. La siguiente gráfica representa esquemáticamente la extensión frecuencial de las 4 *eses* examinadas:[15]



15 En bastantes ocasiones, las zonas de intensidad se solapan, de modo que el final medio de una es superior al inicio medio de la siguiente. En esos casos, la zona se ha agrupado.

4. CARACTERIZACIÓN DEL SESEO VIZCAÍNO: EL INFORMANTE 2

La siguiente tabla muestra los datos generales de las realizaciones sibilantes en español y euskara:

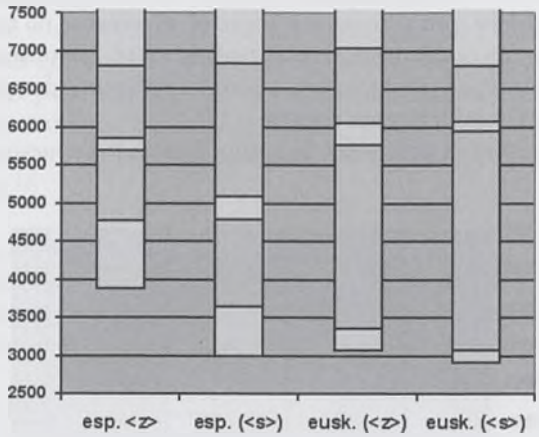
	[s] del español	[s] del euskara
Inicio de la fricación	3.246 (S=616)	2.988 (S=453)
Frecuencia inferior	2.660 (17%)	1.731 (17%)
1ª zona de intensidad	3.692 – 4.817 (52%)	3.158 – 4.310 (48%)
2ª zona de intensidad	5.062 – 6.131 (35%)	4.700 – 5.849 (80%)
3ª zona de intensidad	5.848 – 6.830 (42%)	6.070 – 6.985 (19%)

Se aprecia claramente que la [s] del español presenta unos valores siempre superiores a los de la [s] del euskara. La diferencia no es muy acusada, pero no parece, desde luego, desdeñable.

¿A qué se debe dicha diferencia? La siguiente tabla proporciona la información necesaria para responder a la pregunta:

	/s/ del español		/s/ del euskara	
	Origen: /θ/	Origen: /s/	Origen: /s/	Origen: /ʂ/
Inicio de la fricación	3.890	2.992	3.067	2.913
Frecuencia inferior	3.176	2.461	2.040	1.320
1ª zona de intensidad	4.780 – 5.700	3.652 – 4.785	3.360 – 4.256	3.071 – 4.335
2ª zona de intensidad	4.520 – 5.320	5.093 – 6.177	4.583 – 5.778	4.860 – 5.944
3ª zona de intensidad	5.856 – 6.816	5.841 – 6.842	6.066 – 7.046	6.080 – 6.800

Los datos ponen en evidencia que la *ese* seseante del informante es bien distinta de la canónica. Esta se parece notablemente a las realizaciones sibilantes del euskara, que, por otra parte, no muestran entre sí diferencias significativas. El informate 2 resulta, en definitiva, diferente del anterior, puesto que presenta dos realizaciones de *ese*: una [s] para el euskara y la canónica en español, y una [ʂ] para el seseo. La gráfica siguiente lo ilustra con claridad:



4.1. Segunda encuesta: pares mínimos

Para confirmar la existencia de las dos realizaciones sibilantes, se procedió a realizar una nueva encuesta al Informante 2, con el objetivo de conseguir una colección de posibles pares mínimos que distinguieran la *ese* seseante de la canónica. Se preparó una encuesta de recursos múltiples,[16] que se realizó con los mismos procedimientos técnicos que la anterior.

Se consiguieron 17 pares mínimos, del tipo *peces-peses*. Conviene resaltar que el registro de habla obtenido posee una mayor formalidad que el de la encuesta primera. Los resultados principales del análisis acústico realizado se muestran en la siguiente tabla (se repiten también los datos anteriores, para su oportuna comparación):

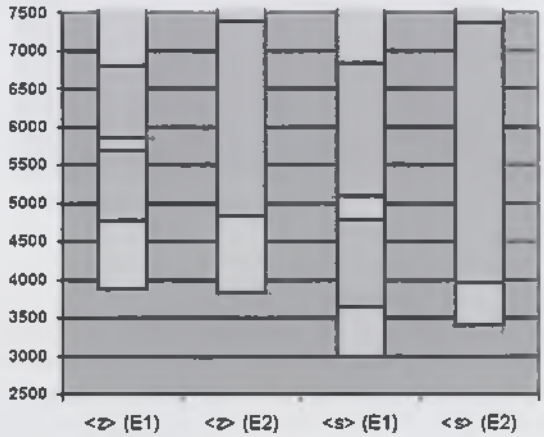
	Primer experimento		Segundo experimento	
	Origen: /θ/	Origen: /s/	Origen: /s/	Origen: /ʂ/
Inicio de la fricación	3.890	2.992	3.835	3.409
Frecuencia inferior	3.176	2.461	3.724	2.717
1ª zona de intensidad	4.780 – 5.700	3.652 – 4.785	4.830 – 6.067	3.960 – 5.441
2ª zona de intensidad	4.520 – 5.320	5.093 – 6.177	5.940 – 7.040	5.256 – 6.928
3ª zona de intensidad	5.856 – 6.816	5.841 – 6.842	6.320 – 7.400	6.400 – 7.360

Las dos [s] seseantes son muy similares: la segunda comienza ligeramente antes, pero sus zonas interiores de energía son más altas. Ambas se diferencian con claridad de las *eses* canónicas. No obstante, éstas no guardan entre sí una gran similitud, puesto que

16 Se emplearon traducciones (*¿cómo se dice en español "josi"?*), preguntas para completar (*¿los nacidos en Suecia son...?*), pictogramas (señal de *stop*) y lectura de palabras. También se introdujeron numerosos ítemes de distracción.

la del segundo experimento resulta notablemente más anterior (sin llegar a los valores de las seseantes). Esto se debe, probablemente, a que el informante no realiza correctamente algunos de los pares, y en ocasiones acerca las realizaciones. No obstante, el experimento demuestra que el informante produce pares mínimos diferenciados por la sibilante, aunque no de una manera absolutamente regular.

Como en las ocasiones anteriores, la gráfica ilustra lo expuesto:



4.2. Experimento complementario: prueba de audición

Por último, se decidió someter el material de la encuesta anterior a una prueba de audición, con el objetivo de averiguar si la distinción producida por el informante se percibe como tal por otro hablante.

Para ello, se aislaron 12 realizaciones de [s] intervocálica junto con sus vocales adyacentes, 6 seseantes y 6 canónicas, procedentes de otros tantos pares mínimos,[17] y se presentaron como un test de discriminación de las sibilantes del euskara a una persona euskaldun y con práctica en las transcripción fonética. Cada ítem se repetía tres veces, y el sujeto debía escribir <z> o <s> en su hoja de respuesta. Los 12 ítemes se repetían (por supuesto aleatoriamente), de modo que el total de estímulos era de 24.

A pesar de la dificultad objetiva de la prueba, el sujeto aceptó las [s] del test como genuinas del euskara, y "transcribió" correctamente 16 de los 24 estímulos. Para la correcta interpretación de los ocho errores cometidos, hay que tener en cuenta lo siguiente:

17 Los pares escogidos fueron: *azada-asada*, *cacé-casé*, *cocer-coser*, *cocido-cosido*, *hacia-asia* y *peces-peses*.

- 6 errores corresponden a 3 estímulos repetidos.
- En dos de ellos, además, el análisis acústico de los pares no había encontrado diferencias apreciables. De hecho, se incluyeron entre los estímulos dos que no servían para establecer la distinción; el error del sujeto, por tanto, debe computarse más bien como acierto.
- Los otros dos errores corresponden a una de las repeticiones de dos estímulos distintos.[18]

En resumen, el test no funcionó correctamente en 3 de los 12 estímulos (4 de 24 ítemes). Estos resultados permiten, por tanto, corroborar la existencia de dos *eses* diferentes en el informante, que parecen ser asimilables a las dos sibilantes del euskara –dorsal y apical–, que el propio informante no posee.

5. CONCLUSIONES

Los dos informantes analizados representan dos realidades distintas en el fenómeno del seseo: la primera responde a las características más presumibles, es decir, una sola realización, de tipo apical; la segunda, por el contrario, abre un panorama más complejo, con dos realizaciones sibilantes: una para la *ese* canónica, compartida además por el euskara, y otra para la *ese* seseante.

A pesar de que la escasez de los datos recogidos obliga a presentarlos como provisionales, la coexistencia de las dos realidades mencionadas podría fundamentar sin embargo una hipótesis explicativa que debe ser de perspectiva diacrónica:

1. En primer lugar, no debe olvidarse que el reajuste castellano que produjo la moderna /θ/ supuso dos procesos distintos: de fricatización (pérdida de africación) y adelantamiento (interdentalización).
2. En un determinado entorno bilingüe, es posible que se produjera únicamente la fricatización, puesto que el resultado de ese proceso se asimilaría a un elemento ya existente en el euskara (aunque, en rigor, no conocemos la distribución diatópica y diacrónica de las sibilantes vascas en las primera etapas del cambio romance).
3. Es posible que el proceso de desfonologización de la oposición *ese* apical-predorsal en áreas de Bizkaia tuviera un reflejo en la evolución del fenómeno romance.
4. Sin embargo, podrían haberse producido desajustes en el ritmo de expansión de la confusión en una y otra lengua.
5. Cabe interpretar los datos recogidos como reflejo vivo de ese desajuste.

---

18 Posteriormente, el sujeto reconoció haber dudado al contestarlos. Aunque es preciso reconocer que también dudó en varios otros casos, que contestó sin embargo correctamente.

En definitiva, el seseo vizcaíno se muestra hoy como un proceso en marcha, con distintas soluciones y en distintos estadios,[19] que exige un estudio detallado, que clarifique no sólo sus implicaciones fonéticas y fonológicas, sino también la variación dialectal y sociolingüística e incluso, en la medida de los posible, la conexión entre los hechos actuales y el plano diacrónico.

No queremos concluir sin poner de relieve los beneficios esperables de una cooperación más estrecha entre los ámbitos de la fonética experimental y la diacrónica.

BIBLIOGRAFÍA

BIZCARRONDO, G.; ELEJABEITIA, A.: *La S en el español de Vizcaya. Análisis acústico*, Bilbao: Universidad de Deusto, 1992.

CARAVEDO, R.: "¿Restos de la distinción /s/ /θ/ en el español del Perú?", *RFE*, LXXII, 1992, pp.639-654.

ECHENIQUE, M.ª T.: *Historia lingüística vasco-románica*, Madrid: Paraninfo, 1997.

GÓMEZ, S.; RAMÍREZ, J.L.: "Notas sobre la lengua de un guipuzcoano emigrado a Indias en el siglo XVIII", *ELUA*, 16, 2003, pp. 331-50.

ISASI, C.: "Seseo vizcaíno en documentos del siglo XVII", *Actas del VI Congreso Internacional de Historia de la Lengua*, Madrid (en prensa).

ISASI, C.: "Seseo, 'ese' sigmática y edición de textos vascos, *Fontes Linguae Vasconum stvdia et docvmenta*, t. XXXI, n.º 81, 1999, pp. 227-240.

ISASI, C.: "Para la historia del castellano de Bilbao", *Actas IV Congreso internacional de Historia de la Lengua*, t. II, Claudio García Turza *et alii* (eds.), Logroño: Universidad de la Rioja, 1998, p. 741-749.

OÑEDERRA, L.: "El español en contacto con otras lenguas: español y vasco", en CANO, R. (coord.): *Historia de la lengua española*, Barcelona: Ariel, 2004, pp. 1103-1115.

TÚRREZ, I.: "Características fonéticas del español en el País Vasco", *Letras de Deusto*, 40, 1988, pp. 65-79.

URRUTIA, H.; ETXEBARRIA, M.; TÚRREZ, I.; DUQUE, J.C.: *Fonética vasca I. Las sibilantes en el Vizcaíno*, Bilbao: Universidad de Deusto, 1988.

19 La convivencia de distintos estadios en los procesos de evolución de las sibilantes del español no es extraña en otras áreas. Véase, por ejemplo, CARAVEDO (1992).

FORMACIÓN Y RENDIMIENTO DE LAS SECUENCIAS VOCÁLICAS EN EUSKARA

ALEXANDER IRIBAR, ITZIAR TÚRREZ
Universidad de Deusto

Las secuencias vocálicas han sido ampliamente tratadas desde el punto de vista acústico en ETXEBARRIA-IRIBAR-TÚRREZ-URRUTIA (*1998*),[1] estudio en el que se analizaba y procesaba un completo material correspondiente a todas las variedades lingüísticas vascas, recogido mediante encuestas de pregunta-respuesta.

En esta ocasión, desde una perspectiva funcional, se retoma el análisis de las secuencias vocálicas con otro objetivo: conocer su rendimiento en el habla continua y formular, más adelante, las reglas de formación que subyacen en su articulación. En definitiva, se pretende conocer el uso que hacen los euskaldunes de las secuencias vocálicas y verificar si, tal como nuestra hipótesis propone, se observan diferencias predecibles por el cambio de registro,[2] tales como la variación en el número y distribución de las secuencias, la aparición de secuencias secundarias y la tendencia a la relajación (reducción silábica).

El trabajo se basa en el análisis de manifestaciones producidas en habla continua, en un registro no formal, perteneciente al corpus oral denominado Euskalkiz Euskalki, recogido por R.M. Pagola, accesible en FonAtari (http://www.fonatari.org) y que abarca todos los dialectos vascos. El corpus recogido consta de 700 unidades ya que se ha procedido a seleccionar 100 secuencias por cada uno de los siete informantes cada uno de los cuales representa a una variedad del euskara. Se tienen en cuenta las siguientes variables: secuencias homosilábicas vs. heterosilábicas, crecientes vs. decrecientes y primarias vs. secundarias.

---

1 A partir de ahora, este trabajo se citará como SVE.

2 Ciertas diferencias vocálicas asociadas al cambio de registro se examinaban en IRIBAR-TÚRREZ (*2001*).

1. REPERTORIO DE SECUENCIAS

En primer lugar, la siguiente tabla recoge las secuencias producidas en cada dialecto:

		Dialecto							Total
		B	G	AN	L	BNoc	BNor	S	
Secuencia	ie	12	11	6	10	13	4	8	64
	ia	10	4	11	15	15	20	16	91
	io	2	1	4	0	5	2	3	17
	iu	1	0	2	0	2	1	1	7
	ei	3	6	16	4	0	9	7	45
	ea	3	14	1	6	17	15	8	64
	eo	0	7	1	0	1	2	1	12
	eu	10	7	0	0	0	0	1	18
	ai	13	12	11	25	21	22	30	134
	ae	1	2	1	0	2	0	0	6
	ao	4	2	3	0	1	1	0	11
	au	13	6	13	7	3	10	1	53
	oi	3	3	5	6	5	2	8	32
	oe	0	7	0	0	1	0	1	9
	oa	15	2	11	2	6	9	6	51
	ou	0	5	0	0	0	0	0	5
	ui	0	0	0	2	0	1	2	5
	ue	7	7	5	11	3	1	0	34
	ua	3	4	9	12	5	1	7	41
	uo	0	0	1	0	0	0	0	1
Total		100	100	100	100	100	100	100	700

Tabla 1

Cabe señalar que solamente ocho secuencias (/i+e/, /i+a/, /e+a/, /a+i/, /a+u/, /o+i/, /o+a/ y /u+a/) son generales, aunque otras tres (/i+o/, /e+i/ y /u+e/) casi lo son, puesto que dejan de registrarse en una variedad.

Las 10 secuencias más frecuentes suponen el 87% de los casos.[3] Ninguna de los otras 10 alcanza una frecuencia apreciable.

Como era de esperar, las secuencias anteriores son notablemente más frecuentes que las posteriores: 57,7 % y 23,1 %, respectivamente.[4]

3 Dos de ellas son no generales (/e+i/ y /u+e/). Las formadas por /a+i/ e /i+a/ suponen ya un tercio del total.

4 Las secuencias antero-posteriores son el 19,1 %.

Al comparar estos resultados con los de SVE, se observa que:

1. Se amplia el repertorio de secuencias, ya que se mantienen las ya registradas (a excepción de la producidas en suletino con el fonema /y/) y aparecen nuevas (aunque su frecuencia, en ocasiones, es muy baja).
2. Las de carácter general se mantienen a excepción de /e+i/, que deja de registrarse en un dialecto.
3. La secuencia /i+e/, que en SVE no se registraba en las variedades bajo navarras –aunque en las otras alcanzaba una frecuencia considerable–, ahora está representada en todos los dialectos.

2. VARIABLES CRECIENTE/DECRECIENTE Y HOMOSILÁBICO/ HETEROSILÁBICO

2.1. Distribución creciente / decreciente

La tabla siguiente muestra los resultados obtenidos en cada dialecto:

	B	G	AN	L	BNoc	BNor	S	TOTAL
Crecientes	52	43	47	56	65	52	48	51,9 %
Decrecientes	47	45	49	42	31	44	47	43,6 %
Homorgánicas	1	12	4	2	4	4	5	4,6 %

Tabla 2

Algo más de la mitad de las secuencias son crecientes. Si se atiende únicamente a las secuencias generales, los resultados de la distribución no varían sensiblemente: aumentan ligeramente tanto las crecientes (54,3%) como las decrecientes (45,7%), puesto que no se registran secuencias homorgánicas generales.[5]

La relación de las secuencias decrecientes con respecto a las crecientes puede examinarse en la tabla siguiente:

	B	G	AN	L	Bnoc	BNor	S	TOTAL
%	90,38	104,65	104,25	75,00	47,69	84,61	97,91	84,00

Tabla 3

2.2. Distribución diptongo / hiato

La distribución entre diptongo e hiato es muy similar en las distintas variedades dialectales. En todos los casos abundan más los primeros, en un porcentaje que oscila entre el 67% y el 75%.[6] Los datos completos pueden verse a continuación:

5 Para el tratamiento de las secuencia homorgánicas, véase TXILLARDEGI (*1982*).

6 La proporción favorable a los diptongos es ligeramente mayor únicamente en las secuencias generales.

	B	G	AN	L	BNoc	BNor	S	TOTAL
Diptongos	72	68	67	75	67	75	69	72,3
Hiatos	28	32	33	25	33	25	31	27,7

Tabla 4

Estos datos permiten señalar otra variación con respecto a SVE, puesto que en ese trabajo el porcentaje de diptongos-hiatos era de 60% frente a 40%. Nos encontramos, por tanto, con que el menor grado de formalidad favorece la reducción silábica, y la consiguiente diptongación.

2.3. Distribución creciente/decreciente vs. diptongo/hiato

Como se muestra en la tabla correspondiente, son diptongos un poco más de la mitad de las secuencias crecientes, los 2/3 de las homorgánicas y la inmensa mayoría de las decrecientes. Por otra parte, un poco más de la mitad de los diptongos son decrecientes, pero los 4/5 de los hiatos son crecientes:[7]

	Diptongo		Hiato		
Creciente	56,5	40,5	43,5	81,4	100
Decreciente	91.8	55,3	8.2	12,9	100
Homorgánico	65,6	4,2	34,4	5,7	100
		100		100	

Tabla 5

Mientras las secuencias decrecientes mantienen una distribución silábica similar a la de SVE, las crecientes se diferencian por un incremento considerable de la diptongación. Así, la esperada reducción silábica hace que en el habla continua la proporción diptongo-hiato pase de 15-85% a 56-43%.

3. VARIABLE PRIMARIO / SECUNDARIO

En general los porcentajes presentan un ligero predominio de las secuencias primarias. En dos de los dialectos (labortano y bajo navarro occidental), la preponderancia de las secuencias primarias es aún mayor. Destaca, como excepción, el guipuzcoano, cu-

7 Ha sido frecuente hablar de la resistencia del sistema a la formación de diptongos crecientes y referirse a ellos como una producción de carácter esporádico. Véase, entre otros, los trabajos de AZKUE (*1919*), GAVEL (*1920*), MOUTARD (*1975-76*), TXILLARDEGI (*1982*) y PAGOLA (*1992*).

yas tres cuartas partes de las secuencias son secundarias. Los porcentajes aparecen en la tabla siguiente:

	B	G	AN	L	BNoc	BNor	S	TOTAL
Primario	51	27	53	66	68	58	45	52,6
Secundario	49	73	47	34	32	42	55	47,4

Tabla 6

3.1. Distribución primario/secundario vs. diptongo/hiato

De mayor interés resulta el cruce de esta variable con la de homosilábico/heterosilábico. En primer lugar, se recogen los datos en la siguiente tabla:

	Diptongo		Hiato		
Primaria	82,1	59,7	17,9	34,0	100
Secundaria	61,4	40,3	38,6	66,0	100
		100		100	

Tabla 7

Todas las secuencias son mayoritariamente diptongos, pero la tendencia a diptongar es mayor en las primarias (4/5) que en las secundarias (2/3). Los diptongos son más primarios que secundarios (2/3), mientras que los hiatos son más secundarios que primarios, en la misma proporción.[8] En las secuencias secundarias aumenta la proporción de hiatos, más en las secuencias decrecientes que en las crecientes.

En resumen, los diptongos son más frecuentemente primarios, mientras que los hiatos lo son secundarios, en ambos casos en la misma proporción (3/2).

3.2. Origen de las secuencias secundarias

El origen de estas secuencias se debe a tres causas principales: el cambio de timbre de una secuencia preexistente (32,3% de los casos), la elisión de algún elemento (40,7%) y los fenómenos de fonosintaxis (26, 9%).

Los primeros, vinculados al fenómeno de armonía vocálica, han sido abundantemente mencionados.[9]

Los producidos por elisión presentan una clara influencia de la variable creciente/decreciente ya que, mientras que el 83% de los decrecientes son diptongos, el 54% de

8 Sobre el carácter secundario de los hiatos, véase MICHELENA (*1977*).

9 Sobre las modificaciones del timbre y su relación con el fenómeno de armonía vocálica véase, entre otros, los trabajos de NAVARRO TOMÁS (*1923*), DE RIJK (*1970*), TXILLARDEGI (*1982*) y HUALDE (*1988*).

los crecientes son hiatos. Es decir: cuando se produce una secuencia por elisión, si dicha secuencia es creciente tiende a mantener su estructura silábica (hiato); por el contrario, si es decreciente, tiende -más- a la reducción silábica (diptongo):

Los hiatos producidos por fonosintaxis refuerzan la conclusión anterior al reproducir el mismo comportamiento.

La siguiente tabla muestra los resultados del cruce de variables para las secuencias generales y para las secundarias:

	Todas las secuencias		Secuencias secundarias	
	Diptongo	Hiato	Diptongo	Hiato
Creciente	56,5 40,5	43,5 81,4	45,9 40,7	54,1 76,6
Decreciente	91.8 55,3	8,2 12,9	83,7 50,5	16,3 15,6
Homorgánico	65,6 4,2	34,4 5,7	64,3 8,8	35,7 7,8

4. CONCLUSIÓN

El trabajo permite, por un lado, describir el uso de las secuencias en el habla continua y, por otro, establecer las diferencias con respecto al modelo más formal de SVE. Se consignan a continuación, de manera esquemática, las ideas más importantes manejadas en las líneas precedentes:

a) Descripción general de las secuencias vocálicas:
   - Se obtiene el repertorio de las secuencias vocálicas en habla continua.
   - Las secuencias crecientes son algo más frecuentes que las decrecientes. No aparecen homorgánicas de carácter general.
   - La mayoría de las secuencias son anteriores.
   - Mayor número de diptongos que de hiatos (75/25%), con una aparición homogénea, sin grandes diferencias dialectales.
   - Son diptongos más de la mitad de las secuencias crecientes, pero la inmensa mayoría de las decrecientes. Del mismo modo, algo más de la mitad de los diptongos son decrecientes, pero la inmensa mayoría de los hiatos son crecientes.
   - Ligero predominio de las secuencias primarias. Los diptongos son más primarios que secundarios; los hiatos son, en la misma proporción, más secundarios.
   - El origen de las secuencias secundarias se distribuye, proporcionalmente, entre: cambio de timbre, elisión y fonosintaxis.
   - Las producidas por elisión y fonosintaxis confirman la importancia de la variable creciente/decreciente. En ellas, además, se advierte una tendencia favorable al mantenimiento silábico (hiato).

b) Diferencias con respecto a SVE:
- Se amplia el repertorio de secuencias. Se mantienen las ya registradas y se introducen nuevas.
- Se produce un considerable aumento de la diptongación:
  - La relación diptongo / hiato pasa de 60/40% a 75/25%.
  - La variación no afecta a las secuencias decrecientes, que mantienen una distribución similar.
  - La variación producida en las crecientes hace que de una relación 15/85% se pase a una 56/43%.

Lo anterior no deja de tener un carácter provisional, puesto que este trabajo es una primera aproximación al estudio del comportamiento de las secuencias vocálicas en habla continua. En todo caso, faltaría, por lo menos, examinar el timbre vocálico e intentar la formulación de reglas de formación de diptongos e hiatos.

En definitiva, el trabajo confirma la hipótesis inicial. Es evidente que el grado de formalidad del registro conlleva cambios en el comportamiento de los hablantes, tal y como confirman las variaciones reseñadas. Habida cuenta de que la mayoría de las descripciones realizadas han estudiado los registros más formales, es conveniente que, en la línea potenciada en los últimos años, se dedique mayor atención a los registros menos formales. Su estudio, además de acercarnos a una articulación lingüística *más real*, nos permite completar las valiosas descripciones precedentes.

BIBLIOGRAFÍA

AZKUE, R.Mª: "Fonética vasca", *Primer Congreso de Estudios Vascos*, Bilbao, 1919-1920, pp. 456-480.

DE RIJK: "Vowel interaction in Bizcayan Basque", *ASJU*, 1972, pp. 149-169.

ELEJABEITIA, A.; IRIBAR, A.; PAGOLA, R.M.: *FonAtari* (on line), Universidad de Deusto, 2001, actualizado 10-2005, [consulta: 5-10-2005]. Disponible en: http://www.fonatari.org.

ETXEBARRIA, M.; IRIBAR, A.; TÚRREZ, I.; URRUTIA, H.: *Fonética vasca 5: las secuencias vocálicas en euskera*, Bilbao, FBV, 1998.

IRIBAR, A.; TÚRREZ, I.: "Algunos factores de variación vocálica", *Letras de Deusto*, XXXI, 92, 2001, pp.229-241.

GAVEL, H.: "Eléments de phonétique basque", *RIEV*, XII, monográfico, 1920, pp. 132-179.

HUALDE, J.I.: *A Lexical Phonology of Basque*, University of Southern California, 1988.

MICHELENA, L.: *Fonética histórica vasca*, San Sebastián, Excma. Diputación Foral de Gipuzkoa, 1977.

MOUTARD, N.: "Etude phonologique sur les dialectes basques-I, II, III", *FLV*, VII, 1975-76, n.º 19, pp. 5-24; n.º 20, pp. 141-189; n.º 22, pp. 9-54.

NAVARRO TOMÁS, T.: "Observaciones fonéticas sobre el vascuence de Guernica", *Tercer Congreso de Estudios Vascos*, San Sebastián, 1923, pp. 49-56.

PAGOLA, R.M.: *Euskal fonetika Nafarroan*, 2 t., Pamplona, Gobierno de Navarra, Departamento de Educación y Cultura, 1992.

TXILLARDEGI: *Euskal Fonologia*, San Sebastián, Ediciones Vascas, 1982.

CREAKY VOICE IN SLOVENE

PETER JURGEC
Scientific Research Centre of the Slovenian Academy of Sciences and Arts, Novi trg 2, 1000 Ljubljana, Slovenia

1. INTRODUCTION

Articulatorily, creaky voice is a non-modal phonation type, usually characterized by laryngeal configuration of inward pressed arytenoid cartilages, so that only anterior portions of glottal folds vibrate; glottal adduction is increased, and longitudinal tension is lower, in comparison to modal voice. Subglottal pressure is lower as well, and both speed and open quotients (cf. Ladefoged 1971: 14, Murry 1971, Laver 1980: 109ff., Ní Chasaide & Gobl 1997: 444–451, Berry 2001, Hanson *et al.* 2001, Gordon & Ladefoged 2001). Acoustically, low F0 has been documented, irregular periods and an increased jitter. A positive spectral tilt has been studied thoroughly; normally, amplitudes of F0 are compared to second harmonic (H2) and harmonic closest to F1 (A1), complemented by corresponding harmonics A2 and A3 (Klatt & Klatt 1990, Hanson 1997, Ní Chasaide & Gobl 1997, Hanson & Chuang 1999, Hanson *et al.* 2001, Gordon & Ladefoged 2001). Perceptually, creaky voice was characterized as »rapid series of taps, like a stick being run along a railing« (Catford 1964: 32).

Regardless of seemingly unproblematic descriptions above, authors have been reluctant to use one uniform terminology. Many partially or completely overlapping terms to denote creaky voice phonation exist: glottal fry, vocal fry, vocal cry, pulse phonation, creak, laryngealization and glottalizaton (not as secondary articulation feature). These differences can be observed from the viewpoint of Ladefoged's larynx feature model. According to Ladefoged (1971) and Ladefoged & Gordon (2001), there is a phonation continuum form completely open (as in voiceless phonation) to completely closed glottis (as in glottal stop). Modal voice being at the central point, creaky voice has somewhat more closed glottal configuration, while being more open than glottal stop at the same time. Following Ladefoged (*l.c.*), the part of the continuum from modal voice to glottal stop (but exclusive of both), is discussed further in the text. This window could easily be

named laryngealization as well.

However, other classifications exist. For example, Laver (1994: 199) acknowledges six basic phonation types, classified in three basic groups (whisper, voicing and voicelessness). On the other hand, more recent studies of Edmonson & Esling (2005), and Esling & Harris (2005) propose actually six independent laryngeal valves, glottal fold adduction being only one of them. Others are ventricular incursion, sphincteric compression forwards and upwards, epiglotto-pharyngeal constriction, laryngeal raising and lowering of the larynx, and pharyngeal narrowing (*l.c.*).

In the present study, creaky voice in Slovene is studied with particular attention to its distribution and function with reference to natural languages in general. Slovene (or Slovenian) is a south Slavic language, spoken by approximately 2 million speakers of the Republic of Slovenia, and by minorities in adjoining areas of the neighboring countries (Italy, Austria, Hungary and Croatia). There are 48 distinct dialects and 7 dialect groups, of which the central dialects of Gorenjska and Dolenjska regions are relevant to the contemporary standardization process, as realized in educated speech of Ljubljana.

In standard Slovene, stress is distinctive. According to the traditional grammar (Toporišič 2000), vowel quantity is also contrastive in word-final stressed vowels. More recent studies of Srebot Rejec (1988, 1998) and Petek *et al.* (1996) argue that there are no significant durational distinctions in contemporary standard Slovene as spoken in Ljubljana any more; all stressed vowels are long. Data on formant frequencies confirm that only /a/ retains significant quantitive distinction (Jurgec, forthcoming-b). In respect to tone, Slovene dialects are either tonal or non-tonal; this is retained in the standard form and so, both varieties are permitted. In the tonal standard Slovene there are two lexical tones (tonemes), namely *acute* and *circumflex*. Traditionally (Vodušek 1961, Toporišič 1967, 1968) acute has been described as rising, or low on the tonic and high on the post-tonic syllable, and circumflex *vice versa*. Morphologically related minimal pairs are abundant, but only 100 or less not lexically related minimal pairs exist. Such examples are listed in Table 1.

Orthographic form	Acute	Circumflex
MONOSYLLABLES		
pot [poːt]	'path'	'sweat'
vrat [wraːt]	'neck'	'door, *gen. pl.*'
DISYLLABLES		
kila [ˈkʰiːla]	'hernia'	'kg'
kura [ˈkʰuːra]	'chicken'	'cure, treatment'
TRISYLLABLES		
šalca [ˈʃaːlit͡sa]	'joke, *dimm.*'	'cup, *coll.*'
šibica [ˈʃiːbit͡sa]	'rod, *dimm.*'	'match, *coll.*'

Tab. 1. Standard Slovene minimal pairs in tone.

Creaky voice has not been mentioned in connection to standard Slovene until very recently (Jurgec 2005b). However, there have been previous studies of Carinthian dialects of Rož and Podjuna, which have [ʔ] as a reflex of Proto-Slavic */k/, and it is in free variation with creaky voice either word-initially or word-internally (Neweklowsky 1970, Priestly 1976, 1980). In connection to standard Slovene, the present author found the same variation in vowels word-initially or as hiatus resolution (Jurgec 2004, 2005b). These phenomena are discussed further in this text (Section 3).

2. METHOD

The collection of three spoken corpora has been selected to evaluate the extent of creaky voice in Slovene. The main corpus consists of 275 mono-, di- and trisyllables, compiled according to segmental and suprasegmental criteria. 10 speakers, representative of standard Slovene (5 per gender, geographically diverse, 5 of them tonal) were selected and instructed to read the words from the computer screen in citation form. The words appeared in random order, each twice non-consecutively. The recordings took place in the studios of the Department of phonetics in Zagreb (Croatia) and at Radio Slovenia in Ljubljana. Digitalized under standard conditions, 204 minutes of recordings (or 5,500 words in total) were examined by ear and confirmed by means of digital acoustic analysis (*Praat* program). The data collected were later averaged and analyzed statistically.

Two supporting corpora were used principally to corroborate the findings in the main corpus. The first supporting corpus consists of five live radio broadcastings (306 min in total) *Studio ob sedemnajstih*, a quality talk show, transmitted by the national radio station (Radio Slovenia). 39 speakers from various parts of Slovenia and of different social origin spoke standard Slovene. The recordings were examined by ear.

The second supporting corpus consists of selection from longer (approximately 4 and 3.5 hours, respectively) guided fieldwork interviews of a northwestern dialect of Zilja (Carinthian dialect group, tonal) as spoken in Kanalska Dolina Valley (Italy), see Jurgec, forthcoming-a for a detailed phonetic analysis. There were two speakers, one female and one male, and the total duration of speech analyzed was 60 minutes (approximately 30 minutes per speaker). The ear examination was reconfirmed acoustically.

3. RESULTS

In the main corpus, 585 words (or 11.6%) had some form of creaky voice in one or more of the segments. In the supporting corpus #2, 61 and 19 cases of creaky voice were found, whereas in the supporting corpus #1 these were numerous (>1,000). However, individual speakers differ considerably in degree of creaky voice present. This was also a conclusion of the previous studies for other languages (Umeda 1978, Streeter 1978, Klatt & Klatt 1990, Holmberg *et al.* 1996, Dilley *et al.* 1996, Hanson 1997, Redi & Shattuck - Hufnagel 2001). However, neither speakers' gender (m – male, f – female)

nor the variety of standard Slovene (t – tonal, n – non-tonal) is statistically significant (p =.453, p =.714) in respect to creaky voice. In Table 2, frequency of creaky voice for individual speakers of the main corpus can be found.

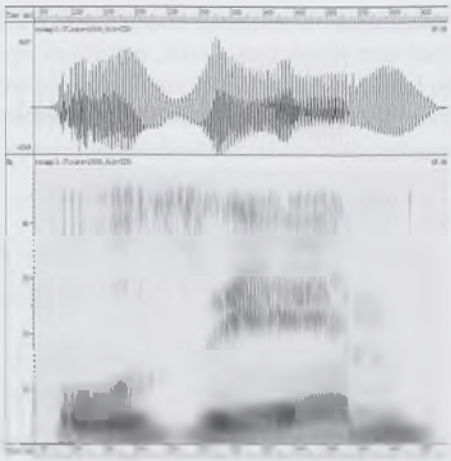
Speaker	No. of instances	Percentile
01fn	31	6.1
02fn	118	23.4
03ft	69	13.7
04ft	100	19.8
05ft	24	4.8
06mn	18	3.6
07mn	47	9.3
08mt	66	13.1
09mt	9	1.8
10mn	103	20.4
Total	585	11.6

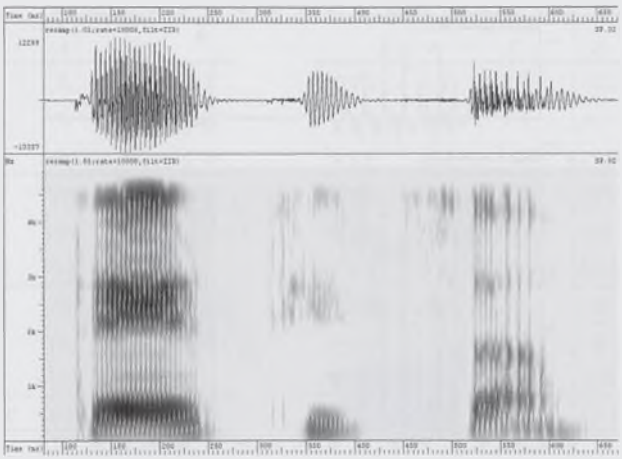
Tab. 2. Creaky voice ratio for individual speakers.

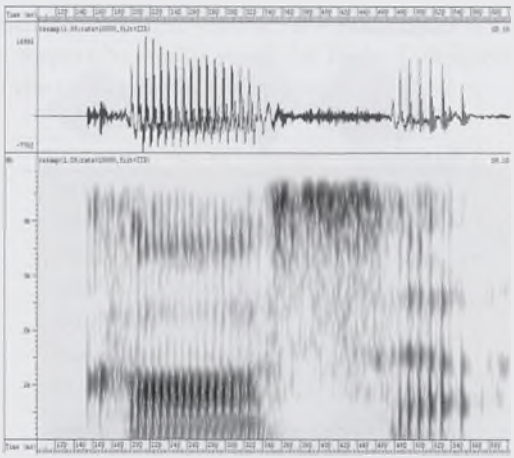
There were three distinctive types of creaky voice in respect to word-position, or function (in a more general sense, as evident from the discussion below).

Type I denotes a boundary type of creaky voice. It is present as an allophonic realization of [ʔ] in word-initial position (to satisfy ONSET constraint), or between two vowels word-medially (to satisfy NOHIATUS constraint). In total, 108 instances of type I creaky voice were detected in the main corpus, and in 30.2% of all words beginning on phonological vowel, either [ʔ] or creaky voice was present. This kind of variation has already been observed in Roz and Podjuna dialects, first noted by Neweklowsky (1970) and later studied by Priestly (1976, 1980), who concludes that "the variation between glottal stop and laryngealization in the Sele Fara dialect depends on speech tempo, and also (directly or indirectly) on the rhythm of the sentence" (1976: 272–273). This is comparable to data for standard German (Kohler 1994) and standard Slovene (Jurgec 2004, 2005b). An example of this type of creaky voice is Figure 1a.

In both supporting corpora, this type I creaky voice is frequently attested. In the supporting corpus #1, numerous compound words like *makroekonomski* 'macro-economic' or *severnoevropski* 'northern European' are frequently pronounced with creaky phase in hiatus. Furthermore, creaky voice is present on a prosodic boundary, most prominently on the beginning of a prosodic unit (sentence, utterance), normally with high sonority segment. However, this phenomenon is highly speaker-dependent; while some speakers in the supporting corpus #1 almost always insert glottal or creaky voiced segment word-initially, other almost never do.







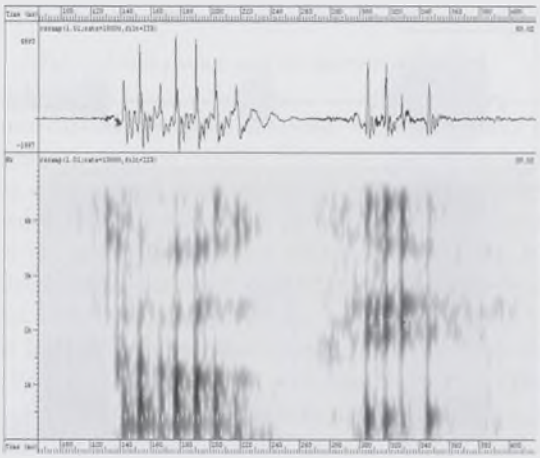


Fig. 1a–1d. Different types of creaky voice. 1a (above): *OZN* [ˌʔozeˈɛ̃n] 'UN'. 1b: *tetica* [ˈtɛːtit͡sa] 'auntie'. 1c: *kosa* [ˈkɔːsa] 'scythe'. 1d: *okej* C [ˈɔ́ːˌkéi] 'OK'.

Creaky voice of type II is word-internal and most prominent in word-final position. In the main corpus, 385 cases of this type were detected (examples are in Figures 1b and 1c). The present author initial hypothesis, based on previous research on other languages (see below) was, that creaky voice is significantly affected by suprasegmental (or prosodic) variables. A detailed analysis of the main corpus was conducted. These factors were considered: stress, quantity, tone and length of the word.

As regards the latter, no monosyllables were found with creaky voice of this type, but no conclusion can be drawn from di- and trisyllables (see Tables 3a and 3b for indi-

vidual combinations). Moreover, stressed syllables are significantly less frequently creaky voiced (see similar findings in Ogden 2001), and this is true for short (stressed) vowels as well. In Slovene, short stressed vowels are limited to word-final position and so, only a few (<5) were detected. Furthermore, the data from the main corpus suggest post-stressed syllables are decisively more prone to creaky voice than both stressed and pre-stressed.

	Disyllables				
	Tonal		Non-tonal		σ
	A	C	A	C	
ˈVV	3	3	1	5	12
ˈVV	9	21	25	22	77
VˈV	0	2	1	1	4
VˈV	3	3	1	0	7
σ	15	29	28	28	100

	Disyllables				
	Tonal		Non-tonal		σ
	A	C	A	C	
ˈVV	1.9	1.9	.63	3.1	1.9
ˈVV	5.6	13.1	15.7	13.8	12.0
VˈV	.00	1.4	1.2	1.2	.91
VˈV	3.8	2.1	1.2	.00	1.6
σ	6.8	9.7	12.7	9.3	18.5

	Trisyllables				
	Tonal		Non-tonal		Σ
	A	C	A	C	
ˈVVV	3	4	0	5	12
ˈVVV	2	2	6	7	17
ˈVVV	37	29	40	34	140
VˈVV	1	2	1	1	5
VˈVV	6	1	1	2	10
VˈVV	13	27	34	21	95
VVˈV	0	0	0	0	0
VVˈV	0	2	0	1	3
VVˈV	0	1	1	1	3
σ	62	68	83	72	285

	Trisyllables				
	Tonal		Non-tonal		σ
	A	C	A	C	
ˈVVV	1.8	2.5	.00	3.3	1.9
ˈVVV	1.3	1.3	4.0	4.7	2.7
ˈVVV	23.1	18.1	25.0	22.7	22.6
VˈVV	.63	1.3	.63	.63	.78
VˈVV	3.8	.63	.63	1.3	1.6
VˈVV	8.1	16.9	21.3	13.1	14.8
VVˈV	.00	.00	.00	.00	.00
VVˈV	.00	1.3	0	.63	.71
VVˈV	.00	.63	2.0	.63	.71
σ	16.8	14.5	22.4	15.3	16.9

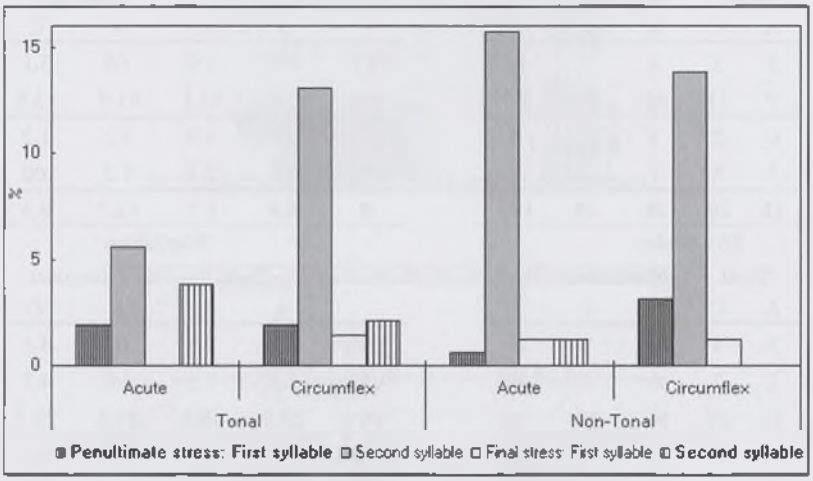
Tab. 3a–3b. Type II creaky voice in the main corpus according to the suprasegmentals (A – acute tone, C – circumflex tone). 3a: Absolute values. 3b: Percentiles.

In connection to tone, there is statistically significant ($p<.001$) difference between acute and circumflex tone in the post-stressed syllables: post-tonic acute is less frequently creaky voiced, for tonal speakers only. This is an expected phenomenon, as low F_0 is frequently associated with creaky voice cross-linguistically.

On the other hand, we cannot explain differences in lexical tones for both tonal and non-tonal speakers; circumflex tone seems to be somewhat less affected by creaky voice than the acute tone (Figures 2a and 2b). Perhaps, as it has been stipulated in more traditional frames (e.g. Toporišič 1967, 2000), circumflex is indeed shorter than acute, and would be more probable to experience creaky voice. No sufficient corroborating acoustic phonetic data exist yet, however.

In the supporting corpus #2 however, creaky voice was more frequent in post-stressed syllables of the acute tone, but it is also attested in the circumflex, so no definite conclusion can be drawn.

In both supporting corpora, this type of creaky voice is considerably more frequent at the end of prosodic units, where F_0 is low by default, and creaky voice has a demarcative function (cf. Umeda 1978, Dilley et al. 1996, Redi & Shattuck - Hufnagel 2001). This is highly speaker dependent as well.



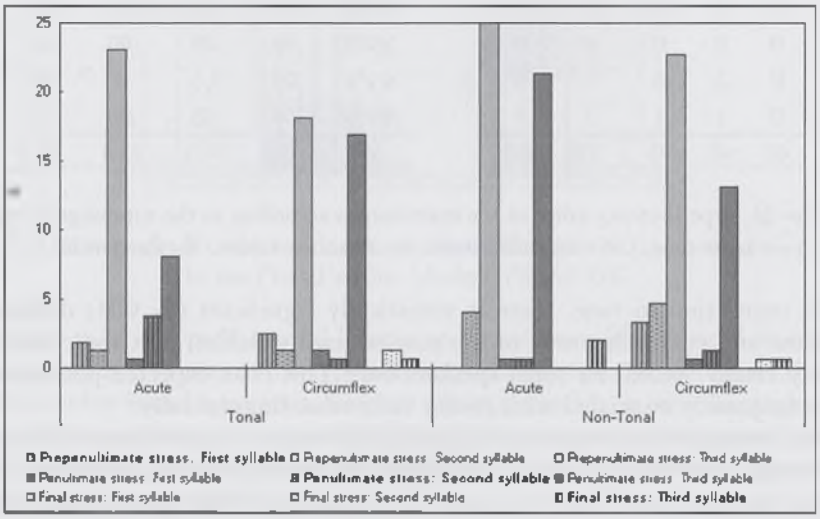


Fig. 2a–2b. Type II creaky voice in the main corpus according to the suprasegmentals (in percentiles). 2a (above): Disyllables. 2b: Trisyllables.

Type III creaky voice applies to whole words, utterances, or their semantically independent fragments. In the main corpus, creaky voice is found in words not predicted by the experiment, e.g. *pardon* 'pardon me', *okej* 'all right', *v redu* 'all right', *ja* 'yes' (see Figure 1d for an example spectrogram and waveform). This kind of paralinguistic use has also been confirmed in the supporting corpora.

4. DISCUSSION AND CONCLUSION

All three types of creaky voice are attested in languages of the world. The most prominent are the phonemic contrasts in creaky vowels of Jalapa Mazatec or San Lucas Quiaviní Zapotec. Creaky voice is also found in sonorants (of Kwakw'ala, Hupa, Montana and Columbian Salish), while obstruents are fully phonemic quite rarely; examples include Chadic languages, e.g. Hausa. See Ladefoged 1971, Laver 1994, Ladefoged & Maddieson 1996, and Gordon & Ladefoged 2001, for a review. It is very improbable to find these contrasts in contemporary Slovene, although breathy voice might exist in the dialect of Rezija (Resia), but acoustic studies so far are fairly inconclusive.

Phonation contrasts can also be suprasegmental. The most well known example is Danish *stød* (Fischer - Jørgensen 1989, Grønnum & Basbøll 2001), while in many other tonal languages creaky voice can be a accompanying feature to a certain tone, e.g. in Chinese (Belotel - Grenié in Grenié 2004), Hmong (Huffman 1987), or Latvian *lauzta intonacija*, Swedish *Accent 1* (Laver 1994: 464) and Croatian *dugosilazni naglasak* (Elen Mari Pletikos, p.c.). In Slovene, this type has been confirmed (so called type II).

A more frequent use of creaky voice cross-linguistically is the allophonic, either replacing a certain consonant phoneme -usually a stop, for example in German (Kohler 1994) or American English (Patterson in Connine 2001)-, or preceding a vowel (German, Slovene).

Creaky voice is also attested as a boundary marker in several languages. Different variables, including word frequency, boundary level, pitch-accent, speed, rhythm, pauses and segmental variables were studied, in English (Allen 1970, Umeda 1978, Dilley *et al.* 1996, Redi & Shattuck - Hufnagel 2001), Swedish, Czech, Finnish, Serbian/Croatian (Lehiste 1965, 1970; Ogden 2001), Chinese (Belotel - Grenié & Grenié 2004) or Huaulta Mazatec (Ladefoged and Maddieson 1996: 74). In Slovene, further studies of this type creaky voice are needed.

Pragmatic variables do influence creaky voice distribution as well, and are common in English (Laver 1980: 126) or other European languages. Paralinguistic use is very diverse in its effect, see *l.c.* In Slovene, embarrassment as well as the end of utterance frequently correlate with creaky voice (the main and the supporting corpus #1). Sometimes, creaky voice can be used ironically.

Sociolinguistically, creaky voice has been detected in Edinburgh English (Ní Chasaide & Gobl 1997: 457) and Copenhagen Danish (Laver 1980: 196). The present investigation could not address this question in full. Female vs. male voice characteristics -male voices being characterized as more creaky than female, on average (Holmerg *et al.* 1988,

Klatt & Klatt 1990, Hanson 1997, Hanson & Chuang 1999, Redi & Shattuck - Hufnagel 2001)- fall in this category as well. In Slovene, creaky voice can be found both in very low-pitched male voices or female voices at certain prosodic positions (see above).

Further studies will address boundary effects and distribution as well as other pragmatic and paralinguistic phenomena of non-modal phonation in Slovene.

BIBLIOGRAPHY

ALLEN, J.: "The glottal stop as a junctural correlate in English", *JASA* 40: 1 (1970), p. 47.

BELOTEL - Grenié, Agnès, and GRENIÉ, Michel: "The creaky voice phonation and the organisation of Chinese discourse" [on-line], *International Symposium on Tonal Aspects of Languages:* Emphasis on Tone Languages, Beijing, 28-30 March 2004, 2004. [Consultation: 7th June 2005]. Available at: http://www.lpl.univ-aix.fr/~sprosig/tal2004/tal2004-Beijing/Belotel.pdf.

BERRY, David A.: "Mechanisms of modal and nonmodal phonation", *JournPhon* 29: 4 (2001), p. 431–450.

BLOMGREN, Michael, CHEN, Yang, NG, L. Manwa & GILBERT, Harvey R.: "Acoustic, aerodynamic, physiologic, and perceptual properties of modal and vocal fry registers", *JASA* 103: 5 (2001), p. 2649–2658.

CATFORD, J. C.: "Phonation types". *In Honour of Daniel Jones*. London: Longmans, Green, 1964, p. 26–37.

DILLEY, L., SHATTUCK HUFNAGEL, S., and OSTENDORF, M.: "Glottalization of word-initial vowels as function of prosodic structure", *JournPhon* 24: 4 (1996), p. 423–444.

EDMONDSON, Jerald A. & ESLING, John H., 2005: "The valves of the throat and their functioning in tone, vocal register, and stress: laryngoscopic case studies". Submitted to *Phonology*.

ESLING, John H. & HARRIS, Jimmy G.: "States of the glottis: an articulatory phonetic model based on laryngoscopic observations", Hardcastle, William J. & Mackenzie Beck, Jenet (eds.), *A Figure of Speech: a Festschrift for John Laver*, Mahwah: Lawrence Erlbaum Association, 2005.

FISCHER - JØRGENSEN, Eli: "Phonetic analysis of the stød in Standard Danish", *Phonetica* 46: 1–3 (1989), p. 1–59.

GERRATT, Bruce R., and KREIMAN, Jody: "Toward a taxonomy of nonmodal phonation", *JournPhon* 29: 4 (2001), p. 365–381.

GORDON, Matthew & LADEFOGED, Peter: "Phonation types: a cross-linguistic overview", *JournPhon* 29: 4 (2001), p. 383–406.

GRØNNUM, Nina & BASBØLL, Hans: "Consonant length and morae in Standard Danish", *Phonetica* 58: 4 (2001), p. 230–253.

HANSON, Helen M.: "Glottal characteristics of female speakers: Acoustic correlates", *JASA* 101: 1 (1997), p. 466–481.

HANSON, Helen M., STEVENS, Kenneth N., KUO, Hong-Kwang Jeff, CHEN, Marilyn Y. & SLIFKA, Janet: "Towards models of phonation". *JournPhon* 29: 4 (2001), p. 451–480.

HANSON, Helen M. & CHUANG, Erika S.: "Glottal characteristics of male speakers: Acoustic corelates and comparison with female data", *JASA* 106: 2 (1999), p. 1064–1077.

HOLLIEN, Harry: "On vocal registers", *JournPhon* 2: 1 (1974), p. 125–143.

HOLLIEN, Harry & MICHEL, John F.: "Vocal fry as a phonational register", *JournSpHearRes* 9 (1968), p. 600–604.

HOLMBERG, Eva B., HILLMAN, Robert E. & PERKELL, Joseph S.: "Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal, and loud voice", *JASA* 84: 1 (1988), p. 511–529.

HOLMBERG, Eva B., PERKELL, Joseph S., HILLMAN, Robert E. & GRESS, Carla: "Individual variation in measures of voice", *Phonetica* 50: 1 (1993), p. 30–37.

HOMBERT, Jean-Marie, OHALA, John J., & EWAN, William G.: "Phonetic explanations for the development of tones", *Language* 55: 1 (1979), p. 37–58.

HOUDE, R. A. & HILLENBRAND, J. M.: "The role of voice pitch in the perception of glottal stops", *JASA* 95: supp. 1 (1994), p. 2872.

HUFFMAN, Marie K.: "Measures of phonation type in Hmong", *JASA* 81: 2 (1987), p. 495–504.

JURGEC, Peter: "Antihiatski pojavi v knjižni slovenščini", *Jezikoslovni zapiski* 10: 1 (2004), p. 125–144.

JURGEC, Peter: *Samoglasniški nizi v slovenščini: Fonološko-fonetična analiza*, Ljubljana: Rokus (Slavistična knjižnica, 8), 2005.

JURGEC, Peter: "Fonetični opis govora Ovčje vasi", 17pp. In print (forthcoming-a).

JURGEC, Peter: "Formant frequencies of Standard Slovene vowels", 19pp. Submitted to *Govor* (forthcoming-b).

KLATT, Dennis H. & KLATT, Laura C.: "Analysis, synthesis, and perception of voice quality variations among female and male talkers", *JASA* 87: 2 (1990), p. 820–857.

KOHLER, Klaus J.: "Glottal stops and glottalization in German", *Phonetica* 51/1–3 (1994), p. 38–51.

KREIMAN, Jody K., GERRATT, Bruce R., KEMPSTER, Gail B., ERMAN, Andrew & BERKE, Gerald S.: "Perceptual evaluation of voice quality: Review, tutorial and framework for future research", *JournSpHearRes* 36 (1993), p. 21–40.

LADEFOGED, Peter: *Preliminaries to Linguistic Phonetics*, Chicago *et al.*, The University of Chicago Press, 1971.

LADEFOGED, Peter: "The linguistic use of different phonation types", *Vocal Fold Physiology: Contemporary Research and Clinical Issues*, San Diego: College Hill, 1983, p. 351–360.

LADEFOGED, Peter, MADDIESON, Ian & JACKSON, Michel T. T.: "Investigating phonation types in different languages", *Vocal Physiology: Voice Production. Mechanisms and Functions*, New York, Raven, 1988, p. 297–317.

LADEFOGED, Peter & MADDIESON, Ian: *The Sounds of the World's Languages*, Malden *et al.*, Blackwell Publishers, 1996.

LAVER, John: *The Phonetic Description of Voice Quality*, Cambridge *et al.*, Cambridge University Press (Cambridge studies in linguistics), 1980.

LAVER, John: *Principles of Phonetics*, Cambridge *et al.*: Cambridge University Press (Cambridge textbooks in linguistics), 1994.

LEHISTE, Ilse: "Juncture", *Proceedings of the 5th International Congress of Phonetic Sciences*, Basel: S. Karger, 1965, p. 172–200.

LEHISTE, Ilse: *Suprasegmentals*, Camridge *et al.*: The M. I. T. Press, 1970.

LIENARD, Jean-Sylvain: "Effect of vocal effort on spectral properties of vowels", *JASA* 106: 1 (1999), p. 411–422.

LUCERO, Jorge C.: "The minimum lung pressure to sustain vocal fold oscillation", *JASA* 98: 2 (1995), p. 779–784.

MCGLONE, Robert E. & Shipp, Thomas: "Some physiologic correlates of vocal-fry phonation", *JournSpHearRes* 14 (1971), p. 769–775.

MICHEL, John. F. & HOLLIEN, Harry: "Perceptual differentiation of vocal fry and harshness", *JournSpHearRes* 11 (1968), p. 439–443.

MONSEN, Randall B., ENGEBRETSON, A. Maynard & VEMULA, Rao N.: "Indirect assessment of the contribution of subglotal air pressure and vocal-fold tension to changes of fundamental frequency in English", *JASA* 64: 1 (1978), p. 65–80.

MURRY, Thomas: "Subglottal pressure and airflow measures during vocal fry phonation". *JournSpHearRes* 14 (1971), p. 544–551.

NEWEKLOWSKY, Gerhard: "Der laryngale Verschlußlaut in Slovenischen", *Wiener Slavistisches Jahrbuch* 16 (1970), p. 104–111.

NEWEKLOWSKY, Gerhard: *Slowenische Akzentstudien: Akustistische und linguistiche Untersuchungen am Material Slowenischer Mundarten aus Kärnten mit 46 Abbildungen und 76 Figuren in Text*, Wien: Verlag der Österreichischen Akademie der Wissenschaften, 1973.

NÍ CHASAIDE, Ailbhe & GOBL, Christer: "Voice source variation", *The Handbook of Phonetic Sciences*, Oxford *et al.*: Blackwell Publishers, 1997, p. 427–461.

OGDEN, Richard: "Turn-holding, turn-yielding and laryngeal activity in Finnish talk-in-interaction", *JournIPA* 31: 2 (2001), p. 139–152.

PATTERSON, David & CONNINE, Cynthia M.: "Variant frequency in flap production", *Phonetica* 58: 4 (2001), p. 254–275.

PETEK, Bojan, ŠUŠTARŠIČ, Rastislav & KOMAR, Smiljana: "An acoustic analysis of contemporary vowels of the standard Slovenian language" [on-line], *Proceedings ICSLP 96: Fourth international conference on spoken language processing, October 3–6, 1996, Philadelphia, PA, USA*, Wilmington: University of Delaware, Apllied science and engineering laboratories, Alfred I. du Pont institute, 1996, p. 133–136. [Consultation: 2nd October 2005]. Available at: http://www.asel.udel.edu/icslp/cdrom/vol1/820/a820.pdf.

PRICE, P. J., OSTRENDORF, M., SHATTUCK - HUFNAGEL, S. & FONG, C.: "The use of prosody in syntactic disambiguation", *JASA* 90: 6 (1991), p. 2956–2970.

PRIESTLY, Tom M. S.: "A note on the glottal stop", *Phonetica* 33: 4 (1976), p. 268–274.

PRIESTLY, Tom: "*k + dative in a Carinthian Slovene dialect", *Folia Slavica* 4: 1 (1980), p. 25–34.

PURCELL, E. T., VILLEGAS, G. & YOUNG, S. P.: "A before and after for tonogenesis", *Phonetica* 35: 5 (1978), p. 284–293.

REDI, Laura & SHATTUCK - HUFNAGEL, S.: "Variation in the realization of glottalization in normal speakers", *JournPhon* 29: 4 (2001), p. 407–429.

SINGH, Sadanand & MURRY, Thomas: "Multidimensional classification of normal voice qualities", *JASA* 64: 1 (1978), p. 81–87.

SMITH, Marshall E., BERKE, Gerard S., GERRATT, Bruce R. & KREIMAN, Jody: "Laryngeal paralyses: Theoretical considerations and effects on laryngeal vibration", *JournSpHearRes* 35 (1992), p. 545–554.

SREBOT REJEC, Tatjana: *Word Accent and Vowel Duration in Standard Slovene: An Acoustic and Linguistic Investigation*, München: Otto Sagner (Slavistische Beiträge, 226), 1988.

SREBOT REJEC, Tatjana: "On the vowel system in present-day standard Slovene" [online], *Jezikovne tehnologije za slovenski jezik: Mednarodna multi-konferenca Informacijska družba* [...], Ljubljana: Institut Jožef Stefan, 1998, p. 22–24. [Consultation: 3rd September 2005]. Available at: http://nl.ijs.si/isjt98/zbornik/sdjt98-Srebot.pdf.

SREBOT REJEC, Tatjana: "Ali je današnja knjižna slovenščina še tonematična?", *Razprave II. razreda SAZU*, 17, p. 51–66.

STEVENS, Kenneth N.: "Physics of laringeal behaviour and larynx modes", *Phonetica* 36: 4 (1977), p. 264–279.

STREETER, Lyn: "Acoustic determinants of phrase boundary perception", *JASA* 64: 6 (1978), p. 1582–1592.

ŠVEC, Jan G., SCHUTTE, Harm K. & MILLER, Donald G.: "A subharmonic vibratory pattern in normal vocal folds", *JournSpHearRes* 39 (1996), p. 135–143.

ŠVEC, Jan G., HORÁČEK, Jaromír, ŠRAM, František & VESELY, Jan: "Resonance properties of the vocal folds: *In vivo* laryngoscopic investigation of the externally excited laryngeal vibrations", *JASA* 108: 4 (2000), p. 1397–1407.

TITZE, Ingo R.: "The physics of small-amplitude oscillation of the vocal folds", *JASA* 83: 4 (1988), p. 1536–1552.

TITZE, Ingo R.: "Phonation threshold pressure: A missing link in glottal aerodynamics", *JASA* 91: 5 (1992), p. 2926–2935.

TITZE, Ingo R. & TALKIN, David T.: "A theoretical study of the effects of various laryngeal configurations on the acoustics of phonation", *JASA* 66: 1 (1979), p. 60–74.

TOPORIŠIČ, Jože: "Pojmovanje tonemičnosti slovenskega jezika", *Slavistična revija* 15/1–2 (1967), p. 64–108.

TOPORIŠIČ, Jože: "Liki slovenskih tonemov", *Slavistična revija* 16 (1968), p. 315–393.

TOPORIŠIČ, Jože: *Slovenska slovnica*, Maribor: Obzorja, 2000. First published in 1976.
TRASK, R. L.: *A Dictionary of Phonetics and Phonology*, London et al., Routledge, 1996.
UMEDA, Noriko: "Occurrence of glottal stops in fluent speech", *JASA* 64: 1 (1978), p. 88–94.
VODUŠEK, Bozo: "Grudsatzliche Betrachtungen über den melodischen Verlauf der Wortakzente in den zentralen Slowenichen Mundarten", *Linguistica* 4 (1961), 20–38.
WENDAHL, R. W., MOORE, G. P. & HOLLIEN, H.: "Comments on vocal fry", *Folia phoniatrica* 15 (1963), p. 251–255.
WHITEHEAD, Robert L., METZ, Dale E. & WHITEHEAD Brenda H.: "Vibratory patterns of vocal folds during pulse register phonation", *JASA* 75: 4 (1984), p. 1293–1297.

ESTUDIO COMPARATIVO DE DOS HABLAS ALPINAS EN EL MARCO DE AMPER[1]

JEAN-PIERRE LAI

El objetivo de este estudio es presentar por primera vez una comparación entre dos hablas alpinas que no forman parte de la misma zona dialectal. Se trata de la variedad de Gap (occitano de los Altos Alpes), que ya ha sido objeto de análisis con el método Amper (Lai, 2005), y de la de Niza (llamada nissart). Al igual que el conjunto de hablas occitanas que no conocen los proparoxítonos, la variedad alpina de Gap tiene una estructura acentual que favorece los paroxítonos y los oxítonos, excepto para la variedad de Niza (P. Bec, 1978, p. 28).

Presentamos a continuación las secuencias prosódicas elegidas para las dos variedades.

NIZA	proparoxítono	paroxítono	oxítono
Sustantivos	[ˈpassera]	[tarˈtuga]	
Adjetivos		[perˈduda]	
Sintagmas preposicionales		[de peˈlaska]	[de la ternˈita]

GAP	proparoxítono	paroxítono	oxítono
Sustantivos		[tsaˈbʀeta]	[bwɛˈsuˈna]
Adjetivos		[pøtsuˈneta]	[astʀɔˈpja]
Sintagmas preposicionales		[dˑ ʀuˈmeta]	[dˑ sigoˈje]

Los esquemas siguientes nos darán una primera idea de las evoluciones de las curvas entonativas de las dos variedades con el fin de facilitar la comparación.

1 Tengo que agradecer a Beatriz Hernández la traducción del texto y también a Noelia Gil Peña.

1. PARA NIZA, CON UN PROPAROXÍTONO AL PRINCIPIODE FRASE: ['passera]

1.1. En cuanto a la duración:

En la modalidad declarativa se mantendrán los valores máximos sobre las tónicas de la última expansión en el SN (graf.1), al igual que en Gap, con un alargamiento particular de la nasal tónica del verbo. Pero en Gap, un paroxítono al final de frontera limita las duraciones en torno a los 90 ms.

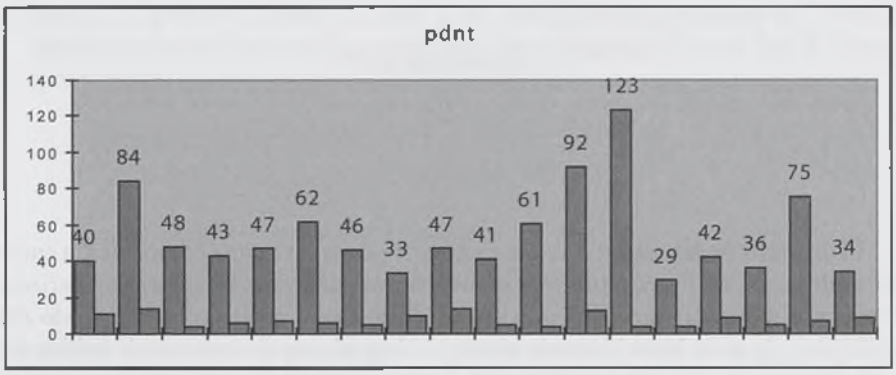


fig.1 Niza [la 'passɛʁa perʁ'duda de la teʁni'ta 'mẽta la taʁ'tuga]

En las interrogativas con expansiones en el SN, las duraciones máximas se observan en la primera y la última tónica del enunciado (graf.3), excepto cuando el adjetivo o el sintagma preposicional es un oxítono ([terni'ta]) (graf.2). En cambio, en Gap, solamente la última tónica del enunciado presenta una duración elevada, además de que la presencia de dos oxítonos implica valores claramente más bajos (graf.2bis).

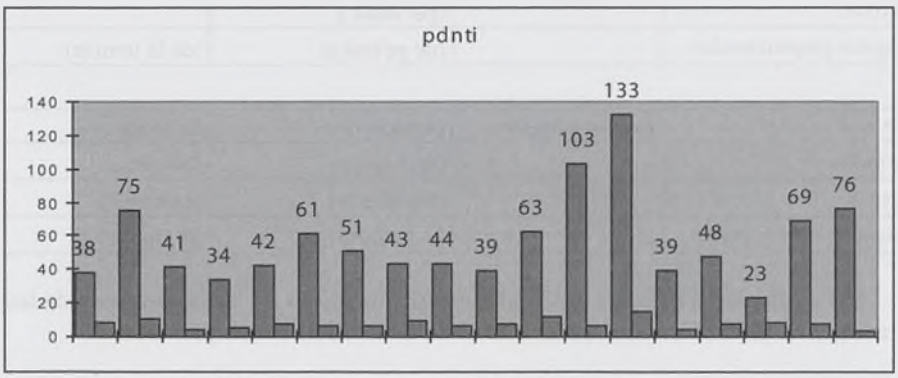


fig.2 Niza [la 'passɛʁa perʁ'duda de la teʁni'ta 'mẽta la taʁ'tuga] ?

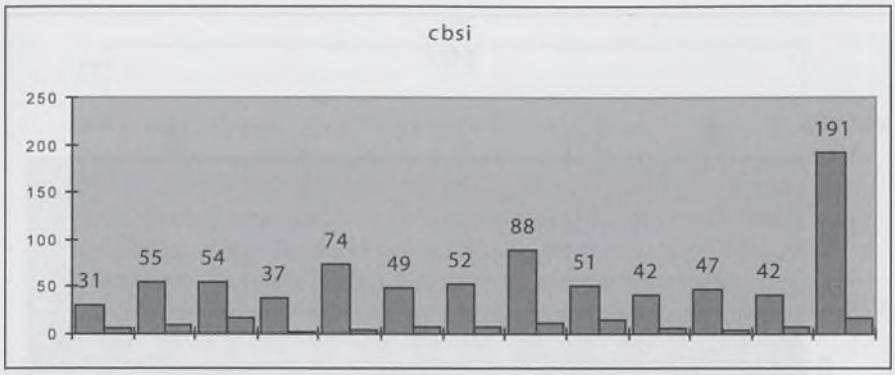


fig. 2bis Gap [la tsa'bʀeta 'tɔtsa la bweisu'na astʀɔ'pja] ?

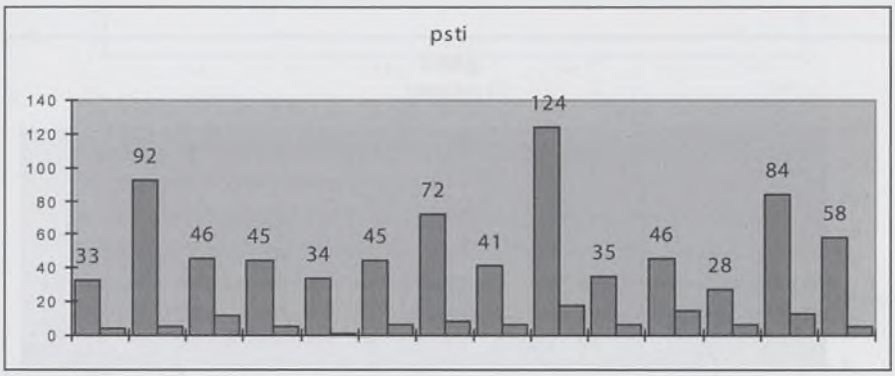


fig.3 Niza [la 'passeʀa de pɛ'laska 'mẽta la taʀ'tuga]?

Cuando hay expansiones en el SV, es frecuente que la duración sea más importante en el adjetivo, pero un oxítono al final implica valores altos (graf.4, graf.5). Por supuesto, se dará siempre un máximo sobre la nasal tónica del verbo.

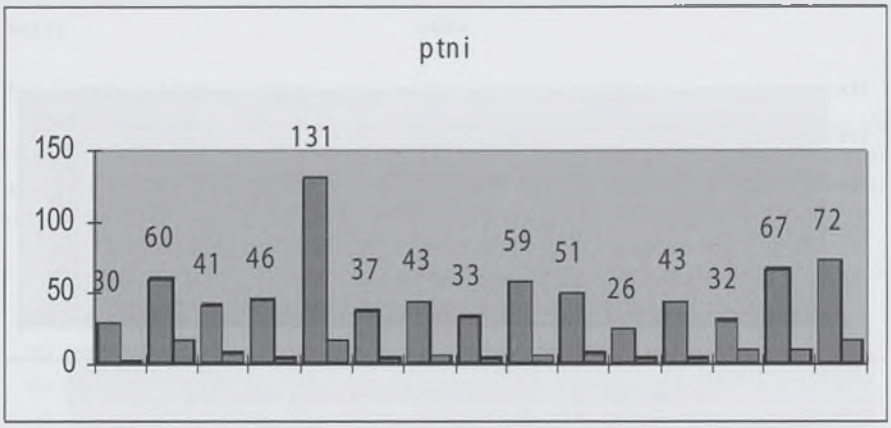


fig.4 Niza [la 'passeʁa 'mẽta la taʁ'tuga de la teʁni'ta] ?

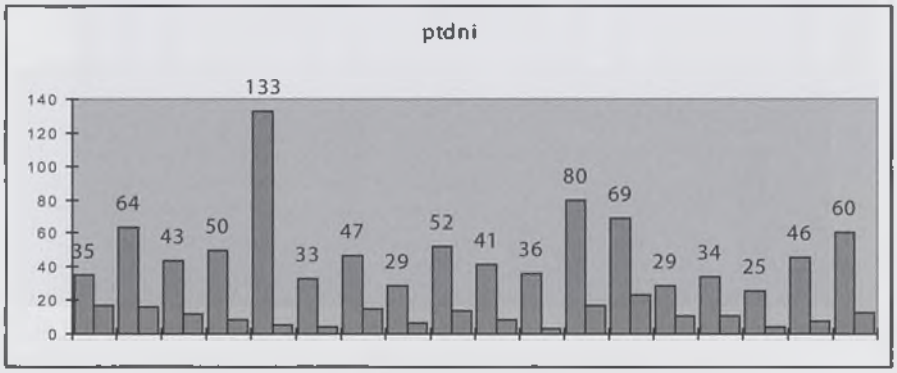


fig.5 Niza [la 'passeʁa 'mẽta la taʁ'tuga peʁ'duda de la teʁni'ta]

1.2. En cuanto a la frecuencia:

Cuando hay 1 ó 2 expansiones en el SN, aparece un pico de frecuencia sobre la tónica de ['passera] (graf.6, fig7: 270 hz mínimo) seguido de un pequeño aumento sobre la tónica del adjetivo ([per'duda]); a continuación se produce un descenso continuo (la curva pasa por debajo del umbral de los 100 hz) antes de una brusca y fulgurante subida en la última tónica, que alcanza al menos los 300 hz seguida, como siempre, de una caída muy importante de la última vocal por debajo de los 100 hz (3 veces menos). De hecho, una buena parte de la curva está por debajo de la frecuencia laríngea media, dibujándose dos valles entre las tónicas del primer y del último elemento del enunciado.

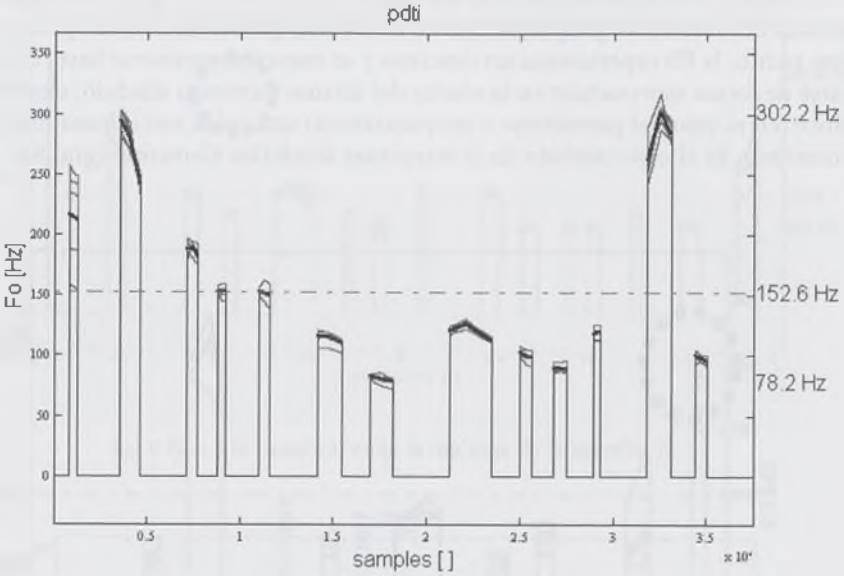


fig.6 Niza [la ˈpasseʁa perʁˈduda ˈmẽta la taʁˈtuga] ?

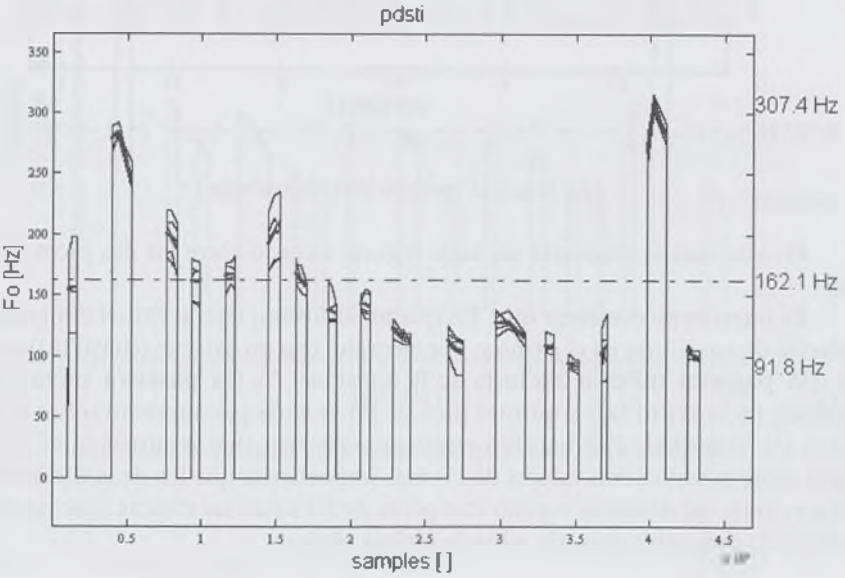


fig.7 Niza [la ˈpasseʁa perʁˈduda de pɛˈlaska ˈmẽta la taʁˈtuga] ?

Cuando la expansión está en el SV, se da algunas veces un pico de F0 sobre la pre-acentuada del sustantivo proparoxítono en el SN ([la ˈpassera]: 270 hz aprox.). A partir de este punto, la F0 experimenta un descenso y se mantiene uniforme hasta que vuelve a elevarse de forma espectacular en la tónica del último elemento añadido, desde la que se produce (en el caso del paroxítono o proparoxítono) una caída vertiginosa final (¡más de dos octavas!). Es el caso también de la estructura sintáctica elemental (graf.8).

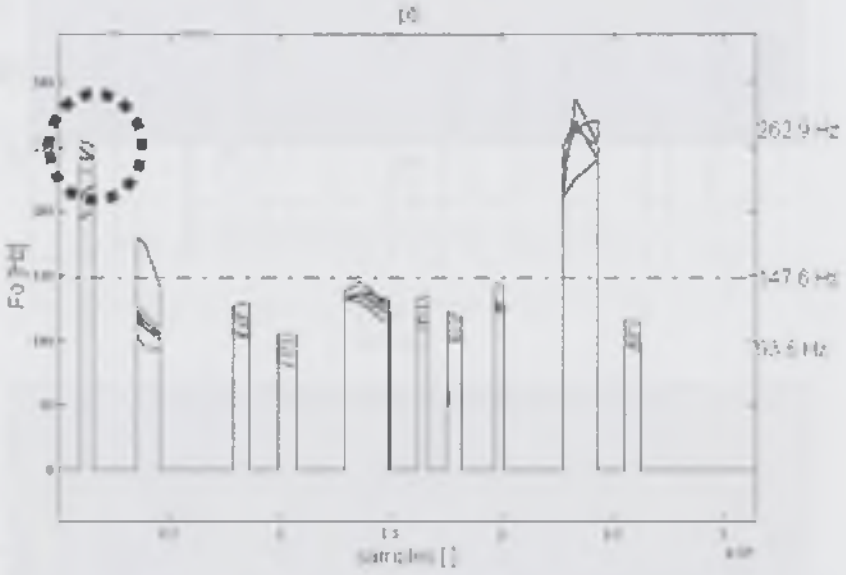


fig.8 Niza [la ˈpasseʁa ˈmẽta la taʁˈtuga] ?

En este caso se observará un valle regular situado entre los dos picos de F0 de la frase.

Es interesante constatar que, En quanto al último tipo acentual del enunciado, se producirá un equilibrio en el ritmo y, por ejemplo, con un oxítono ([tɛrniˈta]) en posición final ([la ˈpasseʁa ˈmẽta la taʁˈtuga de la teʁniˈta] ? y [la ˈpasseʁa ˈmẽta la taʁˈtuga peʁˈduda de la teʁniˈta]) el primer pico de F0 se sitúa por supuesto sobre la pretónica (graf.9, [la ˈpassera]: 275 hz). Sin embargo, con otro tipo acentual (graf.10, graf.11) recaerá sobre la tónica con valores de F0 más importantes que los de la última tónica. Se produce, pues, un descenso regular con picos de F0 sobre las tónicas (que no sobrepasan los 150 hz), para subir después sobre la última tónica.

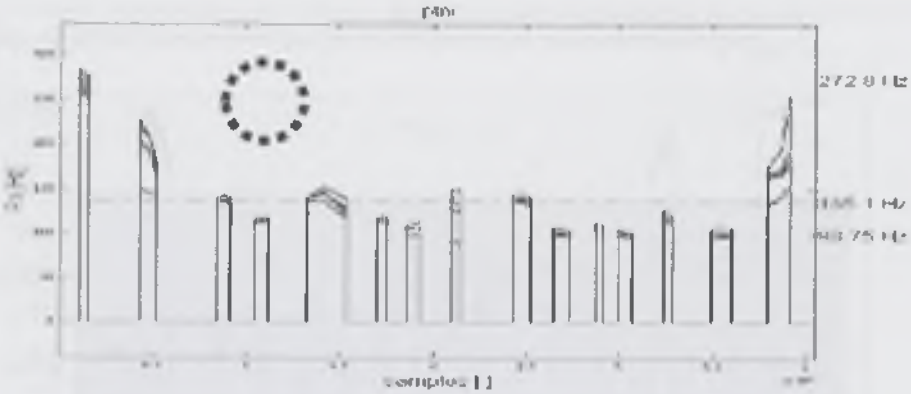


fig. 9 Niza [la ˈpassɛʁa ˈmɛ̃ta la taʁˈtuga de la tɛʁniˈta]?

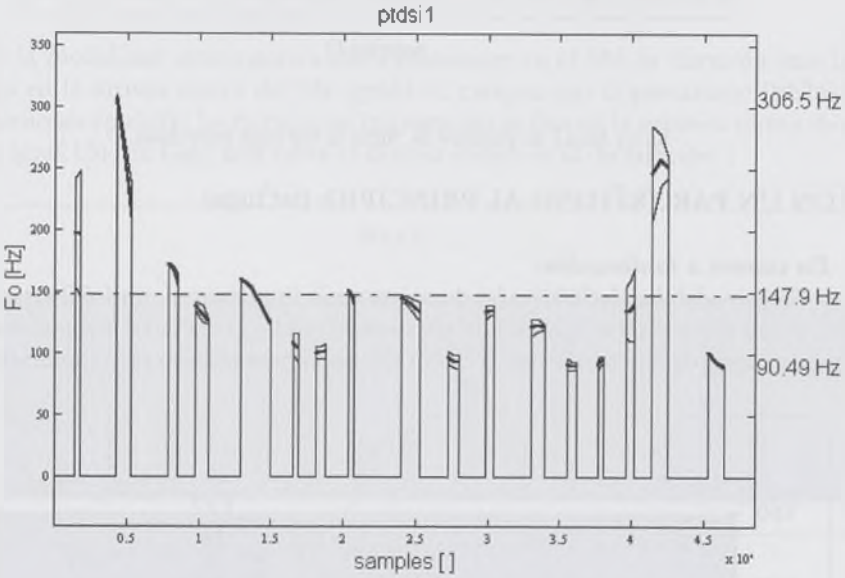


fig.10 Niza [la ˈpassɛʁa ˈmẽta la taʁˈtuga perʁˈduda de pɛˈlaska] ?

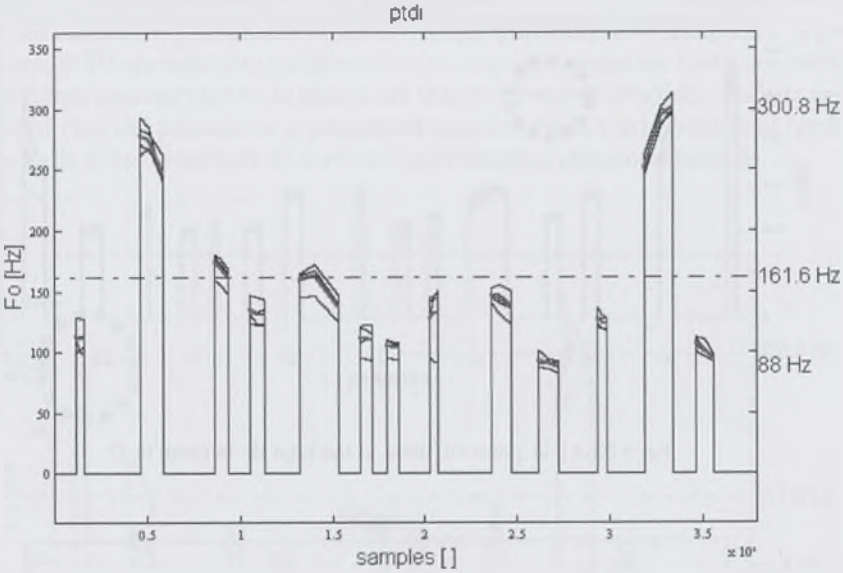


fig.11 Niza [la 'passeʁa la 'mẽta la taʁ'tuga perʁ'duda] ?

2. CON UN PAROXÍTONO AL PRINCIPIO: {tar'tuga}

2.1. En cuanto a la duración:

En la modalidad declarativa las duraciones más importantes, con 1 ó 2 expansiones en el SN, se dan siempre sobre la tónica del elemento añadido (graf.12), tal como se destacó para la variedad sarda de Nuoro, aunque la tónica del verbo presenta siempre el valor más alto.

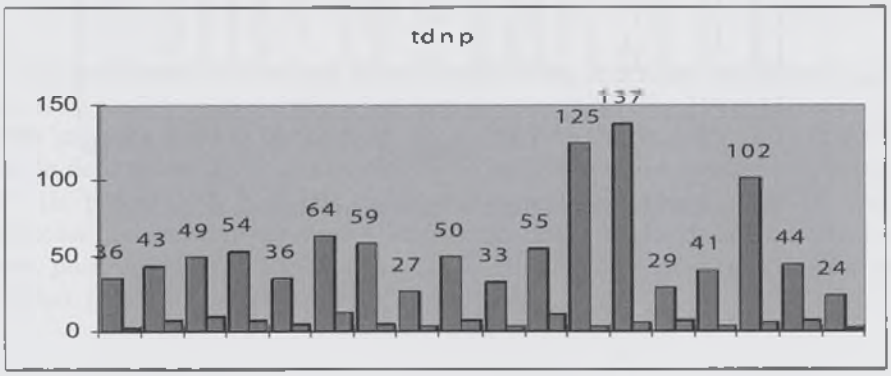


fig.12 Niza [la taʁ'tuga perʁ'duda de la teʁni'ta 'mẽta la 'passeʁa]

Cuando las expansiones se dan en el SV, después del pico en la tónica del verbo, la segunda duración importante aparece en el sustantivo proparoxítono (graf.13).

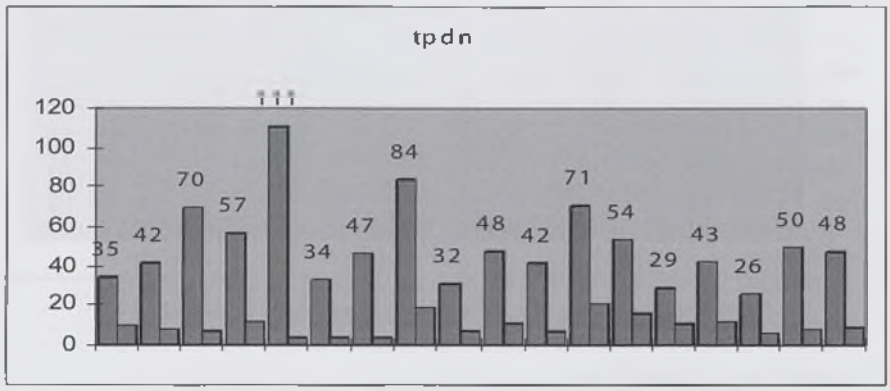


fig.13 Niza [la taʁ'tuga 'mẽta la 'passeʁa perʁ'duda de la teʁni'ta]

En la modalidad interrogativa con expansiones en el SN, la duración más larga se registra en la última tónica del SN (graf.14), excepto con el paroxítono [pe'laska]. Con expansiones en el SV, las duraciones importantes se dan en la primera tónica de este sintagma (graf.15). En Gap, serà sobre el ùltimo tónico de la declaración.

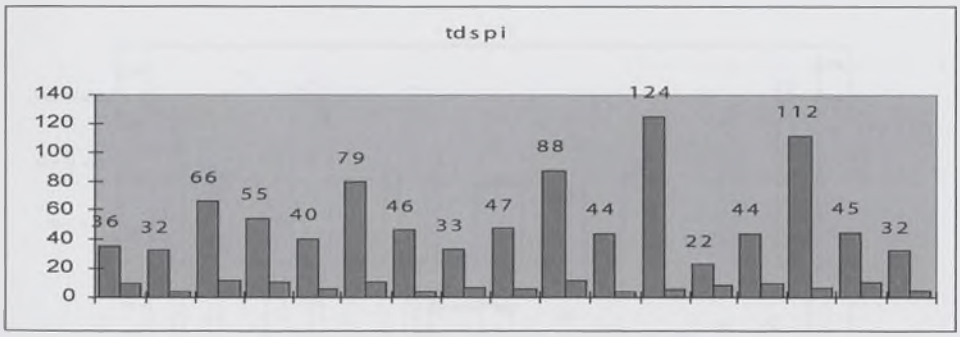


fig.14 Niza [la taʁ'tuga perʁ'duda de pɛ'laska 'mẽta la 'passeʁa] ?

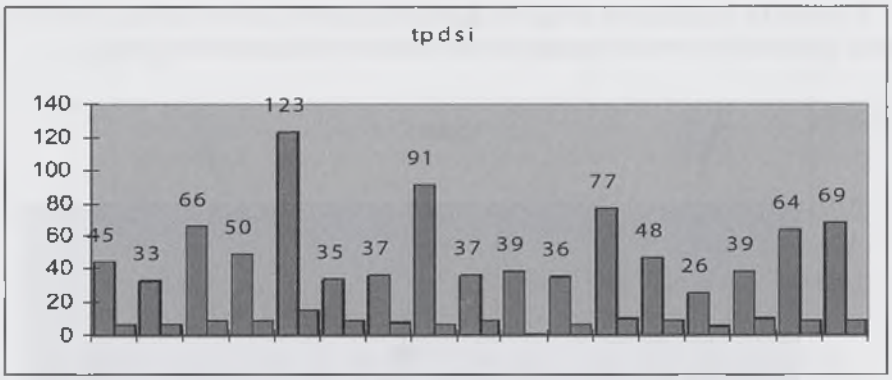


fig.15 Niza [la taʁ'tuga 'mēta la 'passeʁa perʁ'duda de pɛ'laska] ?

2.2. En cuanto a la frecuencia:

En la modalidad interrogativa, en una frase con una estructura sintáctica mínima (graf.16) se producirá inmediatamente, al igual que para el sardo, un pico de F0 sobre la pre-acentuada al principio de frase (240 hz / 150 hz aprox. para la tónica). La aparición de la expansión en el SN (graf.17) implica el desplazamiento del pico de F0 sobre la tónica del sustantivo (en las dos expansiones el grado de declinación es menor), a partir del cual se produce una caída muy fuerte de la línea melódica seguida de una brusca subida sobre la última tónica del enunciado.

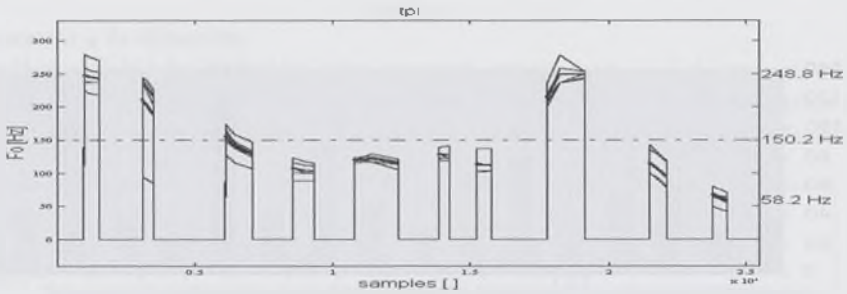


fig.16 Niza [la taʁ'tuga 'mēta la 'passeʁa] ?

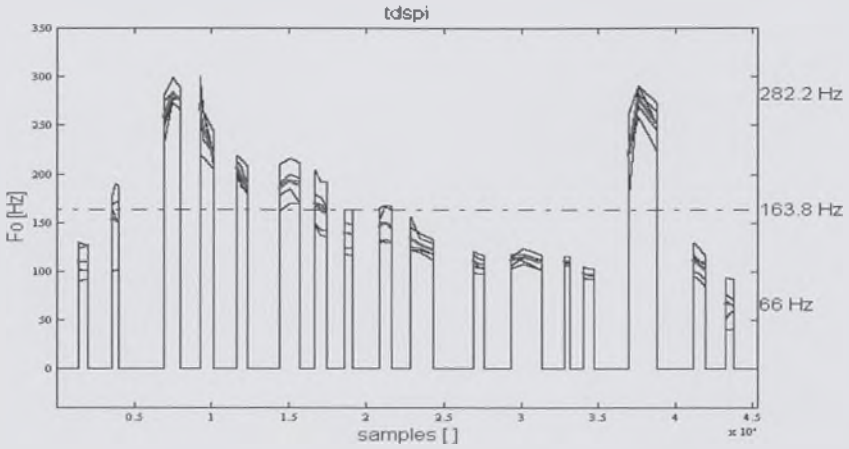


fig.17 Niza [la taʁ'tuga perʁ'duda de pɛ'laska 'mēta la 'passeʁa] ?

Cuando hay un sintagma preposicional oxítono ([tɛrni'ta]) en el SV (graf.18), se da un pico sobre la pre-acentuada del sustantivo del SN (300 hz aprox. / 100 hz para la declarativa), y luego un descenso hasta el verbo desde donde la F0 se mantiene (oscilando entre 100 y 150 hz), con una superposición de las dos curvas (af. e int.) (graf.20), hasta la última tónica de la frase situada en torno a los 300 hz.

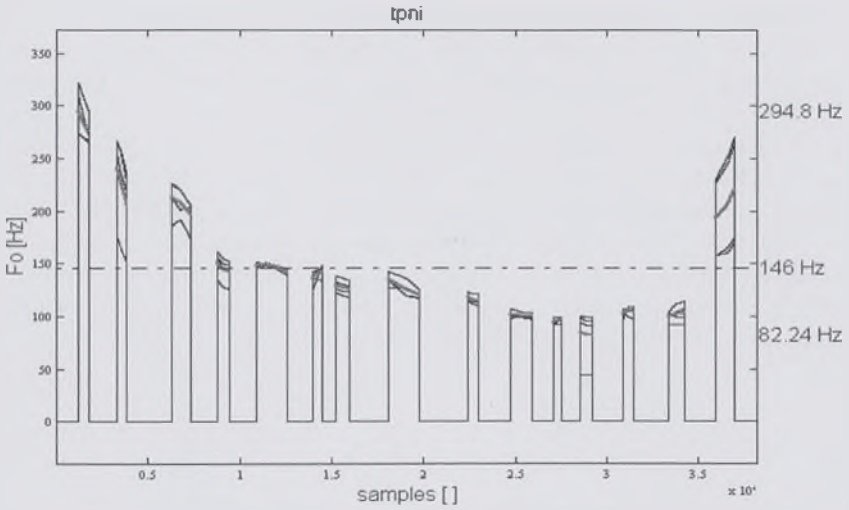


fig.18 Niza [la taʁ'tuga 'mēta la 'passeʁa de la tɛʁni'ta] ?

Finalmente, se produce una caída brutal por debajo de los 100 hz (lo que sigue ocurriendo con los otros tipos acentuales). En los otros tipos acentuales, al final de frase, se observa una subida en la tónica y sobre todo en la última vocal del enunciado, esto es, la postónica del paroxítono [pe'laska] (250 hz) (graf.19).

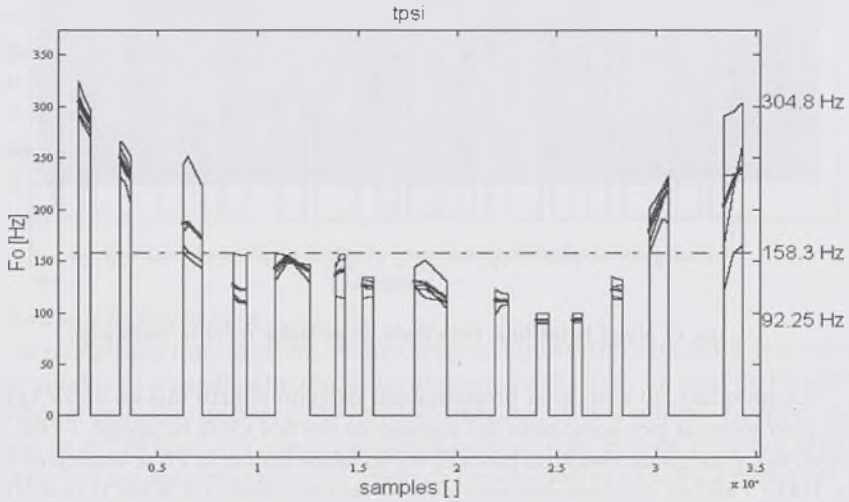


fig.19 Niza [la taʁ'tuga 'mẽta la 'passeʁa de pɛ'laska] ?

Sin embargo, en Gap, con un paroxítono a principio de frase, en la modalidad interrogativa con expansiones en el SN, se obtiene un pico máximo sobre la pretónica del oxítono [bwɛ'su'na] con un descenso de más de una octava entre esta última y la tónica.

En la modalidad interrogativa, se observa una diferencia entre las dos variedades meridionales, en primer lugar En la de Gap (graf.21) los picos de F0 se sitúan sobre las pre-acentuadas de los dos sustantivos, al principio y al final del enunciado. En nissart, en cambio, esto es así solamente en los enunciados que presentan expansiones en el SV (graf.20) o en los que no tienen expansión (graf.22), en este caso, el pico de frecuencia puede alcanzar hasta los 300 hz y, a continuación, la línea melódica experimenta una caída muy fuerte que continúa hasta la última tónica del enunciado, salvo en algunos casos con expansión adjetiva (por ejemplo con [pe'laska]).

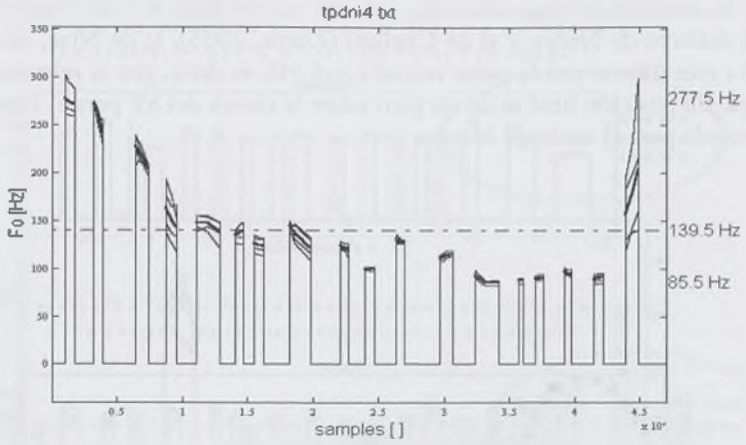


fig.20 Niza [la taʁ'tuga 'mɛ̃ta la 'passɛʁa perʁ'duda de la teʁni'ta]?

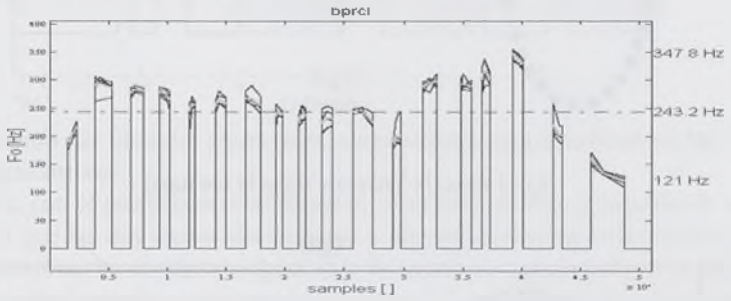


fig.21 Gap [la bwɛ'su'na pøtsu'nɛta də ʁu'mɛta 'tɔtsa la tsa'bʁɛta]?

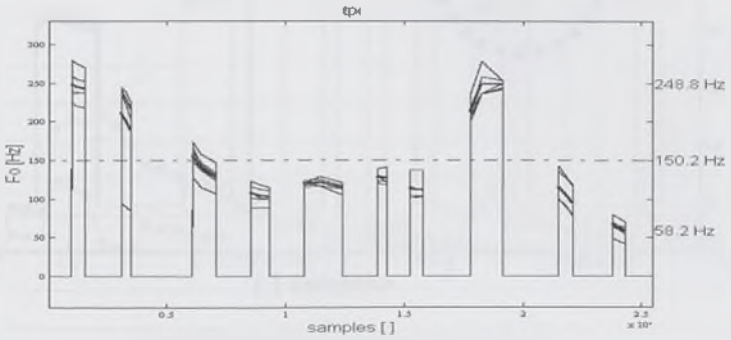


fig.22 Niza [la taʁ'tuga 'mɛ̃ta la 'passɛʁa] ?

Si la variedad de Gap recuerda al sardo por el final de frase (graf.25), concretamente al dialecto de Nuoro y al de Cagliari (Zucca, 2005), la de Niza, en cambio, sólo recuerda a este último por la parte inicial (graf.24), es decir, por la relevancia de la pre-acentuada. En posición final se da un pico sobre la tónica del SV para el nissart y sobre la pre-acentuada para el sardo de Nuoro.

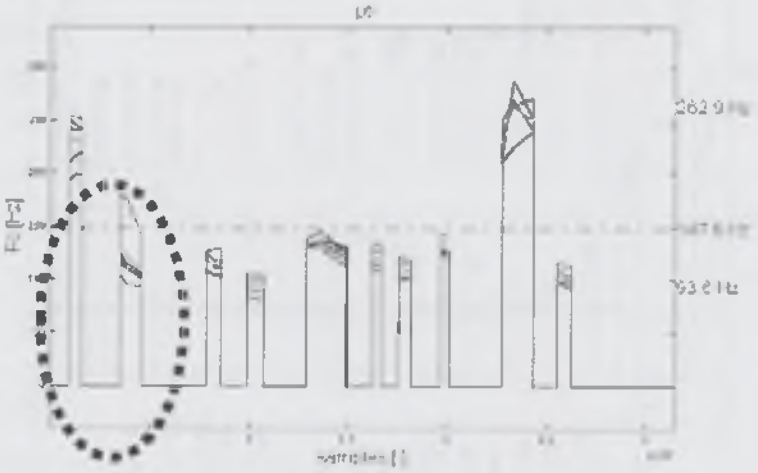


fig.23 Niza [la 'passeʁa 'mɛta la taʁ'tuga] ?

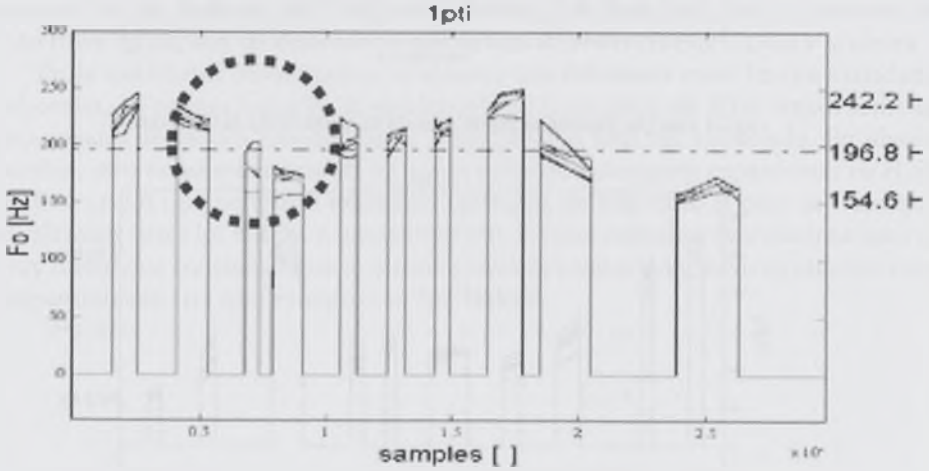


fig.24 Nuoro [sa 'pipera 'tɔka sa tɛ'raka] ?

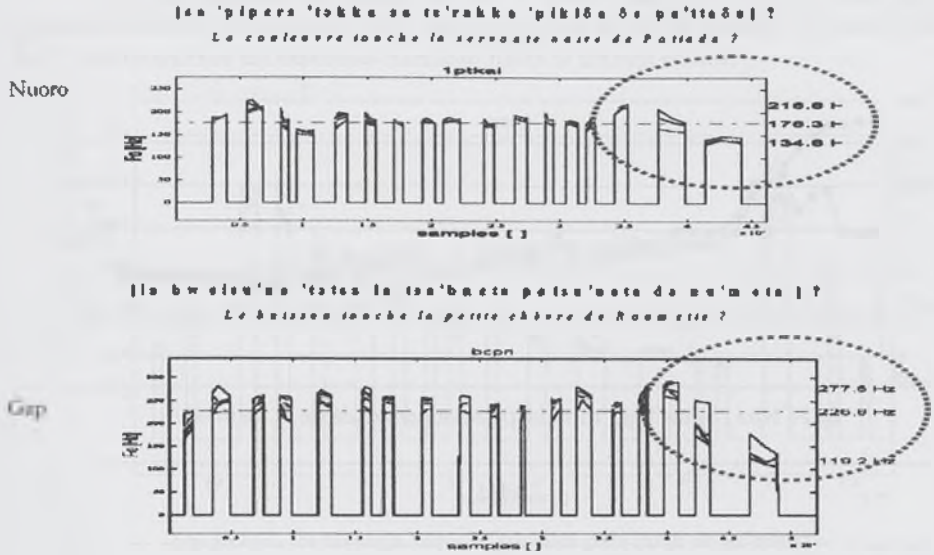


Fig.25

Asimismo, además del final del enunciado, encontramos otra similitud en las "mesetas" entre los dos sustantivos.

En Niza, con el paroxítono [tar'tuga] al principio de frase y seguido de expansiones, se observa que las dos curvas declarativas e interrogativas son relativamente distintas; sólo se aproximan en el adjetivo (graf.26). En cambio, cuando presentan expansiones en el SV, éstas prácticamente se superponen a partir del sustantivo y terminan separándose únicamente en el elemento final, al igual que con ['passera] (graf.27-29).

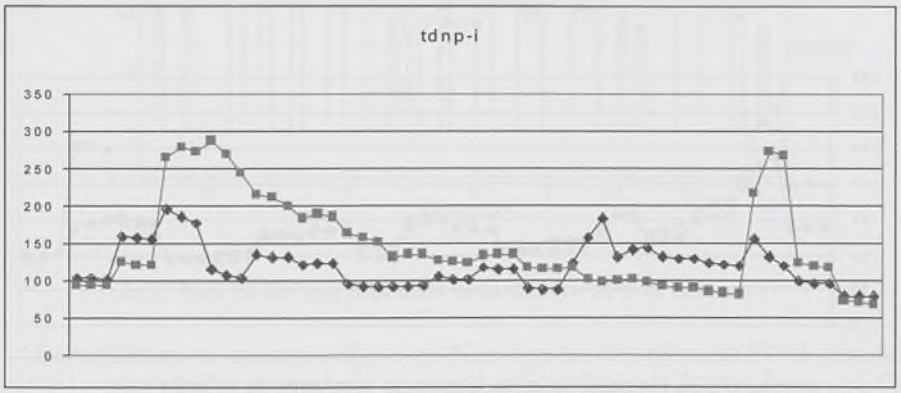


fig.26 Niza [la taʁ'tuga perʁ'duda de la teʁni'ta 'mẽta la 'passeʁa] / ?

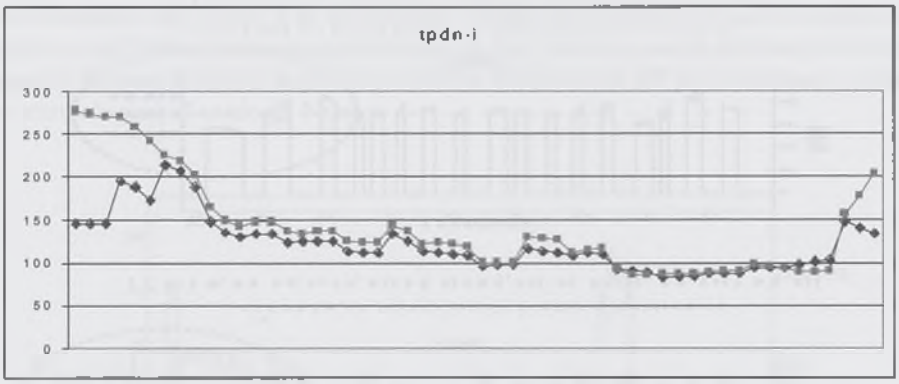


fig.27 Niza [la taʁ'tuga 'mẽta la 'passeʁa perʁ'duda de la teʁni'ta] / ?

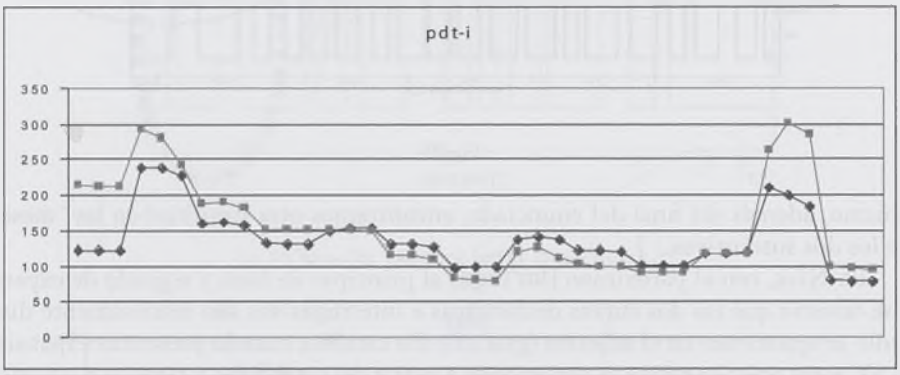


fig.28 Niza [la 'passeʁa perʁ'duda 'mẽta la taʁ'tuga] / ?

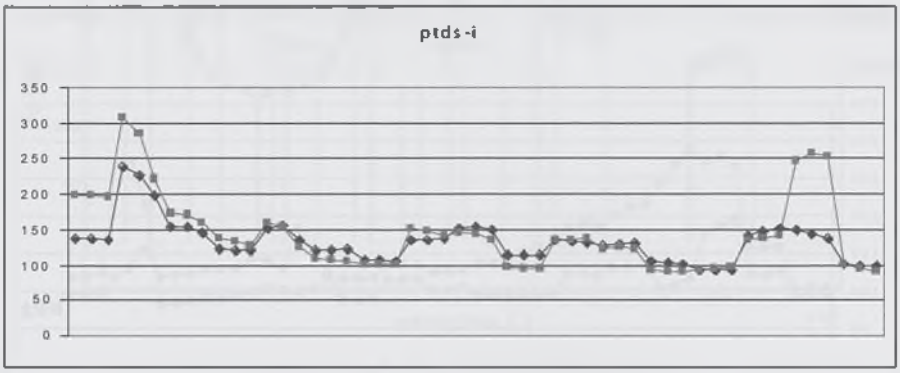


fig.29 Niza [la 'passeʁa 'mẽta la taʁ'tuga perʁ'duda de pɛ'laska] / ?

Otra particularidad es la aparición de 1 (graf.30), 2 (graf.31) ó 3 valles según el tipo acentual final, pero cuando hay expansiones en el SV se obtiene un único valle, es decir, prácticamente un descenso continuo hasta la última tónica.

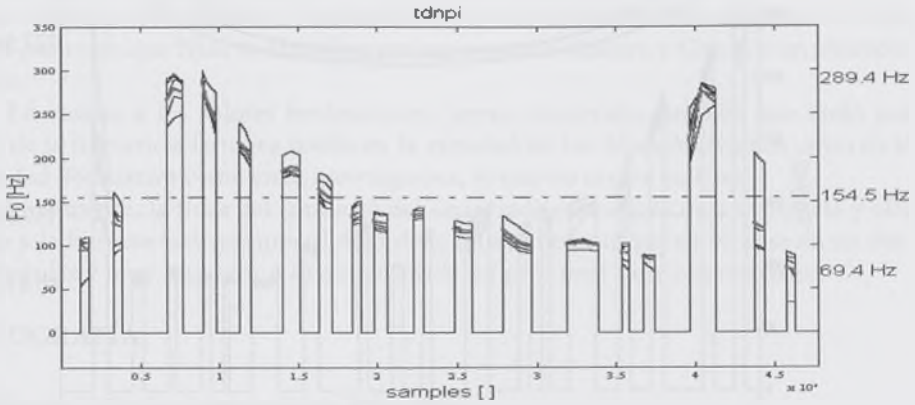


Fig.30 Niza [la taʁ'tuga 'mẽta la 'passeʁa perʁ'duda de pɛ'laska] ?

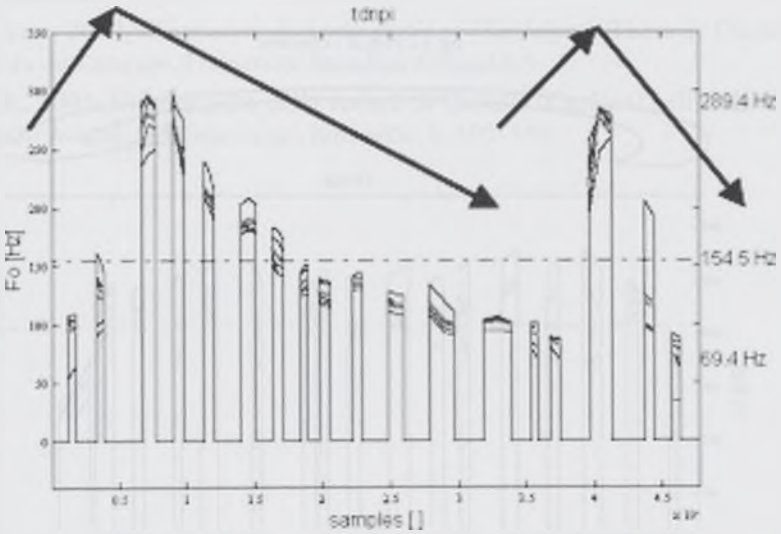


fig.31 Niza [la taʁ'tuga perʁ'duda de la teʁni'ta 'mẽta la 'passeʁa] ?

Encontramos un contorno típico en Niza con los dos picos de F0 al principio y al final del enunciado de manera que, entre uno y otro, la curva de F0 adopta una forma cóncava (graf.32), mientras que en Gap es convexa (graf.33).

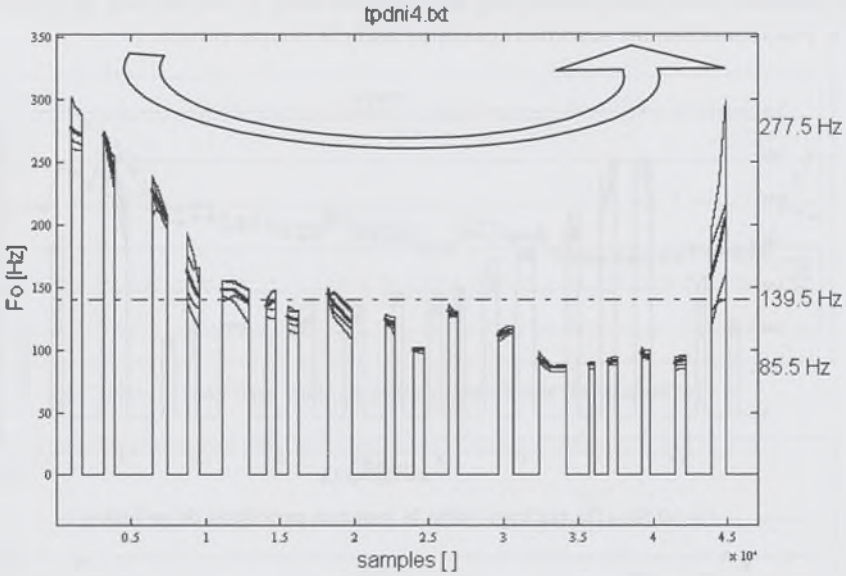


fig.32 NICE : cóncava

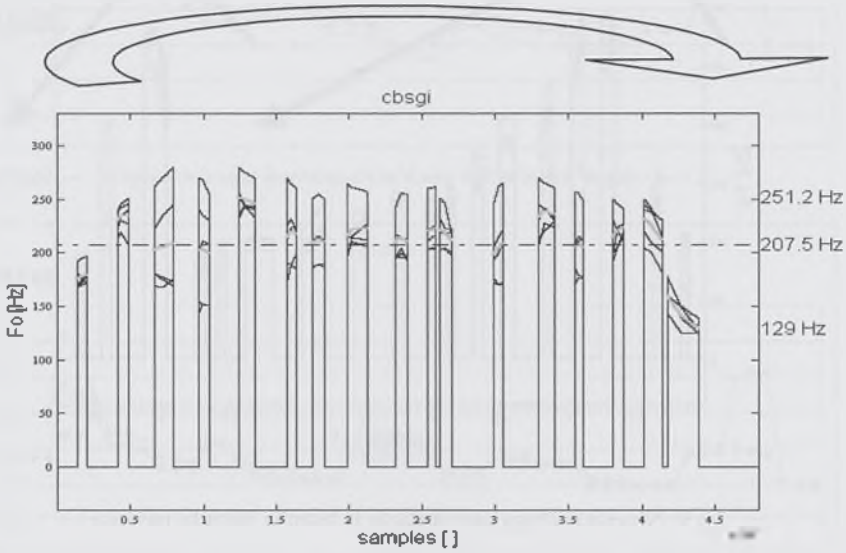


fig.33 GAP : convexa

CONCLUSIONES

Con este trabajo, hemos podido constatar a través del estudio de estas dos variedades occitanas tan distintas que la línea melódica es realmente diferente. No obstante, aunque ambas presenten uno o dos valles, éstos son mucho menos acusados en Gap, hasta el punto de que Niza se identifica por un contorno cóncavo y Gap por un contorno convexo.

En cuanto a los valores frecuenciales, hemos observado también que están por debajo de la frecuencia laríngea media en la variedad de los Alpes Marítimos tanto en la modalidad declarativa como en la interrogativa, lo que no ocurre en Gap.

Finalmente, la línea del conjunto del enunciado es siempre muy compacta y alta en Gap y la frecuencia disminuye al final de la frase, mientras que en Niza se da un descenso regular y muy marcado, a lo que se añade un pico final muy característico.

BIBLIOGRAFÍA

BEC Pierre, 1978, *La langue occitane*, Paris: PUF.

LAI, Jean-Pierre, 2005, « Etude de l'intonation d'un parler occitan : Gap », *Géolinguistique*, hors série 3, 65-86.

LAI, Jean-Pierre, 2002, *L'intonation du parler de Nuoro (Sardaigne).* Thèse de Doctorat en Sciences du Langage, Université Stendhal-Grenoble3.

ZUCCA Iside, 2005, L'intonazione della varietà di Castello (Cagliari) nell'ambito del progetto Amper, *Géolinguistique*, hors série 3, 305-336.

PERCEPTIONAL AND ACOUSTIC STUDY OF ASPIRATION IN ENGLISH

TERESA LÓPEZ-SOTO
Universidad de Sevilla. SPAIN teresals@us.es

1. ASPIRATION IN ENGLISH

Lisker and Abramson (1964, 1967) have shown that VOT is a very effective measure to determine stops with different laryngeal actions. Aspiration has traditionally been associated with voiceless plosive consonants (stops) in very specific collocations: initial position in a stressed syllable. All stops show a period of silence which goes throughout the whole spectrum. During this closure, fully voiced stops (e.g. intervocalic voiced stops) will show a low level of quasiperiodic energy during the closure. That is, there is an absence of energy except at a low frequency. This low-frequency component is what is known as a "voice bar", generally below 250 Hz. Voiced stops are different from voiceless stops in that the latter usually present a higher onset or offset in fundamental frequency (in contact with a following or a preceding vowel). According to Lisker (1957) a fully /b/ (i.e. intervocalic) has an average duration of 75 msec, while an intervocalic voiceless /p/ has an average of 120 msec. One more distinction is that the adjacent vowel to a voiced stop is likely to be marked by a rising bend of F1, which is not marked in the case of the voiceless ones.

However, voiced stops in English /b, d, g/ get devoiced in some specific phonetic contexts: in contact with a voiceless sound, at the beginning and at the end of the utterance/tone unit (i.e. in contact with silence/pause). As a result, the real acoustic and perceptual cues to distinguish voiced and voiceless stops in English are the following:

a) Initially, voiced stops /b, d, g/ show a weak burst of noise associated with the onset of the release stage, while voiceless stops /p, t, k/ lack the gap before the onset of voicing, this is known as **aspiration**.
b) Medially, there is a longer closure period and, as a result, absence of energy, in /p, t, k/.
c) Finally, they influence the duration of the preceding vowel: vowels are lengthened when followed by a voiced stop.

Aspirated voiceless stops are said to occur initially in stressed syllables, that is, we would have aspiration in "table" but not in the first /t/ in "tattoo". There is a weak aspiration in word-final position, as in "that". Some of the most important and productive studies in aspiration in English comes from the work of Leigh Lisker and Arthur Abramson. The main conclusion from their work is that voice onset time (VOT) is often to be taken as the most adequate physical basis for separating homorganic stop categories in a variety of languages. VOT is usually described as the duration of the time interval between the onset of the release explosion (the burst of noise) and the onset of glottal pulsing. This acoustic measure does not evaluate, however, other laryngeal and supraglottal events whose temporal relation it reflects. The VOT measure works this way:

a) The voice bar precedes the release stage in the voiced stops /b, d, g/. VOT has a negative value here.
b) The voice bar shows off simultaneously with the release stage onset. According to the theory by Lisker & Abramson (1964, 1967, 1970) then, Spanish voiceless stops should be included here and be considered as voiced stops too but with a [+fortis, +voice] characterization, in opposition to voiced Spanish stops which have a [-fortis, +voice] characterization. VOT has a 0 value here (neutral VOT). See also López-Soto & Rodríguez-Romero (2002).
c) The voice bar can appear slightly after the release stage (positive VOT, +VOT).
d) The voice bar can appear much after the release stage of the plosive has taken place (+VOT).

So, VOT has to do with the point in time at which the vocal folds start to vibrate. This starting point is measured in relation with the release of the closure stage in plosives. In a fully voiced plosive the vocal folds vibrate throughout, thus having a negative VOT. In voiceless unaspirated plosives, there is a delay (or lag) before voicing starts (case c) above). When the delay is much longer, as described in d) above, aspiration occurs. This amount of delay varies from language to language.

2. IMPORTANCE OF ASPIRATION IN ENGLISH

In spite of its importance, aspiration has been said to be **redundant** by many theoreticians (Akmajian, Demers and Harnish, 1979; Anderson, 1974; Fromkin and Rodman, 1983). Redundant here means that the absence of aspiration in words such as "pin" may merely determine a non-native speech. This statement is contradictory with the fact that in the pair "pin" vs. "spin", the phonetic distinction established shows that the plosive /p/ following initial /s/ has no **distinctive** status as a member of the voiceless /p/ as it could have as a member of the voiced /b/: either a form /spɪn/ or a form /sbɪn/ are equally possible in English. Auditory and acoustic evidence shows that no real distinc-

tion is to be seen between them and that the phonological theory motivating choosing /p/ instead of /b/ for words such as "spin" relies on the spelling convention of standard orthography merely.

A more precise account for the distinction between voiced and voiceless stops in English would be, according to Lisker (1984) that it is aspiration which infers the absence of pre-release voicing and, thus, determines the distinction. Because the absence of aspiration is compatible with both [+voice] and [-voice] closure, the rules we can state are that [+aspirated] leads to [-voice] in English and that [+voice] leads to [-aspirated]. Any other rule would be ambiguous for [-voice] could lead to [±aspirated] and [-aspirated] would lead to [±voice].

The present literature on the phonetic and phonological description of the English language keeps on emphasizing a classification of the aspirated plosives which follows a very weak assumption. It is stated that, considering that a stop is voiceless, it is not preceded by initial /s/, it is word-initial or that the next vowel is stressed and within the same word, it is spelled phonologically /p/ and not /b/, then, we can infer that its release will be aspirated. The evidence, however, shows that the voiceless stop is not aspirated because it is voiceless and initial, but rather it is voiceless because it is aspirated.

This discussion proves the importance in the distinction **aspiration** vs. **voicing** for stops, rather than **presence of voicing** vs. **absence of voicing**.

The study we present here follows this assumption and gives an account on how oral stops and affricates are to be differentiated in English. The investigation, then, applies the same characterization to affricates (aspiration vs. voicing) and studies the influence of the fricative stage and how the plosive release is to be understood in this kind of hybrid sound.

3. JUSTIFICATION OF THE VOT MEASURE

VOT is often taken as the unique measurement to determine whether aspiration exists. This will be the approach we will follow in this study, not because we think it is to be considered the only acoustic perspective to be considered. Other studies have succeeded in analyzing and getting to conclusions after considering other aspects of sounds, namely, F1 transition, as in Stevens and Klatt (1974). Aspiration is acoustically described as an interval of noise, with a broadband frequency distribution that co-occurs with the closure period and extends over the release stage. Aspiration has been considered a cue for voicelessness in English consonants. However, perceptual experiments carried out by Lisker (1978) have not been contradicted yet. Considering perception mainly, stop voicing seems to depend on more than the absolute duration of the interval that separates the burst of noise after the release stage and the onset of voicing and F1. Our choice of VOT measurement also responds to the characteristics of the study itself, where plosives and affricates are analyzed. It would be difficult to apply the F1 transition technique to the affricate sounds, where the frication stage overlaps the release stage of the first part of the hybrid.

4. ASPIRATION ACROSS LANGUAGES

What this study tries to demonstrate is the real function of aspiration in English. Aspiration in English, as we have said, is not distinctive. We do not disagree in this point, but we find that the dichotomy voiced vs. voiceless for the classification of plosive consonants lack scientific basis. As we were stating above, it is aspiration which infers the absence of pre-release voicing and, thus, determines the distinction between homorganic plosives in English. Because the absence of aspiration is compatible with both [+voice] and [-voice] closure, the rules we can state are that [+aspirated] leads to [-voice] in English and that [+voice] leads to [-aspirated]. Any other rule would be ambiguous for [-voice] could lead to [±aspirated] and [-aspirated] would lead to [±voice]. Apart from this clarification, we will not assume that aspiration is a distinctive feature in English. It is distinctive in other languages, as in Nepali, where there is a fourfold classification: aspirated and non-aspirated voiced and voiceless stops. In Hindi, we also find that fourfold distinction, but with the inclusion of post-alveolar affricates (Tej K. Bhatia, 1973; R. Ahmad & S. Agrawal, 1969; J. Gupta et al., 1969; M. Ohala & J. Ohala, 1972; R. N. Srivastav, 1968). Other languages also exhibit this aspirated vs. non-aspirated phonemic distinction, as Korean (Nam-Kil Kim, 1987). In Nar-Phu (a language spoken by 800 hundred people in Nepal) (M. Mazaudon, 1997; M. Noonan, forthcoming) and Chichewa (in Congo) (S. Mchombo, 2005), aspiration is a distinctive feature for both voiceless stops and affricates. Of this cross-linguistic account, it is worth noted the occurrence of aspiration in both stops and affricates in Hindi, Korean, Nar-Phu and Chichewa. We will see how aspiration behaves when it comes to affricates and stops in English.

5. ENGLISH AFFRICATES

Affricates are usually described as being a hybrid sound, meaning that after a quick plosive burst, a "normal" frication follows. The air-pressure behind a complete closure in the vocal tract is gradually released in the case of affricate sounds. The separation which follows is sufficiently slow, and not sudden as in the case of the plosives. This slow separation allows for audible friction to take place. The duration of the friction is not as long as would be the case of a fricative sound. There are two affricate sounds in English, namely /dʒ/ and /tʃ/. The former is described as being [+voice], the latter [-voice] sounds, where the frication stage overlaps the release stage of the first part of the hybrid. There are cases in the language where the phoneme sequences may resemble /dʒ/ and /tʃ/, that is the case of the words "lids" or "cats", for example. Their spectrographic representation may show a brief plosive burst passing immediately into frication. However, the fact that they may occur only in syllable-final position make them be considered as allophones of a plosive plus a fricative. The most frequent cue to distinguish a fricative sound from a frication stage in an affricate is found in the increased asymmetry of the frication rectangle, probably caused by the passage from the plosive. Affricates usually show a bulge in the lower frequencies of the frication stage.

From an acoustic point of view, affricates have traditionally been described taking into account their fricative stage. Cruttenden (1994) states that the most essential perceptual cues to identify the affricates will be in the transition between the preceding vowel and the stop and by the explosive onset of the friction. We have applied the same VOT rule to identify aperiodic noise turbulent in both stops and affricates. We have analyzed voiced and voiceless consonants. The most difficult part was that of isolating the plosive stage. This plosive component is detectable through the single visible bar just to the side of the frication portion of the phoneme. When the plosive is too weak, the affricates can be confused with the fricatives. Affricates are also very likely to be affected by neighboring phonemes as the bands of frication will often extend toward the locus of the following phoneme, passing into voiced formants. This, however, does not affect the VOT analysis on them.

6. THE EXPERIMENT

In this paper we show the results of an acoustic and perceptional study of aspiration in voiced and voiceless consonants in English. The first level of analysis shows an account of the acoustic features of these consonants, with special attention to VOT duration and interval of noise on the release stage of stops. A semi context-free data of a set of 8 consonants (6 stops plus 2 affricates) in initial, medial and final position has been collected. The measurement has been done on prevocalic, and intervocalic stops, mainly. The latter context also includes the occurrence of stops in final position in the syllable. A distinction between strong and weak syllables has also been established. This corpus was recorded by 7 American English native speakers, 1 Irish English native speaker and 3 British English native speakers. The acoustic study shows clear evidence of aspiration in voiceless stops and affricates. The existence of turbulent noise during the release stage can only be partially visualized in voiced consonants in limited contexts. The perceptional study, though, accounts for the existence of aspiration in voiced stops in initial position in conjunction with stress. The results are not as definite as for the voiceless stops in the same position but clearly demands for further acoustic data to fully account for aspiration in English.

A collection of phrases where made where the three English plosives and the two affricates were located in between vowels. Phrases were not longer than a natural tone unit, making compact semantic chunks isolated in context. The target sounds (plosives and affricates) were given the focus and, thus, sentence stress in obvious ways. The selection of phrases was extensive enough as to divide words in the following groups:

1. The target sound is word-initial, weak position, unstressed syllable, intervocalic.
2. The target sound is word-initial, strong position, stressed syllable, intervocalic.

We, therefore, did not pay attention to middle or final position in the word nor in the tone unit/utterance. This study does not analyze the target sounds, plosives and affricates, making special indication of the preceding or following vowel. In every case, the preceding vowel was a weak schwa, the following vowel could be very different from one instance to another. Further study is necessary to complement the conclusions we reached in this one.

7. RESULTS

The perceptional study consisted of an audition carried out by expert phoneticians, native and non-native speakers of English. They listened to the samples in context and gave their impressions on whether aspiration had taken place. The recordings were augmented by alien data consisting of the reading of the same samples by non-native speakers. This data was taped by non-natives with a high command of the language, some of them, bilingual or having lived in an English-speaking country for longer than 10 years. No indication was given whatsoever on which recordings corresponded to native or non-native speakers. They had to evaluate whether aspiration had taken place in the plosives (not in the affricates). Our study extracted their conclusions on the data that had been recorded by the native speakers only.

Their conclusions stated than aspiration was always happening in the case of the alveolar plosive, regardless of the quality and the quantity of the following vowel, and regardless of whether the syllable was stressed or not. The perception of aspiration decreased to 85% in the case of bilabial and velar voiceless plosives. In this group of plosives, aspiration was only 100% of the cases perceived when the syllable was obviously stressed. Aspiration was not perceived in about 15% of the cases when the syllable was unstressed (as in "pedagogical", with a secondary stress only).

As for the acoustic analysis, we paid attention to the burst of turbulent noise considering the VOT parameter. We discovered the evaluators findings in that aspiration was much stronger for the alveolar voiceless plosive than for the bilabial or velar ones. That is, aspiration had a higher positive value in the case of the alveolar. This must explain why the evaluators could perceive it more clearly. Our analysis corroborates others (see M.T. López-Soto & L. F. Rodríguez-Romero, 2002) and the traditional bibliography in that voiced plosives and affricates show no aspiration. The only perceivable and visible energy before and after the stop stage is the voice bar indicating phonation. As for the voiceless affricate /tʃ/, the study did not find any trace of aspiration when the consonant was pronounced in a weak, unstressed syllable. However, aspiration was clearly seen in those cases where the affricate was clearly marked by sentence stress, although this was not consistent throughout the database. Of the 11 speakers, only 3 did pronounce an aspirated voiceless affricate in this context, with a VOT value of an average 0.03 msec after the release stage had taken place.

8. CONCLUSIONS

The perceptional and the acoustic study corroborate the existence of aspiration in the case of voiceless plosive sounds in English. This aspiration is strongly marked in the case of the alveolar plosive, regardless the characteristic of the syllable and stress. An earlier VOT for the bilabial and the velar plosives may indicate that aspiration is coupled by the action of the resonator once the release stage takes place. We should, however, consistently confirm these conclusions taking into account a more exhaustive study of the context of the consonants in the three types of plosives.

The absence of aspiration in the voiced affricate is consistent with the same absence for voiced stops. What is more interesting to highlight here is the appearance of aspiration in the voiceless affricate only when the consonant is in a stressed position. Stress seems to determine the appearance of a stronger aspiration in the case of the voiceless stops, too. This conclusion, however, deserves further study with an extended database so that we can determine if the quality of the vowel has an influence on aspiration in both plosives and affricates and whether stress is the primary conditioning factor in the burst of turbulent noise.

BIBLIOGRAPHY

AHMAD, R. & S, Agrawal. "Significant features in the perception of [Hindi] Consonants". *Journal of the Acoustical Society of America* 3: 758-763. 1969.

AKMAJIAN, A., R. A. Demers & R. M. Harnish. *Linguistics: An introduction to language and communication*. (5th Ed.) Cambridge, MA: MIT Press. 2001.

ANDERSON, S. R. *The Organization of Phonology*. New York: Academic Press. 1974.

CRUTTENDEN, A. *Gimson's Pronunciation of English*. London: Edward Arnold. 1994.

FROMKIN, V. & R. RODMAN. *An Introduction to Language* (7th ed). New York: Heinle. 2002.

GUPTA, J., S. AGRAWAL & R. AHMAD. "Perception of [Hindi] consonants in clipped speech". *Journal of the Acoustical Society of America* 3: 770-773. 1969.

KIM, N-K. "Korean". Bernard Comrie (ed.). *The World's Major Languages*. London: Croom Helm, pp. 881-898. 1987.

LISKER, L. "Closure duration and the intervocalic voiced-voiceless distinction in English". *Lang* 33: 42-49. 1957.

LISKER, LEIGH & Arthur S. ABRAMSON. "A cross language study of voicing in initial stops: Acoustical measurements." *Word* 20: 384-422. 1964.

LISKER, LEIGH & Arthur S. ABRAMSON. "Some effects of context on voice onset time in English stops." *Language and Speech* 10: 1-28. 1967.

LISKER, Leigh. "A qualified defense of VOT". *Language and Speech* 21,4: 375-383. 1978.

LISKER, Leigh. "How is the aspiration of English /p, t, k/ "predictable"?. *Language and Speech* 24, 4: 391-394. 1984.

LÓPEZ-SOTO, M. T. & L. F. RODRÍGUEZ-ROMERO. "Estudio contrastivo de la aspiración fuerte en consonantes oclusivas: inglés-castellano". *Actas del II Congreso de Fonética Experimental*. Sevilla: Laboratorio de Fonética. (pp. 254-257). 2002.

MAZAUDON, M. "An outline of the historical phonology of the dialects of Nar-Phu (Nepal)"..*Linguistics of the Tibeto-Burman Area* 19.1:103-14. 1997.

MCHOMBO, S. *The Syntax of Chichewa*. Cambridge: Cambridge University Press. 2005.

NOONAN, M. (forthcoming). "The Nar-Phu Language." Randy LaPolla and Graham Thurgood (eds.): *The Sino-Tibetan Languages*. Richmond: Curzon Press.

OHALA, M. & J. OHALA. "The problem of aspiration in Hindi phonetics". *Annual Bulletin n. 6*, Research Institute of Bogopedics and Phoniatrics, University of Tokyo, 1972.

SRIVASTAV, R. N. "Theory of morphonematics and aspirated phonemes of Hindi". *Studies in Hindi Linguistics*, AIIS, New Delhi. 1968.

STEVENS, K. N. & D. H. KLATT. "Role of formant transitions in the voiced-voiceless distinction for stops". *Journal of the Acoustical Society of America* 55: 653-659. 1974.

LA ENTONACIÓN EN LAS FRASES CON EXPANSIÓN DEL «TORTOSÍ»

EUGENIO MARTÍNEZ CELDRÁN
martinezceldran@ub.edu

JANINA ESPUNY MONSERRAT
janinaespuny@ub.edu

CARLOS VAN OOSTERZEE
carlosvanoosterzee@hotmail.com
Universitat de Barcelona

1. PRESENTACIÓN[1]

Tortosa y su zona de influencia se sitúa al sur de Cataluña y, por tanto, limita con la provincia de Castellón que pertence a la Comunidad valenciana; por sus características generales esta zona geográfica se considera perteneciente al dialecto occidental del catalán, pero por su personalidad se puede considerar un subdialecto diferenciado de otros subdialectos occidentales.

En ocasiones anteriores ya estudiamos la prosodia en general de las frases sin expansión del "tortosí" (Van Osterzee et alii 2004, Carrera et alii 2004 y Martinez Celdrán et alii 2005). Además, esta zona ya ha sido estudiada, en lo que a las interrogativas se refiere, desde otros presupuestos por Pradilla y Prieto (2001, y Prieto y Pradilla 2003); no obstante, también se trataba de frases breves sin expansión. Por tanto, se puede decir que es la primera vez, si nuestra información es correcta, que se estudian frases con cuatro acentos tonales desde la perspectiva de este subdialecto. En este trabajo nos limitamos a la entonación y en el futuro prestaremos también atención a la duración y la intensidad.

2. CARACTERÍSTICAS METODOLÓGICAS

Se grabó el corpus preparado para el catalán con las características que exige el proyecto AMPER. En este estudio trabajaremos con las frases con expansión en el sujeto,

1 Este trabajo ha contado con un ayuda de la SGPI del MCYT, ref. BFF2003-08487.

como por ejemplo *El capità protestant no porta el passaport* o *el copista intrèpid no porta la caputxa, etc.*, o frases con expansión en el objeto, como por ejemplo *El capità no porta el passaport pintoresc* o *La crítica no porta la caputxa llampegant...* Todas las palabras podrán ser agudas, llanas o esdrújulas, con la excepción del verbo que es llano siempre. Además de la modalidad declarativa, se estudia también dos tipos de interrogativas absolutas: unas encabazadas con la conjunción «que» que obligan a llevar al sujeto a la última posición, por ejemplo *Que no porta la caputxa petita el capità?* o *Que no porta la caputxa el capità intrèpid?*; y también las interrogativas absolutas sin «que» ni inversión del sujeto: *El capità intrepid no porta la caputxa?*

La grabación se hizo a una informante femenina (entre 25 y 50 años) sin estudios superiores que pertenece al subdialecto "tortosí".

Los análisis se han realizado con el programa matlab

3. RESULTADOS

Para describir los resultados nos iremos apoyando en unos gráficos obtenidos con el mismo programa de análisis en los que se representan las evoluciones de los tonos vocálicos a lo largo de su duración. La ordenada está expresada en semitonos. Se parte de la frecuencia media de la hablante que constituye el 0 semitonos y después se calculan las diferencias de cada tono con esa media y se coloca en el gráfico. Así, pues, tenemos una estandarización de la frecuencia que permite perfectamente la comparación de todas las frases y de todas las modalidades.

3.1. Declarativas con expansión en el sujeto

En general (Figura 1):

- aparecen tres picos en el pretonema coincidiendo con la sílaba postónica (en los proparoxítonos, en la segunda postónica)
- el pico más alto se asocia con el final del SN, que está constituido por la expansión. Sube entre 4 y 6 semitonos por encima de la media
- el tonema, sufre un descenso final de unos 8 semitonos, 6 desde la media
- no se observan grandes diferencias debidas a la estructura acentual de las palabras

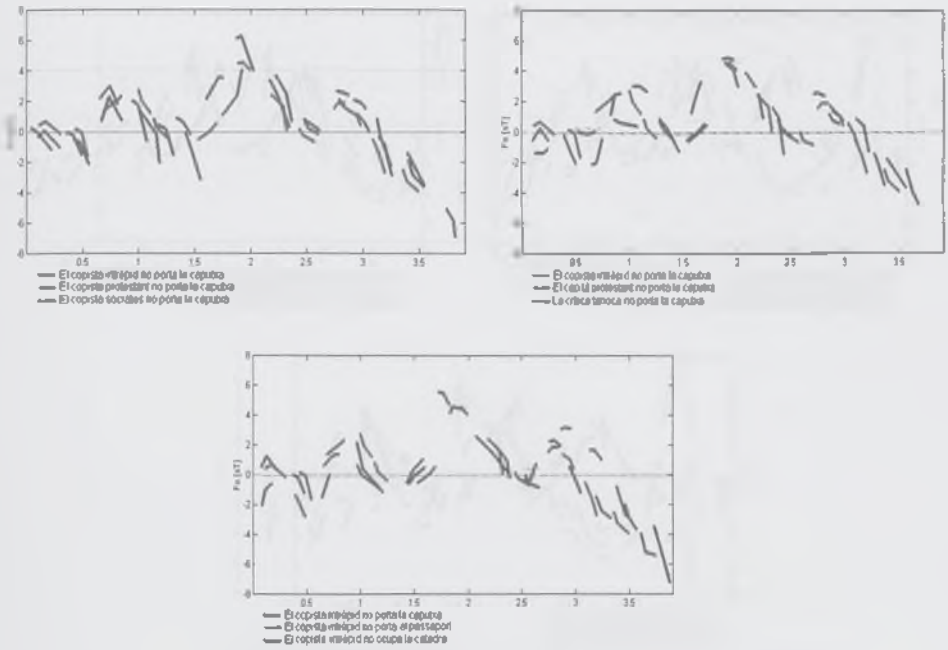


Figura 1. Declarativas con expansión en el sujeto

3.2. Declarativas con expansión en el objeto

En general (Figura 2):

- aparecen tres picos en el pretonema coincidiendo con la sílaba postónica (en los proparoxítonos, en la segunda postónica)
- el pico más alto no siempre se asocia con el final del SN. Sube unos 4 semitonos por encima de la media
- el tonema, sufre un descenso final de unos 5 semitonos, 4 desde la media
- no se observan grandes diferencias debidas a la estructura acentual de las palabras

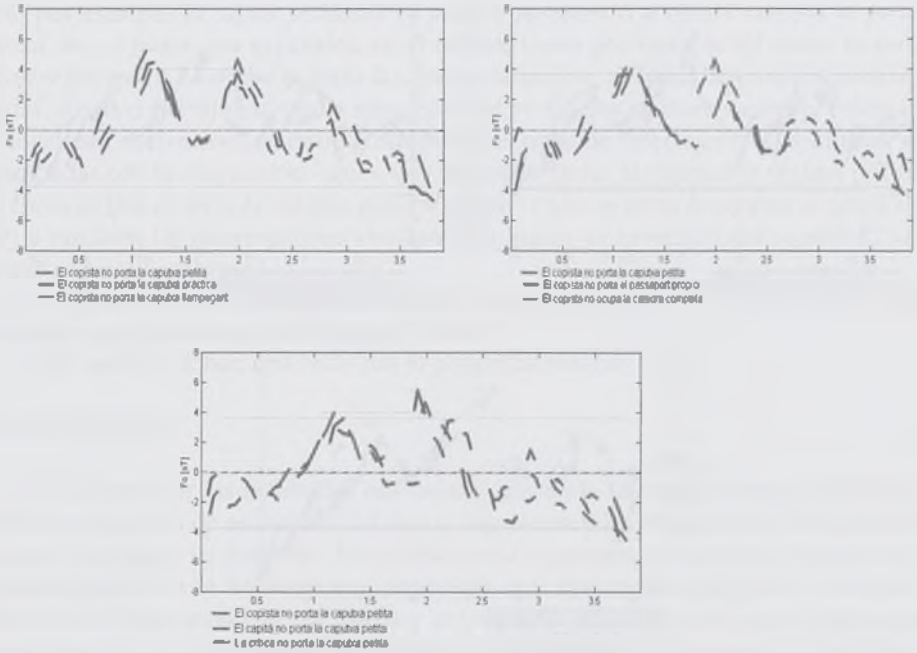


Figura 2 Declarativas con expansión en el objeto

3.3. *Interrogativas con expansión en el sujeto*

En general (Figura 3):

- aparecen dos picos en el pretonema coincidiendo con la sílaba postónica, el tercer acento tonal está desacentuado, corresponde al verbo.
- el segundo pico siempre se asocia con el final del SN. Sube unos 5 o 6 semitonos por encima de la media
- el tonema, sufre un ascenso final de unos 14 semitonos, 8 desde la media
- no se observan grandes diferencias debidas a la estructura acentual de las palabras

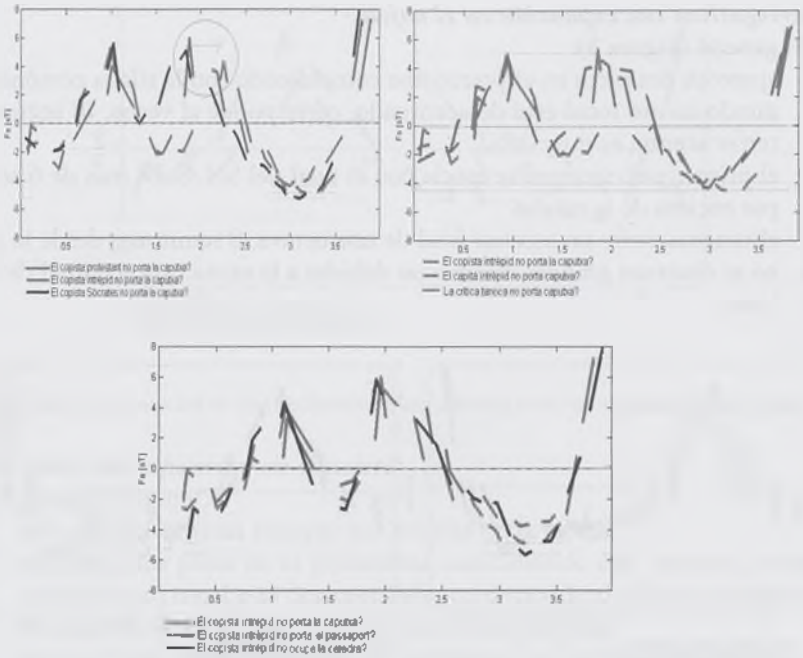


Figura 3 Interrogativas con expansión en el sujeto

Comparación de una declarativa (en el gráfico 4 sin cuadradito)y una interrogativa (en el gráfico 4 con cuadradito) con expansión en el sujeto y estructura acentual paroxítona. Las diferencias se sitúan en los dos últimos acentos. Pico en la declarativa, que está ausente en la interrogativa, en el penúltimo y caída en el tonema frente a la interrogativa, que posee un ascenso considerable.

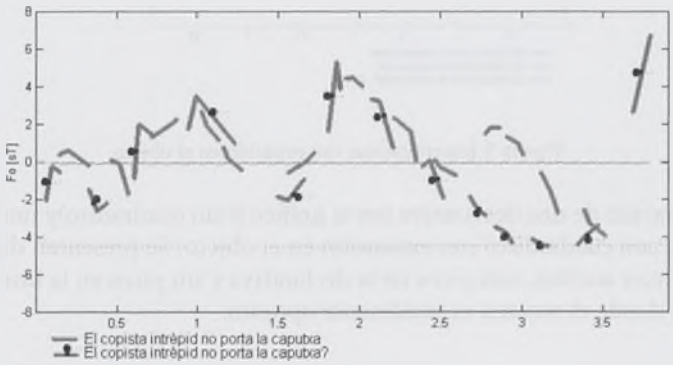


Figura 4 Comparación entre una declarativa y una interrogativa con expasión en el sujeto

3.4. Interrogativas con expansión en el objeto

En general (Figura 5):

- aparecen dos picos en el pretonema coincidiendo con la sílaba postónica, el segundo acento tonal está desacentuado, corresponde al verbo. El segundo pico, tercer acento, es muy débil.
- el primer pico siempre se asocia con el final del SN. Sube más de 6 semitonos por encima de la media
- el tonema, sufre un ascenso final de una octava, 8 semitonos desde la media
- no se observan grandes diferencias debidas a la estructura acentual de las palabras

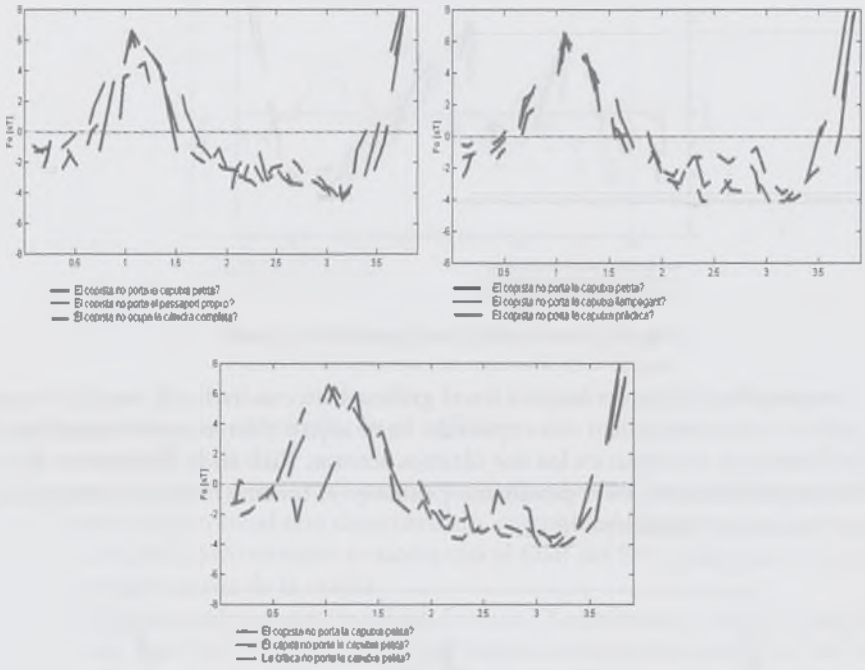


Figura 5 Interrogativas con expasión en el objeto

Comparación de una declarativa (en el gráfico 6 sin cuadradito)y una interrogativa (en el gráfico 6 con cuadradito) con expansión en el objeto. Se presentan diferencias en el segundo y el tercer acentos, con picos en la declarativa y sin picos en la interrogativa y en el tramo final, donde el tenema es totalmente opuesto.

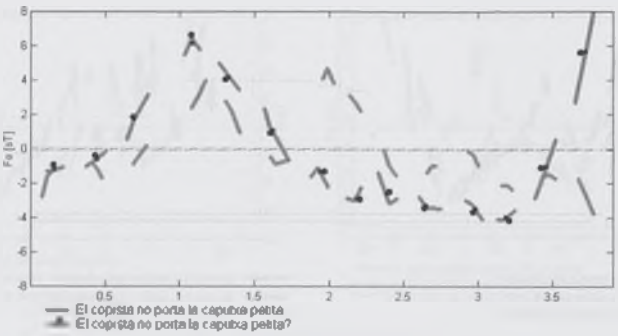


Figura 6 Comparación de una declarativa y una interrogativa con expansión en el objeto

3.5. «Int-que» con expansión en el sujeto

En general (Figura 7):

- el comienzo se sitúa siempre por encima de la media
- aparecen dos picos en el pretonema coincidiendo con la sílaba postónica, el tercer acento tonal está desacentuado, corresponde al núcleo del sujeto
- el segundo pico siempre se asocia con el final del OD
- los dos picos se sitúan entre 4 y 6 semitonos por encima de la media
- el tonema, sufre un ascenso final de una octava aproximadamente, entre 6 y 8 semitonos desde la media
- no se observan grandes diferencias debidas a la estructura acentual de las palabras

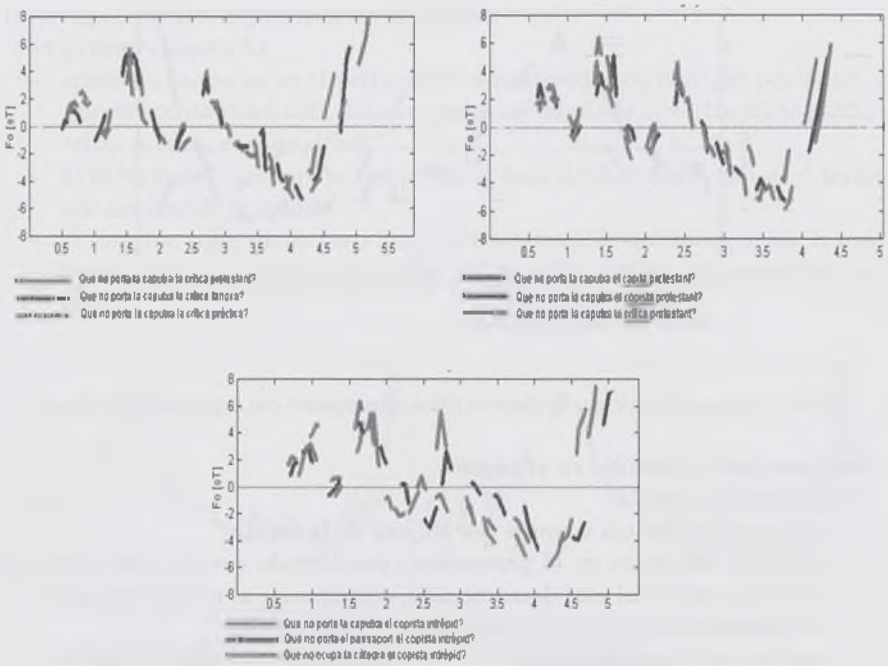


Figura 7 Interrogativas con «que» con expansión en el objeto

3.6. «Int-que» con expansión en el objeto

En general (Figura 8):

- el comienzo se sitúa siempre por encima de la media
- aparecen dos picos en el pretonema coincidiendo con la sílaba postónica, el segundo acento tonal está desacentuado, corresponde al núcleo del objeto.
- el tercer pico siempre se asocia con el final del OD, que es en este caso la expansión.
- el primer pico llega a 6 semitonos por encima de la media; el segundo a 4 semitonos
- el tonema, sufre un ascenso final de una octava aproximadamente, 8 semitonos desde la media
- no se observan grandes diferencias debidas a la estructura acentual de las palabras

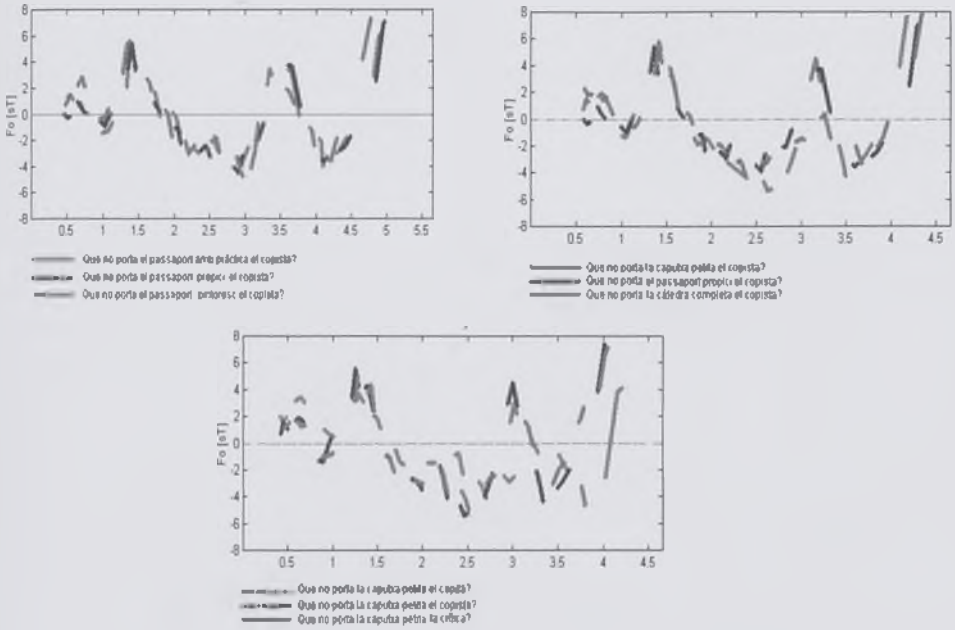


Figura 8 Interrogativas con "que" con expansión en el sujeto

En las interrogativas con «que» con extensión en el objeto se observa claramente que desacentúan el núcleo del objeto y aparece un pico destacado en la extensión; mientras que las que no poseen dicha extensión tienen ese pico en el núcleo del objeto; por tanto, el segundo pico destacado coincide siempre con el final del objeto directo (Figura 9).

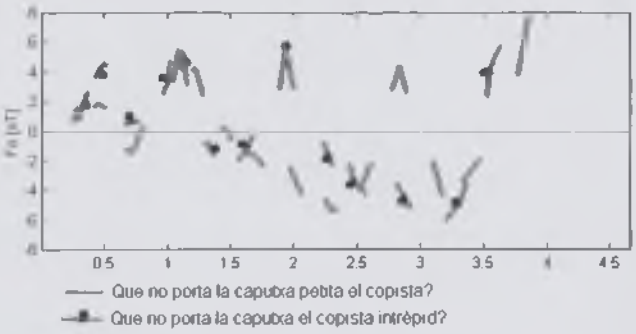


Figura 9 Comparación de interrogativas con "que": sin punto, expansión en objeto; con punto, sin expansión en objeto.

En la comparación entre una interrogativa con «que» y una sin que (Figura 10), aparte de la velocidad mayor en la «Int-que», la principal diferencia entre ambas se sitúa en el comienzo: la «Int-que comienza 3 semitonos por encima del comienzo de la interrogativa sin «que».

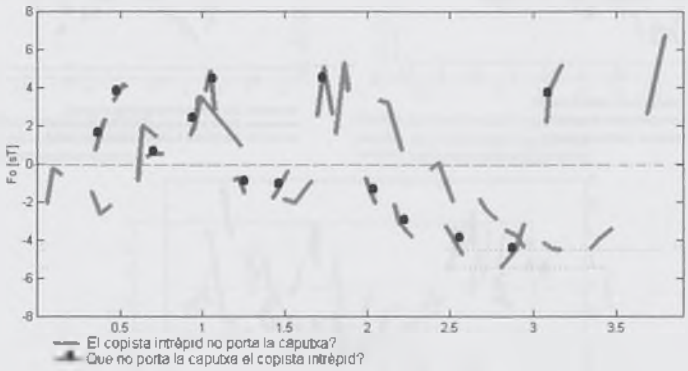


Figura 10 Comparación entre interrogativa sin «que» y con «que» e inversión

4. CONCLUSIONES

Inicio de la frase: «Int-que» comienza siempre por encima de la media del hablante; las demás siempre por debajo de esa línea media.

Primer acento: coincide casi siempre con la postónica del primer acento léxico de cada frase.

Segundo acento: hay pico en las declarativas, interrogativas con expansión en el sujeto y en las «Int-que» cuando el objeto no posee expansión.

Tercer acento: está presente siempre en las declarativas, es débil en la interrogativa con expansión en el objeto y fuerte en las «Int-que» con la expansión en el objeto.

Tonema: descendente en declarativas y ascendente en los dos tipos de intertogativas.

Hay una cierta congruencia con la sintaxis, por ejemplo, al marcar con un pico el final del objeto directo en la interrogativas con «que». Además se puede decir que desacentúan siempre la palabra que precede a la expansión: el núcleo del sujeto en el caso de las frases con expansión en el sujeto y el núcleo del OD en las frases en que la expansión aparece en el objeto. En las interrogativas sin «que», la palabra desacentuada es el verbo en en ambos casos.

La interpretación fonológica, según la teoría AM, es la que se muestra en la tabla siguiente, donde se resumen perfectamente las conclusiones:

	Inicio	*(SN +*	*Expansión)*	*Verbo*	*OD*
Dec-Ex-Suj	%L	L*H	L*H	L*H	H*L L%
Int-Ex-Suj	%L	L*H	L*H	-----	L*H H%
		SN	*Verbo*	*(OD +*	*Expansión)*
Dec-Ex-Obj	%L	L*H	L*H	L*H	H*L L%
Int-Ex-Obj	%L	L*H	-------	(L*H)	L*H H%
		Verbo	*OD*	*(SN +*	*Expansión)*
Int-que-Ex-Suj	%H	L*H	L*H	-----	L*H H%
		Verbo	*(OD +*	*Expansión)*	*SN*
Int-que-Ex-Obj	%H	L*H	------	L*H	L*H H%

REFERENCIAS BIBLIOGRÁFICAS

CARRERA SABATÉ, J.; VAN OOSTERZEE, C.; FERNÁNDEZ PLANAS, A. M.; ROMERA BARRIOS, L.; ESPUNY MONSERRAT, J. y MARTÍNEZ CELDRÁN, E. (2004): "Les interrogatives al *tortosí* i al *lleidatà*. Un element diferenciador de subdialectes", *Estudios de Fonética Experimental*, XIII. Barcelona: Laboratori de Fonètica UB, pp. 157-179.

MARTÍNEZ CELDRÁN, E; FERNÁNDEZ PLANAS, A.M; CARRERA SABATÉ, J. y ESPUNY MONSERRAT, J. (2005): "Approche à la carte prosodique dialectal de la langue catalane en Catalogne", *Projet AMPER. Géolinguistique-Hors série 3*. Grenoble: Centre de Dialectologie, Université Stendhal Grenoble 3, pp. 103-152.

PRADILLA, M. À. y P. PRIETO (2002): Entonación dialectal catalana: la interrogación absoluta neutra en catalán central y en tortosino. *Actas del II Congreso de Fonética Experimental* (2001). Sevilla: Universidad de Sevilla, pp.291-295.

PRIETO, P. y PRADILLA, M. À. (2003) Les inflexions finals en l'entonació del tortosí. M. A. Pradilla (ed.) *Llengua i literatura a les comarques de la diòcesi de Tortosa*. Becarló: Onada Edicions: 275-286.

VAN OOSTERZEE, C.; FERNÁNDEZ PLANAS, A. M.; ROMERA BARRIOS, L.; CARRERA SABATE, J.; ESPUNY MONSERRAT, J. y MARTÍNEZ CELDRÁN, E. (en prensa): Proyecto AMPER: estudio contrastivo de frases interrogativas sin expansión en tortosí y en lleidatà, *Actas del VI Congreso de Lingüística General*. Santiago de Compostela, 2004.

ON AUTOMATIC EUROPEAN PORTUGUESE SYLLABIFICATION

CATARINA OLIVEIRA
Center for Language and Cultures, University of Aveiro, Aveiro, Portugal coliveira@dlc.ua.pt
LURDES CASTRO MOUTINHO
Center for Language and Cultures, University of Aveiro, Aveiro, Portugal lmoutinho@dlc.ua.pt
ANTONIO TEIXEIRA
Dep. Electronics and Telecom./IEETA, University of Aveiro, Portugal ajst@det.ua.pt

INTRODUCTION

Motivated by our ongoing work on articulatory-based synthesis and our interest in developing a complete TTS[1] system based on this approach, we are developing the necessary linguistic modules.

Syllabification in TTS conversion is important for two reasons: first, it helps the implementation of certain letter-to-phoneme rules such as the diphthong generation in Arabic grapheme-to-phoneme conversion (usually named g2p) or the treatment of pharyngealization in Arabic phoneme-to-phone conversion (El-Imam, 2004); second, syllabification is essential in enhancing the quality of speech produced by synthesizers, since detecting the syllable will help in using them to model phone durations and as carriers of certain acoustic traits like intensity and duration to improve the synthesized speech intonation. Even for data driven approaches to grapheme-phone conversion, using for example pronunciation by analogy, it was recently demonstrated the usefulness of syllabic information (Marchand, 2005).

Syllabification is also useful in Automatic Speech Recognition (ASR). A first effort in using syllables for European Portuguese ASR was Meinedo (2000).

The usefulness of the syllabification information for the formulation of certain phonological rules, combined with the generally reported syllabic base of EP hyphena-

1 A TTS system is a program or a machine that automatically converts a written text into speech.

tion, motivated our explorations in the development of automatic syllabification methods having graphemes as input.

Automatic syllabification was approached recently using different theories and methods, such as: the use of optimality theory (e.g. Hammond, 1995) and the automatic acquisition of syllable grammars (Mueller, 2002).

The present work follows the line of linguistic research arguing that linguistic theories should be implementable and tested (Tatham and Morton, 2004).

The paper is structured as follows: the next two sections summarize knowledge about EP syllable and present, briefly, a syllabification algorithm proposed by two leading EP Phonology researchers; section 4 presents our implementation of the algorithm in two different versions, one for orthography input, the other for the sequence of phonemes; sections 5 and 6 present evaluation methodology and respective results; the final section presents the conclusions.

EUROPEAN PORTUGUESE SYLLABLES

In classic generative phonology theory, the syllable is not the basis for the application of phonological rules. However, for the later generative multilinear models, the syllable is an important linguistic unit, hierarchically organized. In this theoretical framework, the concept of syllable based on the binary branching model with rhyme (usually called "Onset-Rhyme" model) has proved to be productive and efficient in the European Portuguese description (Mateus and d'Andrade, 2000; Mateus, 1994). According to this perspective "syllable is a multidimensional object with an internal structure that has a hierarchical organization where the onset and the rhyme constitute a branching structure" (Mateus and d'Andrade, 2000).

Like this, a syllable (σ) branches in onset (O) and rhyme (R) and this last one branches in nucleus (N) and coda (C). Each syllabic constituent is associated to one or, in the maximum, two skeleton positions, represented by X in the syllabic tree. The segments of the lexical representations are associated to the skeleton level.

The onset can be simple (not branching) and be connected to a skeletal position (e.g. pé "foot"), or branching (complex) if it is connected to two skeletal positions (e.g. prato "dish"). Any PE consonant can occupy this position, but it's possible that segmental material isn't connected to the onset (empty onset - árvore "tree").

Regarding complex onsets, the consonantic groups constituted by stop+liquid and by fricative+liquid respect the sonority principle and dissimilarity condition, even though the first sequences are much more frequent (Vigário and Falé, 1993).

Other consonantic groups (e.g. pn – pneu "tire"; ps – psicologia "psychology"; pt – raptar "to kidnap"; ft - afta "aftae"; mn - amnésia "amnesia") clearly violate the universal principles. To explain these apparent violations, Mateus and d'Andrade (2000) propose the existence of subjacent empty nucleus. Like this, the consonantic groups are not considered real branching onsets, as the consonants now belong to different syllables, with the introduction of a vowel that destroys the consonantic group.

The same way the syllabic rhyme constituent may be simple (e.g. bola "ball"); or branching, when it is made up by the Nucleus and Coda (e.g. porta "door").

The nucleus may be associated to a minimum of one and the maximum of two positions in skeleton level. In EP, the syllabic nucleuses are always vowels. Like this, the nucleus can be filled by any vowel (non-branching nucleus - casa "house") or by a diphthong (branching nucleus - pai "father"). In this case, both elements of the falling diphthong belong to the nucleus (Mateus and d'Andrade, 2000).

The rising diphthongs (GV) create difficulties regarding the determination of the syllabic limit. The problem does not place itself in the base level (phonologic level) syllabification - the glide of the diphthong is a vowel in underlying structure and a syllabic boundary exists between the two elements of the diphthong (V.V) - but in the resyllabification, which results in the loss of the syllabic feature and semi-vocalization of the first diphthong element with the consequent join of the two syllables (GV). At the phonetic level, the glide integrates the rhyme or belongs to another syllable constituent? Some authors (Mateus, 1993; d'Andrade and Viana, 1993; Mateus and d'Andrade, 2000) propose that G must integrate the onset, once that its integration in the next phonetic nucleus would increase, in some cases (e.g. reauscultar "re-sound" – [RjawSkultar]), the number of segments allowed in the rhyme (the rhyme in EP should contain, in the maximum, three segments) and " the number of the elements in the onset is irrelevant for the maximum number of elements in the rhyme" (Mateus and d'Andrade 2000, p.51). Besides that, when the glide occurs before a nasal vowel or a nasal diphthong, it is not nasalized, contrary to what happens with falling diphthongs. This analysis has the disadvantages of creating tripositional onsets in words like criado ("created" [krjadu]) and generate CG onsets, that do not exist in any other context (Freitas, 1997). On the other hand, the acquisition information (Freitas, 1997) doesn't permit to support this analysis, as far as children associate the diphthong glide to the right adjacent vowel nucleus, on the contrary of processing it as part of a branching onset.

The phonetic alternation GV/VV is not verified, in some contexts which involve a semi-vowel [w] preceded by velar stop [k] or [g]. In this case, we can consider that: a) the structure is the only EP rising diphthong (Cunha and Cintra, 1997) and the glide belongs to a branching nucleus; b) the structure is associated to a non-branching onset, being the sequence velar consonant+glide a monophonematic unit, i.e. a labialized velar stop ([kw], [gw]) (d'Andrade and Viana, 1993; Bisol, 2001).

Another case that shows ambiguity in the syllabification process respects VGV structures (e.g. areia "sand"). In Vigário and Falé (1993) and d'Andrade and Viana (1993) various hypothesis of analysis are listed: 1) VG.V, in which G belongs to the nucleus and the second vowel is preceded by an empty onset; 2) V.GV, in which G is associated to the onset; 3) V_G_V, G is ambisyllabic, being connected to the first syllable nucleus and to the onset of the second, "a transition sound between the vowels" (d'Andrade e Viana, 1993, p.40). The most part of the studies consider G ambisyllabic (d'Andrade and Viana, 1993; Mateus and d'Andrade 2000), although under the perspective of the Optimality theory the VG.V and V_G_V syllabification be thought out as optimal candidates.

Regarding Coda, only three consonants, /l, s, r/, with its different realizations, can occupy this position (Mateus, 1994). In previous analysis it is admitted, although, the occurrence of complex Codas. Barbeiro (1986) considers the sequences /rs/,/ls/,/ns/, being the last one the most frequent. Mateus (1993) gives some examples of syllables with complex Codas (abstrair "to abstract"; perspectiva "perspective"), emphasizing the fact these sequences always present the C/s/ form and may happen, mainly in pre-stressed syllables, with non-branching nucleus. Afterwards, the author suggests another explanation for these structures, postulating the existence of empty nucleus between consonants considered in previous analyses as Coda. The consonant /s/ is also the only one able to follow a complex nucleus (claustro "cloister"), increasing to three the number of segments associated to the rhyme.

Statistic studies about the EP syllabic structure carried out in the 1990's (Vigário and Falé 1993; d'Andrade and Viana 1993) permit us to conclude that the CV structure is, by far, the most frequent in polysyllables as in monosyllables. The percentages confirm the tendency of EP language to the syllabic opening, greatly certified by various authors.

MATEUS AND D'ANDRADE ALGORITHM FOR EP SYLLABIFICATION

The syllabification process results from the application of conventions that try to give syllabic roles (rhyme level) to the segments (segment level).

Having in mind the formalization of this base syllabification process, Mateus (1994) and later Mateus and d'Andrade (2000), in the context of "Onset – Rhyme" syllable theory, proposed a rule based syllabification algorithm, considered more adequate to syllabic structure creation in EP than the template-matching approach.

According to the adopted theory each syllabic constituent is associated to at least one skeletal position.

The traditionally called 'all nuclei first' approach seems to be considered the most appropriate way to build up syllable structure in Portuguese. Under this perspective, the first step is the construction of rhymes, in accordance with the language's restrictions. The next conventions concern syllable onsets and codas.

The last version of the complete algorithm (Mateus e d'Andrade, 2000) is presented below.

1) Nucleus Association Convention

The first convention applied regards the nucleus, to which all X [-cons] (not lexically marked as not stressable and not preceded by another [-cons][2]) are associated.

The second part of the rule adjoins the rest of the X [-cons] to the adjacent nucleus leftwards. With the creation of the nucleus, automatically a rhyme is build up.

2) Onset Association Convention

The second rule implies an association of each X [+cons] preceding a nucleus to an onset.

A sequence of two consonants is integrated in the same onset, as long as it respects the sonority principle and the dissimilarity condition.

3) Empty Nucleus Creation Convention

The presence of extra-syllabic consonants (X [+ cons]) not associated to any constituent and not integrated in the syllabic structure gives rise to the creation of empty positions.

Thus, an empty nucleus to the left of the onset is created (e.g. pneu "tire" - pV.neu), with the corresponding position in the skeleton.

A re-application of the second rule allows the association of the consonants without position in the skeleton to an onset, now placed before an empty nucleus.

4) Empty Onset Creation Convention

In the case of a rhyme not being preceded by an onset, the creation of this last and the correspondent position in the skeleton line is demanded.

The authors assume that the syllable in the EP is obligatorily constituted by an onset and by a rhyme, although one of these constituents may be empty or have no phonetic realization.

5) Coda Association Convention

This rule allows associating the remaining [+cons] not integrated in the syllabic structure with coda of the preceding rhyme.

Previous work in automatic syllabification of EP

Opposing other languages such as English, in European Portuguese case, we found very few works about automatic syllabification of words.

In the scope of his PhD thesis, dedicated to the development of a speech synthesis system, based in rules, Oliveira (1996) defines a set of 11 rules, directly applied to orthographic level, aiming the automatic syllabification of words that is important for the prosody module. Despite the present tendencies recognize the advantages of syllabic information utilization, not only in prosodic modelling, but also in the improvement of recognition systems performance, the author believes that "in synthesis system perspective, the exactly definition of syllable boundaries wasn't extremely important" (p.89). In this way, his information about the implementation of the automatic syllabification system is very scarce.

Teixeira's work (2004) attempts to produce a prosody model for TTS systems in EP. Since is commonly accept by the specialists of prosody models that the syllable is useful in the determination of prosodic parameters, the author proposes an algorithm of

2 If a vowel is lexically marked as not being able to get the stress, cannot be a nucleus *per se* and is integrated in the nucleus leftwards. At the phonetic level, it is realized as a glide (e.g. *pai* "father") and integrates the nucleus. According with these theoretical principles, "in most of the falling diphthongs the phonetic glide is, phonologically, an underspecified vowel that has to be lexically marked as a trough" (Mateus, 1998).

syllabic splitting, conceived for application not only to the text written, but also to the sequence of phonemes produced in the locution of this text.

The rules applied aiming syllabic splitting of written text are based in the possible schemes of the Portuguese syllables (V, VC, CV, CVC,...). More complex cases demand complementary rules. The syllabic splitting of spoken text is also based in the principle that any syllable belongs to one of the syllabic types. However, due to the problems caused by vowel suppression phenomena an additional set of rules are needed.

Both versions of the algorithm were tested, using a corpora different from the one used in the development of the process. In both applications of the algorithm the error rate is very low: 0.06% (written text) and 0.89% (spoken text). The difference between error rates in the two versions is justified by difficulties related with the vowel suppression in spoken text.

The author regrets the impossibility of compare his results with another implementations, because he don't have knowledge of other published works about automatic syllabification with measured results.

Meinedo (2000) developed different methods of automatic syllable segmentation as a mean to improve the performance of continuous speech recognition systems. The results achieved show that the application of syllable segmentation information in speech recognizers improves their overall performance, opening good perspectives for the use of this kind of information in other contexts.

Linguateca[3], a resource center for computational processing of Portuguese language, offers their algorithm and his code to public utilization[4]. However, we notice a lack of written documentation and detailed information about it.

Implementation of Mateus and d'Andrade (M&A) algorithm

We implemented two versions of the algorithm: one handling orthography as input, the other capable of processing phone sequences. Being the second one a more direct implementation of the proposed algorithm, and similar in the adopted implementation to the first, we only present here information regarding the grapheme based algorithm.

Several steps, corresponding roughly to the M&A algorithm, implemented, for now, as separate programs, allow careful inspection of the results of each algorithm step.

The input for the first version of the algorithm is plain text. The output of the method and of all the several steps is structured in an XML document. XML provides a powerful, flexible and intuitive way to structure data.

XML processing is done in Perl using XML::DOM implementing the XML Document Object Model (DOM). Similar approaches were adopted in the development of TTS systems such as Mary (Schroder and Trouvain, 2003) or Prosynth (Ogden et al., 2000).

3 http://www.linguateca.pt
4 http://www.cpan.org

Step 0 - Nucleus

Having as input a list of words, a preliminary step creates a node in the hierarchy for each word. A second preliminary step processes each word, splitting it in graphemes and creating a sub-node for each. This sub-node is a potential syllable. To each potential syllable, in addition to the grapheme, several attributes are added regarding stress and consonantic and syllabic features. It is also in this step that grapheme groups are detected; being treated similarly to individual graphemes (e.g. double "r" is grouped and associated to only one potential syllable).

As an example the XML fragment (showing only the relevant information) generated for words claustro ("cloister") and afta ("aftae") is:

```
<WORD SPELLING="claustro"
<syllable>
<letter Stress="-" Consonantic="+" Syllabic="-"
PartSyllable="?">c</letter>
</syllable>
<syllable>
<letter Stress="-" Consonantic="+" Syllabic="-"
PartSyllable="?">l</letter>
</syllable>
<syllable>
<letter Stress="1" Consonantic="-" Syllabic="+"
PartSyllable="?">a</letter>
</syllable>
<syllable>
<letter Stress="-" Consonantic="+" Syllabic="+"
PartSyllable="?">u</letter>
</syllable>
<syllable>
<letter Stress="-" Consonantic="+" Syllabic="-"
PartSyllable="?">s</letter>
</syllable>
<syllable>
<letter Stress="-" Consonantic="+" Syllabic="-"
PartSyllable="?">t</letter>
</syllable>
<syllable>
<letter Stress="-" Consonantic="+" Syllabic="-"
PartSyllable="?">r</letter>
</syllable>
<syllable>
<letter Stress="-" Consonantic="-" Syllabic="+"
PartSyllable="?">o</letter>
</syllable>
</WORD>
```

```
<WORD SPELLING="afta"
<syllable>
<letter Stress="1" Consonantic="-" Syllabic="+"
PartSyllable="?">a</letter>
</syllable>
<syllable>
<letter Stress="-" Consonantic="+" Syllabic="-"
PartSyllable="?">f</letter>
</syllable>
<syllable>
<letter Stress="-" Consonantic="+" Syllabic="-"
PartSyllable="?">t</letter>
</syllable>
<syllable>
<letter Stress="-" Consonantic="-" Syllabic="+"
PartSyllable="?">a</letter>
</syllable>
</WORD>
```

Step 1 - Nucleus

After the two preliminary steps, the first step of M&A algorithm is performed, assigning the value Nucleus to the PartSyllable attribute of [+syllabic] nodes. This is performed in the two sub-steps proposed by the M&A algorithm.

Continuing with our examples:

```
<WORD SPELLING="claustro"
<syllable>
<letter Stress="-" Consonantic="+" Syllabic="-"
PartSyllable="?">c</letter>
</syllable>
<syllable>
<letter Stress="-" Consonantic="+" Syllabic="-"
PartSyllable="?">l</letter>
<syllable>
<letter Stress="1" Consonantic="-" Syllabic="+"
PartSyllable="NUCLEUS">a</letter>
<letter Stress="-" Consonantic="+" Syllabic="+"
PartSyllable="NUCLEUS_2">u</letter>
</syllable>
<syllable>
<letter Stress="-" Consonantic="+" Syllabic="-"
PartSyllable="?">s</letter>
</syllable>
<syllable>
<letter Stress="-" Consonantic="+" Syllabic="-"
PartSyllable="?">t</letter>
</syllable>
<syllable>
<letter Stress="-" Consonantic="+" Syllabic="-"
PartSyllable="?">r</letter>
</syllable>
<syllable>
<letter Stress="-" Consonantic="-" Syllabic="+"
PartSyllable="NUCLEUS">o</letter>
</syllable>
</WORD>
```

```
<WORD SPELLING="afta"
<syllable>
<letter Stress="1" Consonantic="-" Syllabic="+"
PartSyllable="NUCLEUS">a</letter>
</syllable>
<syllable>
<letter Stress="-" Consonantic="+" Syllabic="-"
PartSyllable="?">f</letter>
</syllable>
<syllable>
<letter Stress="-" Consonantic="+" Syllabic="-"
PartSyllable="?">t</letter>
</syllable>
<syllable>
<letter Stress="-" Consonantic="-" Syllabic="+"
PartSyllable="NUCLEUS">a</letter>
</syllable>
</WORD>
```

Step 2 - Onset

The second step of the algorithm creates onsets. In our implementation, first, the [+consonantic] before a nucleus are marked as being Onset in the PartSyllable attribute, second, the [+consonantic] before the recently created Onset are evaluated regarding the sonority principle. If the two have a difference of sonority high enough (Vigário and Falé, 1993), increasing toward the nucleus, this second node is marked as onset and associated to the syllable of the other onset.

Step 3- Coda

Third step processes remaining [+consonantic] marking the ones allowed to occupy coda position with the PartSyllable value of Coda. In the definition of both the table of sonority values, for second step, and the list of allowed codas, for the third step, we processed a list of around 100k words with these steps, monitoring the graphemes processed and decisions made. With this information, the table and list were corrected and completed.

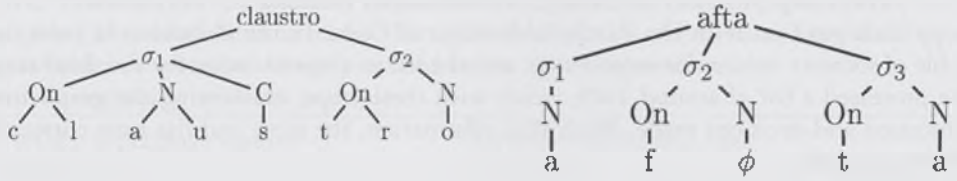
Step 4- Empty Nucleus

The final step deals with empty nucleus creation and finalization of the algorithm.

```
<WORD SPELLING="claustro"
<syllable>
<letter Stress="-" Consonantic="+" Syllabic="-"
PartSyllable="ONSET">c</letter>
<letter Stress="-" Consonantic="+" Syllabic="-"
PartSyllable="ONSET">l</letter>
<letter Stress="1" Consonantic="-" Syllabic="+"
PartSyllable="NUCLEUS">a</letter>
<letter Stress="-" Consonantic="+" Syllabic="+"
PartSyllable="NUCLEUS_2">u</letter>
<letter Stress="-" Consonantic="+" Syllabic="-"
PartSyllable="CODA">s</letter>
</syllable>
<syllable>
<letter Stress="-" Consonantic="+" Syllabic="-"
PartSyllable="ONSET">t</letter>
<letter Stress="-" Consonantic="+" Syllabic="-"
PartSyllable="ONSET">r</letter>
<letter Stress="-" Consonantic="-" Syllabic="+"
PartSyllable="NUCLEUS">o</letter>
</syllable>
</WORD>
```

```
<WORD SPELLING="afta"
<syllable>
<letter Stress="1" Consonantic="-" Syllabic="+"
PartSyllable="NUCLEUS">a</letter>
</syllable>
<syllable>
<letter Stress="-" Consonantic="+" Syllabic="-"
PartSyllable="ONSET">f</letter>
<letter Stress="-" Consonantic="+" Syllabic="-"
PartSyllable="NUCLEUS">V</letter>
</syllable>
<syllable>
<letter Stress="-" Consonantic="+" Syllabic="-"
PartSyllable="ONSET">t</letter>
<letter Stress="-" Consonantic="-" Syllabic="+"
PartSyllable="NUCLEUS">a</letter>
</syllable>
</WORD>
```

The automatically generated graphic representation for our two examples is:



EVALUATION

Evaluation Methodology

The evaluation of the two developed algorithms was performed in terms of the number of agreements in the syllable boundaries and the number of non-existent boundaries inserted, both expressed as percentages. For now, only syllable boundaries were evaluated, leaving a more detailed evaluation of each of the syllable constituents for later. These evaluation parameters were obtained automatically, using a specific developed program in Perl. The problematic words were manually analyzed and classified in terms of the error type, giving further insights in the methods weaknesses.

Corpora

Besides the development corpus (of 1006 words), for evaluation, and due to the non-public availability of a corpus having syllable boundaries and pronunciation for EP, we created two corpora. First, consists in 2076 common words, corresponding to a fraction of the so-called "Português Fundamental" corpus (Fundamental Portuguese corpus) (Nascimento et al., 1987).

Second, consists of 1303 words randomly selected from the Público corpus, created by the Portuguese Project Linguateca (see section "Previous work in automatic syllabification of EP") from the newspaper Público editions. This list contains words of longer length and higher complexity.

Results

The results of three evaluation metrics (percentage of correct syllable boundaries, percentage of insertions and word accuracy) for the 2 versions are presented in Table 1. The results are presented separately for the two corpora.

Table 1- Results (in percentage) of the evaluation of the syllabification methods against two manual segmented corpora. Table presents both the percentage of correctly positioned syllable boundaries (%Correct), the percentage of insertions (%Insertions),

and percentage of correctly syllabified words (%Word Acc)

Corpus		M&A Grapheme based	Phone based
Test1	%Correct	99.77	98.46
	% Insertions	0.08	0.26
	% Word Acc	99.57	97.88
Test2	%Correct	99.59	98.80
	% Insertions	0.15	0.03
	% Word Acc	98.85	96.47

Clearly the extension of Mateus e d'Andrade algorithm to process the words orthography attains a very good performance. Not only in terms of correctly detected syllable boundaries, but also by the high number of correctly syllabified words and low percentage of insertions. The phone-based version has a good but lower performance. Decrease affects particularly Word Accuracy.

Looking in more detail to the words incorrectly syllabified by each method: in our method based on graphemes several errors are due to error in stress assignment, prefixation, incorrect insertion of empty nucleus.

Results of a comparative evaluation of the method presented in this paper with another syllabification method previously developed by us, using finite state transducers (Oliveira et al., 2004), and the publicly available system previously mentioned (see section "Previous work in automatic syllabification of EP") was presented in Oliveira et al. (2005).

CONCLUSION

In this paper we presented two versions, for phonemes or graphemes as input, of a method for EP automatic syllabification, the implementation of an adapted version of Mateus and d'Andrade syllabification algorithm proposal. An evaluation of the performance of the developed methods results in a clear demonstration that the extension to grapheme input of the originally proposed algorithm for base syllabification was very successful.

Future developments of the syllabification method will address the lexical stress assignment and morphologic analysis of words, particularly to handle syllabification errors in prefixes. We are also looking at the more problematic areas of syllabification such as the [kw]/[gw].

Being the work presented motivated essentially by our efforts in the development of a complete text-to-speech system, incorporating the most recent knowledge

and linguistic theories, the new grapheme based syllabification method will be used to provide information to the g2p module, in phase of re-implementation. We intend to use the same approach, based on DOM and use of XML, in this new version of the g2p module.

ACKNOWLEDGEMENTS

The first author thanks Universidade de Aveiro and FCT (Portuguese Research Agency) the PhD scholarships supporting her work in Linguistic Processing for TTS systems. This line of research is funded in part by FCT project HERON (POSI/PLP/57680/2004).

BIBLIOGRAPHY

ALMEIDA, J., SIMÕES, A. and ROCHA, P., "Lingua-PT-PLN-0.06," 2003. Online available: http://www.cpan.org.

D'ANDRADE, E. and VIANA, M. C., "Sinérese, diérese e estrutura silábica", in Actas do IX Encontro da Associação Portuguesa de Linguística, Coimbra, Colibri, 1993.

BARBEIRO, L. F., Estrutura silábica do português. O papel da sílaba na análise dos processos fonológicos e fonéticos, Master Thesis, Lisboa, Universidade de Lisboa, 1986.

BISOL, Leda (org.), Introdução a estudos de fonologia do Português Brasileiro, Porto Alegre, 2001, 3ª Ed.

CUNHA, C., CINTRA, L., Nova Gramática do Português Contemporâneo, Lisboa, Edições João Sá da Costa, 13ª ed., 1997.

EL-IMAM, Y. A., "Phonetization of arabic: rules and algorithms", Computer Speech & Language, nº18, October 2004, pp. 339–373.

FREITAS, M. J., Aquisição da Estrutura Silábica do Português Europeu, PhD thesis, Universidade de Lisboa, 1997.

HAMMOND, M., Syllable parsing in English and French, Ms., Tucson, University of Arizona (ROA-58), 1995.

MARCHAND, Y. and DAMPER, R. I. "Can syllabification improve pronunciation by analogy?", Natural Language Engineering, 2005 (in press).

MATEUS, M. H. M. and D'ANDRADE, E., Phonology of Portuguese, Oxford, OUP, 2000.

MATEUS, M. H. e D'ANDRADE, E., "The Syllable Structure in European Portuguese", DELTA, vol.14, n.1, S. Paulo, Feb. 1998, pp.13-32.

MATEUS, M. H. M., "A silabificação de base em português", in Actas do X Encontro da Associação Portuguesa de Linguística, Evora, Colibri, 1994.

MATEUS, M. H. M., "Onset of portuguese syllables and rising diphtongs", in Proceedings

of the Workshop on Phonology, Coimbra, Colibri, 1993.

MEINEDO, H., Utilization of syllabic information in the recognition of continuous speech, Master thesis, Lisboa, Univ. Técnica de Lisboa/IST (in Portuguese), 2000.

MUELLER, K., Probabilistic syllable modeling using un-supervised and supervised learning methods, PhD thesis, Stuttgart, University of Stuttgart, Institute of Natural Language Processing (IMS), 2002.

NASCIMENTO, F., MARQUES, L. and SEGURA, L., Português fundamental: Métodos e documentos, Lisboa, Tech. Rep., INIC-CLUL, 1987.

OGDEN, Richard et al., "ProSynth: an integrated prosodic approach to device independent, natural-sounding speech synthesis", Computer Speech and Language, 14, pp.177–210, 2000.

OLIVEIRA, L. C., Síntese de Fala a Partir de Texto, PhD Thesis, Lisboa, University of Lisbon, 1996.

OLIVEIRA, C., PAIVA, S., MOUTINHO, L. and TEIXEIRA, A., "Um novo sistema de conversação grafema-fone parao Português Europeu baseado em transdutores," in Actas do II Congresso Internacional de Fonética e Fonologia, 2004 (in press).

SCHRÖDER, M. and TROUVAIN, J. "The German text-to-speech synthesis system MARY: A tool for research, development and teaching" in Intl. J. Speech Technol., vol. 6, 2003, pp. 365–377. Online available: http://mary.dfki.de.

TATHAM, M. and MORTON, K., Speech Synthesis: Developments, Chichester, Wiley, 2004.

TEIXEIRA, J. P., A Prosody Model to TTS Systems, PhD Thesis, Faculdade de Engenharia da Universidade do Porto (FEUP), 2004.

VIGÁRIO, M. and FALÉ, I., "A sílaba no português fundamental: uma descrição e algumas considerações de ordem teórica", in Actas do X Encontro da Associação Portuguesa de Linguística, Coimbra, Colibri, 1993.

OBSERVACIONES SOBRE LA ESTRUCTURA MELÓDICA EN ENUNCIADOS DECLARATIVOS

ANTONIO PAMIES BERTRÁN
Universidad de Granada

1. INTRODUCCIÓN

Frente a la idea consagrada, tanto en la tradición estructuralista como en la fonología métrica autosegmental, según la cual los picos entonativos corresponden a sílabas acentuadas, se ha objetado que la alineación entre pico y acento resulta muy discutible en español (cf. p.ej. Canellada & Madsen 1987, Pamies 1996). Dentro de la propia corriente autosegmental, trabajos empíricos sobre la entonación española señalan el *overshooting* (L*H) como el esquema mayoritario frente a la alineación (Sosa 1995; 1999:142; Prieto 1995; Face 2000, 2002, 2003), y los primeros trabajos realizados por miembros del grupo AMPER coinciden bastante en este punto (p.ej. Martínez Celdrán & Fernández Planas 2002; Martínez Celdrán et al. 2003; Amorós 2003& 2004). El modelo fonológico de Philippe Martin (1987, 1995, 1997) relaciona en cambio la estructura entonativa de una frase con su composición sintáctico-semántica (cf. Cresti, Martin & Moneglia 2002; Boulakia, Deulofeu & Martin 2002). P.ej., en francés, los movimientos tonales, sean ascendentes o descendentes, culminarían en el final de un componente sintáctico: la segmentación prosódica viene determinada por la sintáctica (**fig.1**):

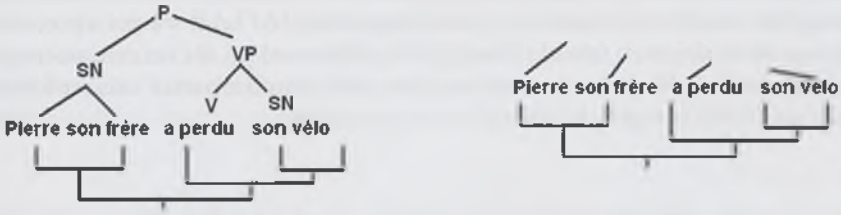


Fig.1 (Martin 1995:647)

Martin localiza aun así el pico tonal en la sílaba tónica (1995:644), pero esta precisión no es imprescindible para que se cumpla el principio esencial del sistema: es la palabra y no la sílaba lo relevante (por otra parte, en francés estándar, tónica y final de palabra coinciden). El problema de las postónicas en otras lenguas románicas, donde el tono sigue subiendo tras el acento, sólo afectaría a la realización fonética para Martin (1999a; 1999b), y no a la estructura lingüísticamente relevante que sería básicamente compartida por varias de estas lenguas.

Un experimento sobre el catalán, realizado por Estebas-Vilaplana (2003), abordó específicamente esta cuestión, la comprobada discoincidencia entre pico y sílaba tónica no se interpreta como L*H, porque este concepto no permite predecir a qué distancia de la tónica aparece el pico. Si el pico coincide con la tónica en palabra aguda, se "retrasa" una sílaba en llana, dos en esdrújula y tres en sobresdrújula, es que tal vez tiene más poder predictivo el final de palabra que el acento (Estebas 2003:105). En un trabajo anterior sobre la entonación granadina (Pamies & Amorós 2005), también hemos criticado la taxonomía *ToBI* por su carácter cíclico: ninguna combinación bitonal posible queda excluida del modelo teórico, con lo que la dependencia entre tono y acento resulta inverificable por definición.

Desde el punto de vista empírico, los últimos trabajos de *AMPER-España* (2003, 2004), comprobaron, en varios dialectos del castellano, del gallego y del asturiano[1], que los cambios de distribución de los acentos (en el pretonema) no provocan ningún desplazamiento de los picos tonales, lo que probaría la independencia de ambos factores. El pico tonal en el pretonema parece alinearse más bien con el final de palabra o sintagma. La idea que Estebas defendió para el catalán bien podría ser extrapolable a otras lenguas románicas peninsulares, no sólo resultaría mucho más simple que el sistema ToBI, sino que tendría una capacidad predictiva de la que éste carece por completo. El llamado "acento tonal" sería entonces metalingüísticamente innecesario además de fonéticamente inexacto.

2. OBJETIVOS

La prudencia recomienda aun así asegurarse de que la propia metodología empleada por todo el equipo de AMPER no haya influido excesivamente en sus resultados. Haciendo de abogado del Diablo contra nuestro propio método, se podría cuestionar, por ejemplo, que nuestro corpus fijo contenga palabras demasiado cultas y abstractas como para propiciar una lectura "natural", que el programa *MATLAB* sea poco preciso en sus mediciones automáticas, o que el centro exacto de la vocal no sea necesariamente el más idóneo para medir la F0 de los tonos circunflejos, cuya cumbre parece más representativa. Para salir de dudas, nos pareció necesario un experimento que estuviera libre de estos po-

1 Dorta & Hernández 2004; Calleja Azpiazu 2004; Elejabeitia, Iribar & Pagola 2005; Zamora, Carrera & Meléndez 2005; Castañer, González & Simón 2005; Ramírez Verdugo 2005; Fernández; Escourido, Caamaño & Xuncal 2005; López, González, Cuevas, Díaz & Muñiz 2005.

sibles reparos metodológicos para confirmar (o criticar) con más seguridad nuestro propio trabajo, complementando las observaciones aportadas al respecto por Martínez Celdrán y Fernández Planas (2005). Si los resultados son aun así similares, no sólo se confirmaría nuestra hipótesis, sino que se demostraría que la metodología común de AMPER es suficientemente adecuada como para no sesgar los resultados.

3. EXPERIMENTO

El corpus está calcado del *corpus fijo* de AMPER en cuanto a la estructura de las frases: son oraciones declarativas de tipo N_1+V+[prep+N_2], con variantes de acentuación aguda, llana y esdrújula tanto en el sujeto como en el predicado, a las que se añaden expansiones del sujeto (N_1+Adj) y del predicado (N_2+prep+N_3), también con sus 3 variantes acentuales (el verbo no cambia). El contenido léxico se ha modificado para evitar las palabras cultas y los significados abstractos. En total son 126 frases (21 frases x 3 tomas x 2 hablantes). El corpus ha sido el siguiente:

a) oraciones sin expansión
los [/el] rábanos [/tomates /limón] se compra(n) en Cáceres [/ Sevilla /Badajoz]

b) oraciones con expansión en el sujeto
los (/el) rábano(s) [/tomates /limón] pálido(s) [/temprano(s)/ español] se compra en Cáceres

c) oraciones con expansión en el predicado
los tomates se compran en Cáceres (/ Sevilla /Badajoz] y en Motril

Las grabaciones se realizaron con DAT (micro *Shure* SM-58), con tres tomas por hablante, captadas con el programa *Goldwave* 4.25 y tratadas con *Matlab* para extraer la F_0 en los puntos iniciales, centrales y finales de cada vocal. Estas grabaciones también fueron segmentadas y analizadas "manualmente" con *Speech Analyzer*, para verificar y eventualmente corregir las tablas automáticas generadas por *Matlab*. En los tonos circunflejos, la medición de F0 se ha realizado en la cumbre tonal y no en el centro exacto de la vocal (a diferencia de *Matlab*).

Finalmente, para comprobar la influencia del factor dialectal, se han analizado paralelamente dos dialectos muy diferentes en vez de centrarse en uno. Las hablantes son dos jóvenes, de edad y nivel sociocultural similar pero con dialectos muy diferentes. La primera es andaluza (C.R.), procedente de Órgiva (Granada), la segunda es colombiana (Y.O.), de Pradera (Valle del Cauca). Se añadió un pequeño test perceptivo sobre el factor dialectal.

4. RESULTADOS Y DISCUSIÓN

4.1. Experimento acústico

4.1.1. Sujeto sin expansión

Como podemos ver en la estilización de la curva tonal (media de las tres grabaciones), el pico tonal sólo se alinea con el acento en palabra aguda, donde éste coincide con un final de palabra y de sintagma[2]. Hay variación entre los hablantes en la palabra aguda: para la hablante colombiana el pico se sitúa incluso a la derecha del límite sintáctico (en el clítico *se*) (fig. 2).

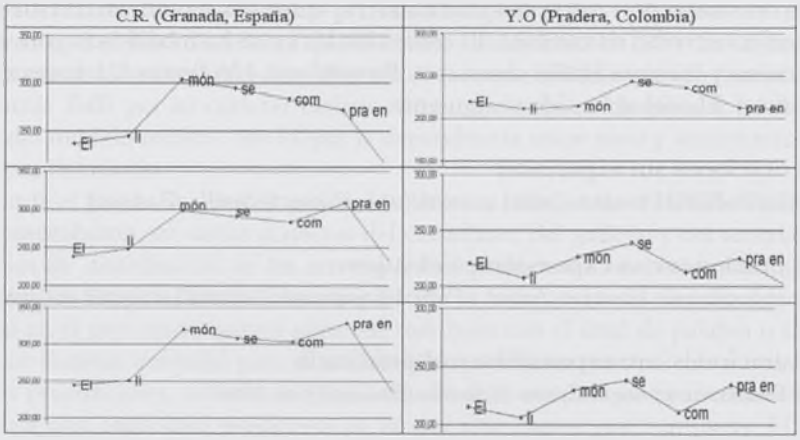


Fig.2

En cambio, si el sujeto es una palabra llana o esdrújula, el pico tonal sigue coincidiendo con el límite sintáctico-semantico (final del sintagma sujeto), no con el acento (fig.3 & 4).

2 El gráfico corresponde a la estilización de la media de las 3 tomas para *x se compra en Cáceres/ Sevilla/ Badajoz.*

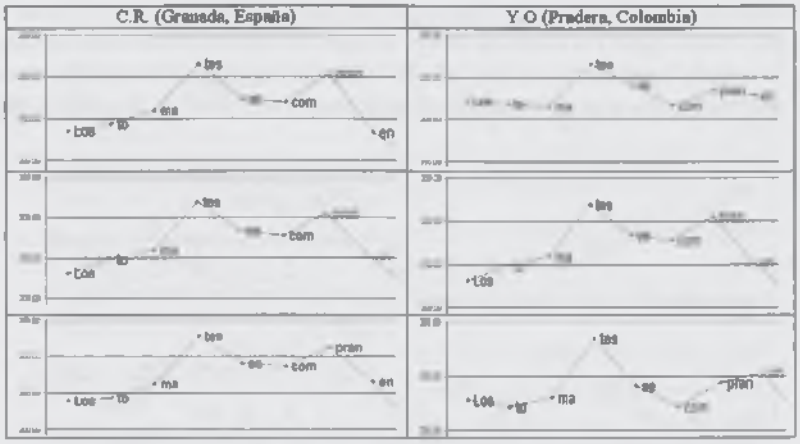


Fig 3

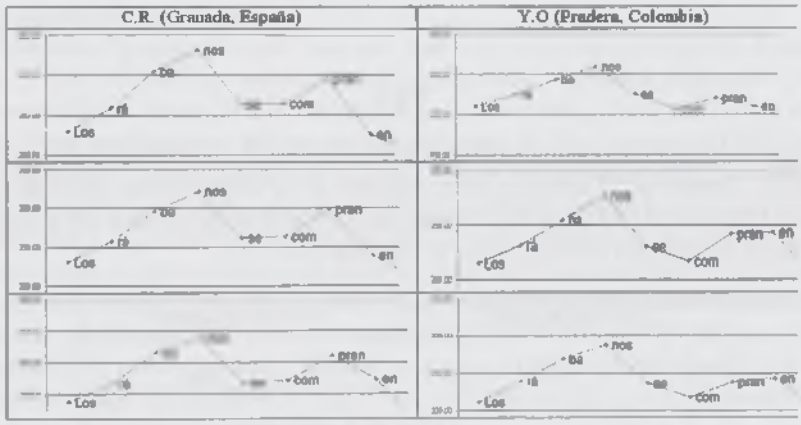


Fig.4

4.1.2. Sujeto con expansión

Si observamos el adjetivo que ahora ocupa el final del sintagma sujeto, vemos que se comporta exactamente como en lo que acabamos de ver para el nombre sujeto sin expansión. El pico tonal sólo coincide con el acento si el adjetivo es agudo (*español*). La única variación al respecto, entre ambas hablantes, es que en una de las tres frases de la hablante colombiana, la subida tonal continúa y recae al otro lado del límite sintáctico, el clítico *se*, tal como ya hemos visto que ocurría en el sujeto nominal agudo sin expansión.

Si observamos el sustantivo *limón*, que ahora se sitúa en el interior del sintagma sujeto, el pico tonal sigue sin alinearse con el acento -ni siquiera tratándose de una palabra aguda-, y la subida tonal continúa después del acento, cosa que no ocurre con el sustantivo llano y el esdrújulo, que siguen teniendo su pico tonal en el final de la palabra (fig.5).

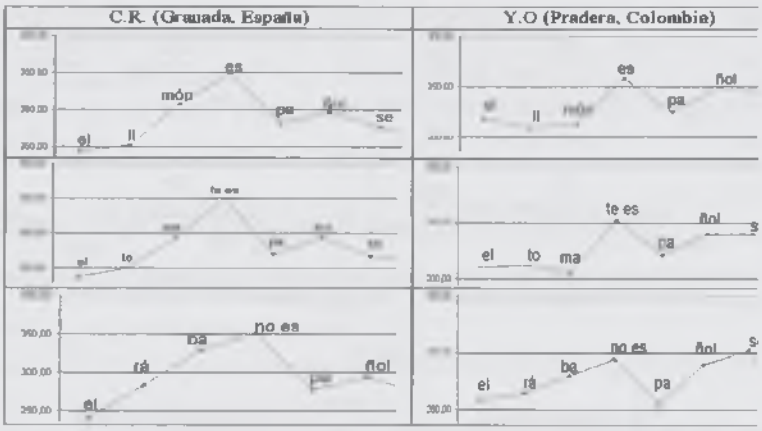


Fig.5

Cuando el adjetivo es una palabra llana (*temprano*[*s*]) o esdrújula (*pálido*[*s*]), el pico tonal ignora claramente el acento y recae siempre en la última sílaba (final de palabra y de sitagma) (fig.6 & 7). Hay menos regularidad del pico en el interior del sintagma que en su final en el esquema entonativo de la frase declarativa.

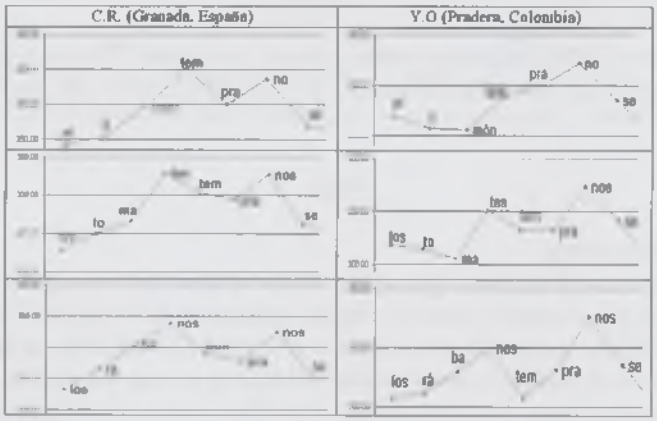


Fig.6

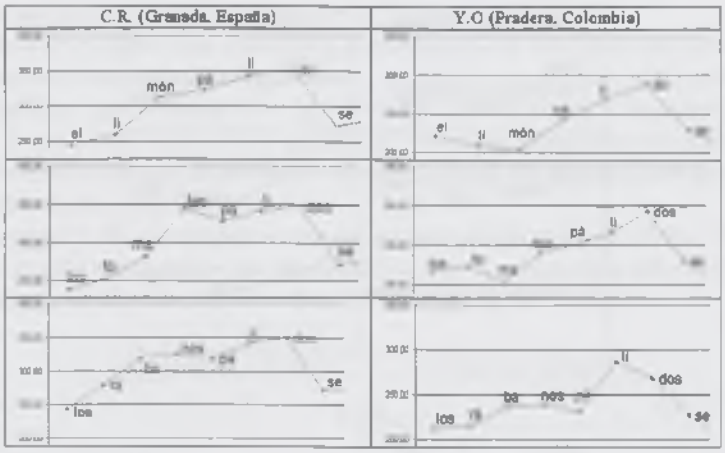


Fig.7

4.1.3. Predicado sin expansión

Cuando el predicado es de una sola palabra, ésta ocupa la posición fonológica de tonema en frase declarativa neutra. Como corresponde al tonema afirmativo español, el contorno global es descendente. Sin embargo, observamos cierta influencia del acento que da lugar a tres variantes del tonema descendente. En final agudo, la tónica interrumpe la caída general con un un levísimo ascenso tonal en la tónica final. Lo mismo hace el tonema llano en la penúltima, aunque el desnivel tonal es ínfimo y no se observa en todas las realizaciones. En la esdrújula, obviamente, la caída tonal de la frase arrastra a la palabra, cuya tónica no interrumpe para nada un descenso que empieza en el clítico que está a su izquierda, aunque también se observa un repunte final del tono en las esdrújulas de la hablante colombiana. Por tanto, el tonema sí parece sufrir cierta influencia del acento en la entonación, con "piquitos" dudosos en cuanto a su percepción[3], pero cuya presencia en la producción es bastante recurrente y -en tonema esdrújulo- no tiene exactamente el mismo perfil en ambos dialectos (**fig. 8, 9 & 10**).

3 No olvidemos que, en español, el umbral perceptivo mínimo de una diferencia tonal está en 1,5 semitonos (Pamies, Fernández Planas, Martínez Celdrán, Ortega Escandell & Amorós et al 2002).

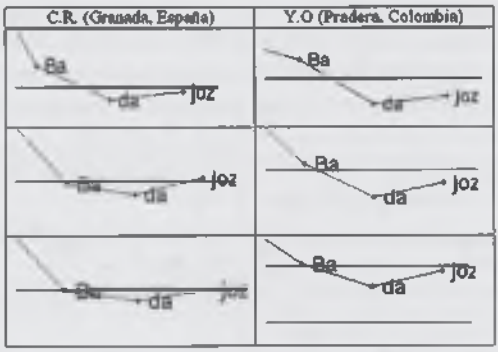


Fig.8

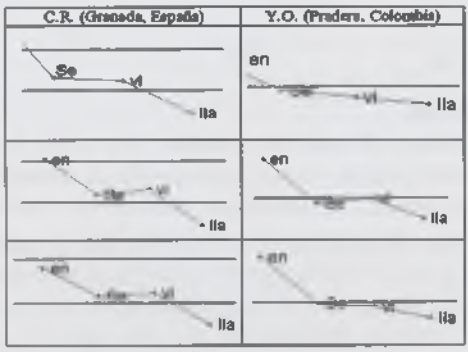


Fig.9

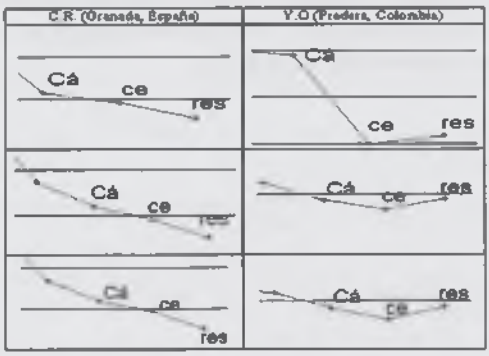


Fig.10

4.1.4. Predicado con expansión

Cuando se le añade una expansión al predicado, su núcleo ya no forma parte del tonema y su contorno entonativo se invierte: se vuelve ascendente, independientemente de que sea palabra aguda, llana o esdrújula[4]. El pico tonal vuelve a disociarse claramente del acento, buscando la frontera de palabra al igual que ocurría en el sujeto. Es la expansión la que ahora adopta el perfil descendente propio del tonema, con el mencionado "piquito" en la tónica (fig.11).

En resumen, la correlación entre los picos tonales y las fronteras sintáctico-semánticas es mucho más estable que su alineación con el acento, que sólo se realmente en palabra aguda (y no siempre), que es donde ambos criterios coinciden.

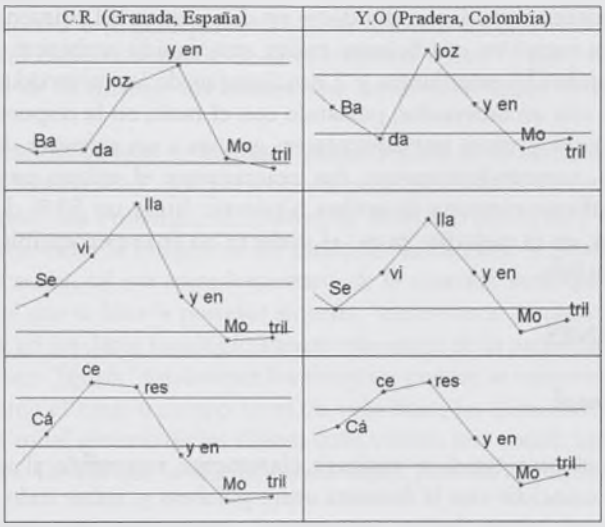


Fig.11

4.2. Test perceptivo

La diferencia entre dialectos es discretamente apreciable en los esquemas analíticos. Sólo cabe señalar que, en el pretonema, tras palabra palabra aguda, el tono sigue subiendo en el clítico siguiente en la hablante colombiana; y que, en tonema esdrújulo, se aprecia que la átona final es un poco más alta que la penúltima sólo en la hablante colombiana. Ésta también presenta a menudo un tono inicial descendente en la primera sílaba de la frase. Pero en lo esencial, ambos dialectos parecen coincidir. El test perceptivo pretende calibrar mejor el estatus de la variación dialectal. Para ello usamos una señal re-

4 El verbo no entra en el experimento puesto que es acentualmente invariable (*se compra*[*n*]) por analogía con el corpus fijo de AMPER, donde siempre es *se toca*.

sintetizada a partir de la voz real (fonemas irreconocibles que sólo conservan la prosodia). En principio, si los entrevistados reconocen su dialecto oyendo sólo la melodía "pura" será señal de que los rasgos prosódicos dialectales están entre los que se mantienen en la resíntesis. Si los oyentes no reconocen su dialecto, será señal de que los elementos que permiten reconocerlo están entre los eliminados por la síntesis.

Elevamos todo el tono global en una de las hablantes para que no se distingan las tesituras por sus picos máximos. Veinte frases del corpus fueron mezcladas aleatoriamente (10 de cada hablante). Los entrevistados debían contestar ***sí***, ***no*** o ***no lo sé*** a la pregunta: ***¿es granadina esta entonación?*** tras oir cada frase. En el recuento, la respuesta *no lo sé* computa lo mismo que un error, su función es evitar posibles respuestas al azar (que habrían contado con 50% de posibilidades de acertar por mera casualidad). El test fue sometido a 27 oyentes granadinos, divididos en dos grupos. El primero (12 estudiantes) oyó las frases (dos veces) en condiciones reales, con sonido ambiente y contestando en un papel. El segundo (13 estudiantes y 2 funcionarias de la universidad) fue examinado individualmente, con un ordenador, pulsando con el ratón en la respuesta elegida (*sí, no, ?*) y oyendo las frases 2 veces por auriculares, gracias a un software diseñado para este fin[5]. El resultado, sorprendentemente, fue exactamente el mismo para ambos grupos, y, además, no confirma ninguna de ambas hipótesis: hubo un 51% de aciertos, lo que parece indicar que, en la melodía "pura" el dialecto no es imperceptible pero tampoco es fácilmente perceptible.

5. CONCLUSIONES

5.1. Nivel funcional

Dado que el pico tonal se muestra claramente insensible al acento, y que, en cambio, tiende a coincidir con la frontera entre palabras y, sobre todo, entre sintagmas (prácticamente en el 100% de los pretonemas), parece que hay elementos razonablemente suficientes como para proponer la hipótesis provisional de una posible **función demarcativa** de la entonación en español, como alternativa a la teoría autosegmental. Siguiendo la propuesta de Philippe Martin, podemos identificar un esquema fonológico para la entonación de la frase asertiva, que sería linear y binario[6] y basado en las fronteras de sintagma (fig.12). Si descendemos al interior del sintagma, el mismo esquema demarcativo se podría conservar, aunque los datos ya no lo confirman con tanta claridad, pues la segmentación prosódica es menos nítida entre palabras de un mismo sintagma.

---

5 © Javier Parra 2005.

6 No tenemos en cuenta pequeños desniveles tonales en el interior del tonema dada la discutible perceptibilidad de los mismos.

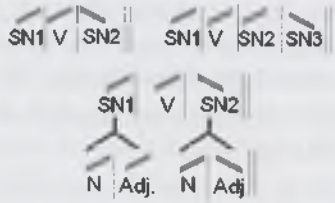


Fig.12

5.2. Nivel dialectal

El resultado del test perceptivo (51% de aciertos en la distinción) indica cierta perceptibilidad del dialecto a partir de la entonación "pura"[7], pero tambén demuestra la gran dificultad de la misma: los errores son demasiado numerosos para ser casuales (49%)[8].

Una explicación de esta "dificultad" puede residir en el estilo de habla aquí analizado (corpus de laboratorio), que presenta en principio menos variación diatópica que el habla espontánea. Sin embargo, el trabajo de Martínez Celdrán y Fernández Planas (2005) relativiza bastante la influencia del estilo de habla sobre la prosodia.

Otra explicación, tal vez complementaria de la anterior, sería que análisis realizado con *Matlab*, y en el que se basa la posteror síntesis, "desfavorece" la distinción entre dialectos porque se basa en los datos fonológicamente relevantes de la prosodia[9]. Pero parece que este análisis tampoco "iguala" totalmente los dialectos porque se conservaron algunos datos fonéticos, tales como el *range* o campo tonal de cada frase, las duraciones e intensidades, y algo del contorno tonal interno de las sílabas (tres valores por vocal), aparte de los niveles fonológicos tonales. De ahí que la distinción sea difícil pero no imposible.

BIBLIOGRAFÍA

AMORÓS CÉSPEDES, M.C. 2003 *La relación entre entonación y acento en español*. Tesis doctoral (inédita) dirigida por Antonio Pamies. Dept. de Lingüística General, Universidad de Granada.

AMORÓS CÉSPEDES, M.C. 2004 "Sincronización entre pico tonal y acento: resultados según posición métrica y morfológica", *Estudios de Fonética Experimental*, XIII, pp. 203-224.

7 Hubo una oyente que obtuvo 18 aciertos sobre 20.

8 Hubo dos oyentes con sólo 4 aciertos.

9 La razón de ser del empleo de curvas entonativas estilizadas en secuencias binarias alto/bajo fue precisamente la extracción de los rasgos lingüísticamente relevantes frente a los expresivos, idiolectales, dialectales, etc. (Rossi et al. 1981, t'Hart, Collier & Cohen 1990, Garrido 1991; Llisterri et al. 2003).

BOULAKIA, G.; DEULOFEU, H.J.; MARTIN, Ph. 2002 "Prosodic features finish off ill-formed French utterances, do'nt they?", in Díaz García, J. (ed.), *Actas del II Congreso Nacional de Fonética Experimental.* Sevilla: Universidad; pp. 107-112.

CALLEJA AZPIAZU, N. 2004 "Alineamiento fonético de acentos tonales", *Estudios de Fonética Experimental*, XIII, pp. 39-64.

CANELLADA, M.J. & MADSEN, J.K. 1987 *Pronunciación del español. Lengua hablada y literaria*. Madrid: Castalia.

CASTAÑER, R.M.; GONZÁLEZ M.P. & SIMÓN, J. 2005 "Aproximación al estudio de la entonación aragonesa", *Estudios de Fonética Experimental*, XIV pp. 273-294. (en prensa).

CRESTI, E.; MARTIN, Ph. & MONEGLIA, M.: 2002 "L'intonation des illocutions naturelles représentatives: analyse et validation perceptive», Firenze: *LABLITA*, 3 [consulta, julio 2005] (disponible en http://lablita.dit.unifi.it/italian/preprint.html).

DORTA, J & HERNÁNDEZ, B. 2004 "Prosodia de las oraciones SVO declarativas e interrogativas en el español de Tenerife", *Estudios de Fonética Experimental*, XIII, pp. 225-274.

ELEJABEITIA, A. IRIBAR, A. & PAGOLA, R.M. 2005 «Notas sobre la prosodia del castellano en Bizkaia», *Estudios de Fonética Experimental*, XIV pp 247-272. (en prensa).

ESTEBAS-VILAPLANA, E. 2003 «The modelling of prenuclear accents in Central Catalan Declaratives», *Catalan Journal of Linguistics*, 2: 97-114.

FACE, T.L. 2000 *Intonational markings of contrastive focus in Madrid Spanish.* Berlin: Lincom-Europa.

FACE, T.L. 2002 "Spanish evidence for pitch-accent structure" *Linguistics*, 40/2, pp. 319-345.

FACE, T.L. 2003 "Intonation in Spanish declaratives: differences between lab speech and spontaneous speech", in Prieto, Pilar (ed.) *Romance Intonation* [(monografía) *Catalan Journal of Linguistics*, 2], pp. 115-132.

FERNÁNDEZ, E.; ESCOURIDO, A.; CAAMAÑO, M & XUNCAL, L. 2005 "A entoación dunha fala de santiago: fronteira prosódica entre suxeito e predicado. *Estudios de Fonética Experimental*, XIV pp. 141-166. (en prensa).

FERNÁNDEZ PLANAS, A.M.; MARTÍNEZ CELDRÁN, E.; SALCIOLI, V.; TOLEDO, G. 2001 "Taxonomía autosegmental en la entonación del español peninsular", *Actas del II Congreso Nacional de Fonética Experimental*. Sevilla: Universidad, pp. 180-186.

FERNÁNDEZ PLANAS, A.M. & MARTÍNEZ CELDRÁN, E. 2003 "El tono fundamental y la duración: dos aspectos de taxonomía prosódica en dos modalidades de habla del español", *Estudios de Fonética Experimental*, XII. pp. 165-200.

LÓPEZ, M.J.; GONZÁLEZ, R.; CUEVAS, M.; DÍAZ, L. & MUÑIZ, C. 2005 "Rasgos prosódicos del centro de Asturias: comparación Oviedo-Mieres", *Estudios de Fonética Experimental*, XIV pp. 167-200. (en prensa).

LLISTERRI, J.; MACHUCA, M.J.; MOTA, C. de la; RIERA, M. & RÍOS, A. 2003 " Entonación y tecnologías del habla ", in Prieto, P. (ed.) *Teorías de la entonación*. Barcelona: Ariel, 209-243.

MARTIN, Ph. 1987 "Prosodic and Rhythmic Structures in French", *Linguistics*, 25/5, 925-949.

MARTIN, Ph. 1995 "The prosody of Mauritian Creole: some experimental aspects", Stockholm: *ICPhS95*, vol.3: pp. 644-647.

MARTIN, Ph. 1997 "L'intonation: analyse instrumentale et modeles», Firenze: *LABLITA*, preprint 4. (http://lablita.dit.unifi.it/papers/97coll04.pdf)

MARTIN, Ph. 1999a «L'intonation du français et du portugais: phonétique et phonologie», Revista de l'ANPOLL (Sao Paulo), vol 5/7, pp. 255-295.

MARTIN, Ph. 1999b «Prosodie des langues romanes: analyse phonétique et phonologique», *Recherches sur le français parlé* (Aix-en-Provence), vol. 15, pp. 233-253.

MARTÍNEZ CELDRÁN, E.; FERNÁNDEZ PLANAS, A.M. & FULLANA RIVERA, N. 2003 «Pre-Nuclear Tonal Inventories of Spanish Intonation» *15th ICPhS*. International Phonetic Association & Universitat Autónoma de Barcelona.

MARTÍNEZ CELDRÁN, E. & FERNÁNDEZ PLANAS, A.M. 2002 «Taxonomía de las estructuras entonativas de las modalidades declarativa e interrrogativa del español estándar peninsular según el modelo AM en habla de laboratorio», *Actas del I coloquio internacional de tonía*. México.

MARTÍNEZ CELDRÁN & FERNÁNDEZ PLANAS, A.M. 2005 «Estudio metodológico acerca de la obtención del corpus fijo en el proyecto», *Estudios de Fonética Experimental*, XIV pp.29-66. (en prensa).

PAMIES BERTRÁN, A.: 1996 «Consideraciones sobre la marca acústica del acento fonológico», *Estudios de Fonética Experimental*, VIII, 11-49.

PAMIES BERTRÁN, A. & AMORÓS CÉSPEDES, M.C. 2003 "The relation between stress and tonal peaks", in Solé, M.J.; Recasens, D. & Romero, J. (eds.): *15th ICPhS*. International Phonetic Association & Universitat Autónoma de Barcelona.

PAMIES BERTRÁN; A. & AMORÓS CÉSPEDES, M.C. 2005 "Pico tonal, acento y fronteras morfo-semánticas: experimento con hablantes granadinos" *Estudios de Fonética Experimental*, XIV pp. 201-224 (en prensa).

PAMIES BERTRÁN, A.; FERNÁNDEZ PLANAS A.M.; MARTÍNEZ CELDRÁN, E.; ORTEGA ESCANDELL A.; AMORÓS CÉSPEDES, M.C.: 2002 "Umbrales tonales en español peninsular", en Díaz García, J. *Actas del II Congreso Nacional de Fonética Experimental*. Sevilla: Universidad, pp. 272-278.

PRIETO, P. 1995 "Tonal alignment patterns in Spanish", *Journal of Phonetics*, 23: 429-451.

PRIETO, P. (ed.) 2003a *Romance Intonation*. Bellaterra: Universitat Autónoma de Barcelona.

PRIETO, P. (ed) 2003b *Teorías de la entonación*. Barcelona: Ariel.

QUILIS, A. 1981 *Fonética acústica de la lengua española*. Madrid: Gredos.

RAMÍREZ VERDUGO M.D. 2005 "Aproximación a la prosodia del habla de Madrid", *Estudios de Fonética Experimental*, XIV pp. 309-326. (en prensa).

ROSSI, M.; DI CRISTO, A.; HIRST, D.; MARTIN, PH. & NISHINUMA, Y. 1981 *L'intonation. De l'acoustique à la sémantique*. Paris: Klincksieck

SOSA, J.M. 1995 "Nuclear and pre-nuclear tonal inventories and the phonology of Spanish declarative intonation", *ICPhS 95*, Stockholm, vol. 4, pp. 646-649.

SOSA, J.M. 1999 *La entonación del español*. Madrid: Cátedra.

ZAMORA, F.J.; CARRERA, M. & MELÉNDEZ M.M. 2005 "Variación geoprosódica en dos entonaciones de Castilla y León: análisis de frases declarativas e interrogativas sin expansión", *Estudios de Fonética Experimental*, XIV pp. 125-140. (en prensa).

IS INTRINSIC F0 SEGMENTAL OR SUPRASEGMENTAL? A CONTEXTUALLY INDUCED INVERSION OF THE VOWEL HEIGHT - F0 HIERARCHY

Olivier Piot
CNRS – University Paris III olivier.piot@gmail.com

1. INTRODUCTION

It is well known, and widely accepted, that the fundamental frequency of vowels is significantly higher for high vowels than for their low counterparts. This effect has been reported in more than 30 languages, and is found either in isolated vowels, isolated words, and in words in a carrier sentence [1]. Several explanations have been proposed for this phenomenon, which are either physiological, acoustical or cognitive (see [2] for a survey). Nevertheless, it has been shown that the cricothyroid (CT) muscle activity is increased in correlation with vowel height [3] [4] [5]. And since CT is known to be a major performer in raising F_0 [6] [7], intrinsic F_0 (IF_0) could well be the mere consequence of a difference in laryngeal muscles activities. But no satisfactory explanation has been proposed for such hypothesis. To begin with, the explanation by a deliberate acoustical enhancement of the vowel height dimension, which was proposed by [8] [9] [10], is contradicted by a number of studies. Firstly, it appears incompatible with the fact, advocated by [1] to be universal, that IF_0 disappears in the low end of a speaker's F_0 range, as well as at the end of utterances [12]. It is more than doubtful that such acoustical enhancement would be less useful in the low range of speakers, or at the end of their utterances. Secondly, IF_0 is found in the babbling of prelinguistic infants [11], suggesting that it is an automatic consequence of articulation. Thirdly, the EMG evidence cited in favor of IF_0 being deliberate is contradicted by [13], which shows that the (EMG, F_0) relationship is significantly different between isolated high and low vowels.

Besides this seemingly unintelligible hypothesis of IF_0 being the consequence of differing laryngeal muscles activities, the most convincing explanation is certainly the one known as the tongue-pull hypothesis [14]. It assumes that for high vowels, the

tongue exerts a pulling action on the larynx through some of the laryngeal tissues, with the consequence of tensing the vocal folds, and thus of raising F_0. This hypothesis, in addition to its explanatory value, is in accord with the apparent automaticity of IF_0. But it is also contradicted by some facts. Firstly, it stands in contradiction with the absence of IF_0 at the bottom of a speaker's F_0 range [1], or near the end of utterances [12]. Secondly, tongue height has been shown not to be a good predictor of IF_0 in German, where "tense" and "lax" vowels have similar IF_0 values, while tongue height is noticeably higher for the former than for the latter [2].

Clearly, new data is needed to develop this discussion any further. And since current explanations for IF_0 appear to be contradicted by various experimental studies, the way should remain as open as possible to alternative explanatory hypotheses. In [15] was advocated the existence of trade-off relationships between muscular activities responsible for prosodic parameters such as F_0, speech rate, but also amplitude of articulatory gestures. Under this hypothesis, F_0 can be expected to decrease if the speech rate or articulatory amplitude increases, all else being equal. While this expectation is met when comparing the F_0 of /a/ and /i/ in /CV/ syllables (the articulatory amplitude is higher and F_0 is lower for /Ca/ than for /Ci/, whatever the consonant C), no phonemic context has been used so far where /i/ would be produced with a bigger articulatory amplitude than a /a/. More precisely, the question is: what would happen if we compared the F_0 of /i/ and /e/ in /aiai/ and /aeae/ sequences? The tongue-pull hypothesis, as well as the "deliberate planning" hypothesis, predict that the IF_0 hierarchy is left unchanged in such context (i. e. F_0(/i/) > F_0(/e/)). Conversely, a trade-off effect would involve F_0(/e/) to be closer to, or bigger than F_0(/i/) in such a context. In order to measure IF0 in this new phonemic context, we carried out an experiment comparing the F_0 of the three French unrounded front vowels /i/, /e/ and /ɛ/, in the contexts /dVdV/ and /aVaV/.

2. METHOD

The experiment to be described was conducted in order to investigate the effect of articulatory amplitude, and speech rate, on IF_0. In this latter purpose, we wanted the variations of speech rate to be elicited by the context of communication. More precisely, we wanted to avoid a higher speech rate condition to be realized through emotional/stressful speech, i. e. with a higher arousal, leading to an augmentation of all prosodic parameters (as in the expression of anger or joy [16]). We thus used three contexts of communication, where the speaker was answering a question asked by the experimenter. In the first condition, the speaker was to answer the question in a "neutral" (or "natural") way, using her own casual speaking style. In the second condition, the hearer was supposed to be a native American, not speaking French very well: the speaker was asked to use a lower speech rate to facilitate the perception of the foreign hearer facing him/her. In the third condition, the speaker was requested to use a high speech rate, because she could do so - in part because the hearer was supposed to be a native speaker of French.

The stimuli were nonsense words, /didi/, /dede/, /dɛdɛ/ in the carrier sentence: "*il a dit /dVdV/ déjà*" ("he already said /dVdV/"), and /aiai/, /aeae/, /aɛaɛ/ in the carrier sentence: "*il a dit /aVaV/ à haute voix*" ("he said /aVaV/ in a loud voice"). The carrier sentences were chosen in order that the phoneme immediately following the target word was the one alternating with V (i. e. /d/ in the first case, /a/ in the second). This choice was motivated by the fact that this would extend the tongue movement cycle to the syllable following the target word, thus preventing the influence of anticipatory coarticulation on the realization of the target syllable. The latter was chosen to be the second syllable of the target word, to limit the possible influence of coarticulation with the preceding word. For each of the three speech rates (normal, high and low), each target word was uttered 7 times. A block of 6 sentences, corresponding to the 6 target words, was thus presented 7 times, each time in a randomized order, and insuring a regular alternation of /dVdV/ and /aVaV/ target words across the recording. The experiment was divided into 3 parts, each part corresponding to one speech rate. The first part was always the one with a neutral speech rate, and the order of the remaining two parts (high and low speech rate) was swapped across subjects.

Eight speakers, four male (M1, M2, M3 and M4) and four female (W1, W2, W3 and W4), were recorded for this experiment. They are all students in linguistics, monolingual native speakers of French, with French parents, and having always lived in the same "prosodic area" (and mainly in Ile de France). All speakers were first explained the task both visually (through reading an instruction sheet) and verbally. They were asked to answer the experimenter's question, instead of just reading the answer. This made the task simpler. Indeed, when the experimenter was supposed to be an American hearer, the experimenter used an American prosody, which freed the speaker from having to remember that he/she had to speak as if the hearer was a foreigner not very familiar to French language. This made the interaction more realistic and easy. In the high speech rate condition, the experimenter asked the questions using a high speech rate, just as was expected from the speaker, which again made the task easier and less artificial. Indeed, it was a way of reproducing the corresponding situation of communication, i. e. one between two natives chatting in a low noise environment.

The recordings took place in an anechoic room, using an Audiotechnica professional microphone positioned about 20 cm from the speaker's mouth, and was made on a DAT recorder. Each sentence was written alone on a plasticized sheet of paper, each one following another in a file. The questions asked by the experimenter was "*Il a dit quoi déjà?*" ("what did he already say?") in the /dVdV/ case, and "*Il a dit quoi à haute voix?*" ("what did he say in a loud voice?") in the /aVaV/ case, so as to elicit a focus on the new element of information, i. e. the target word. Speakers were made aware that in the communication context which was simulated, the target word could be /dadi/ or /dide/ all the same, so that both syllables were potentially informative (and the second syllable was not inferable from the first). The expected intonation was a hat pattern on the target word, which they accepted as a natural way of answering these particular questions in this particular context. No pause was allowed between the target word and the

end of the utterance, because of its potential influence on the articulatory rendering of the target syllable.

The speakers were successful overall in following the instructions. However, a very few times, an utterance was "wrongly" accepted, both by the experimenter and by the speaker. This mistake was in the use of an inconsistent intonation pattern, due to an audible intra-word downstep, or from a pause inserted just after the target word. In such a case, the utterance was excluded from the measurements. The number of excluded tokens, over a total number of 126 per speaker, was 0 for W2, M2, and W4, 2 for W3, 3 for M3, 6 for W1 and M1, and 7 for M4.

Before the recording, some training was made on a few examples, to make sure that the speaker felt comfortable with the task, and that she was performing it properly. During the recording session, whenever the experimenter or the speaker felt the experimental task was not fully respected, the recording (including the preceding question) was performed again, until an acceptable performance was produced. Such "faux pas" could be caused by a word error, an inconsistent intonation (an intra-word downstep for instance), an unwanted pause, or by some background noise.

The recordings were numerically transferred from DAT to wave files, at the original format (44.1 kHz, 16 bits). The second syllable of each target word was phonemically labeled by hand, using the PRAAT program [17]. F_0 was taken at the center of the vowel, and was measured using the standard autocorrelation function. Each time an unexpected value of F_0 was found, it was corrected by hand, and was always due to a miscalculation at one octave below the actual F_0 value.

3. RESULTS

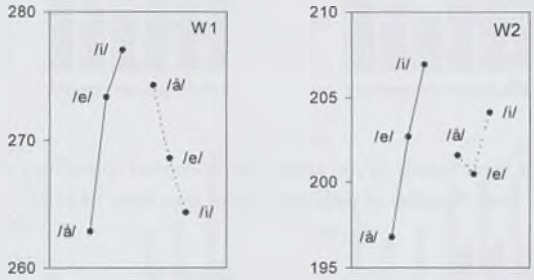
The mean F_0 values for each speaker, averaged over the three speech rates, are shown on Fig. 1. It can be seen that for the /dVdV/ target words, the classical F_0-aperture relationship is globally found for each subject, except for M1 who, on average, does not display any noticeable IF_0 differences. For the /aVaV/ target words, IF_0 differences appear to be either attenuated or neutralized (subject W2, W3, M2 and M3), or even reversed (subjects W1, and maybe partially W4, M1 and M4). To check whether these tendencies are significant or not, several analyses of variance (ANOVA) were performed, using F_0 as the dependent variable, and context and vowel category as independent variables. Paired one-tail t-tests were also used to assess the influence of vowel category on F0 within the /aV/ context (i. e. for target words of the type /aVaV/). The choice of this test comes from the structure of the experience (series of randomized blocks of all 3 vowels × 2 contexts = 6 items).

The effect of context (/aV/ or /dV/) on IF_0 is confirmed for all speakers except M1, by the aperture-context interaction term of a two-factor (aperture × context) ANOVA, as can be seen in table I. In the /dV/ context, the IF_0 hierarchy is also respected for all subjects [$F(2,55 \leq n \leq 60) \geq 5.5$, $p < 0.01$] except M1 [$F(2,55) = 0.48$, $p > 0.6$]. This means that M1 does not have any significant IF0 effect, neither in the /dV/-context nor in the

/aV/-context. This also shows that all 7 other speakers display a "classical" IF0 effect in the /dV/-context, and also a significant "reversing influence" of the /aV/-context on IF0. In particular, IF0 is *reversed* for W1 in the /aV/-context [$F(2,56) = 5.67$, $p < 0.01$], and this reversed hierarchy is attested for each of all three pairs of vowels ($p < 0.05$). This is, we believe, the first time that such an IF_0 inversion is found. IF0 is also reversed for W4 in the /aV/-context [$F(2,60) = 10.45$, $p < 0.0001$], although the reversed hierarchy only concerns 2 pairs, as it does not extend to the /e/-/ɛ/ pair ($0.1 < p < 0.2$). As mentioned above, M1 does not show any IF_0 effect in /a/-context [$F(2, 59) = 3.00$, $0.05 < p < 0.1$], however his F_0(/e/) and F_0(/ɛ/) are significantly reversed with respect to F_0(/i/) in this context ($p < 0.01$).

Speaker	F	p
W1	$F(2,120) = 18.44$*	$p < 0.0001$
W2	$F(2,126) = 4.65$*	$p < 0.02$
W3	$F(2,124) = 21.37$*	$p < 0.0001$
W4	$F(2, 126) = 12.69$*	$p < 0.0001$
M1	$F(2,120) = 2.29$	$0.1 < p < 0.2$
M2	$F(2,126) = 5.85$*	$p < 0.01$
M3	$F(2,123) = 10.89$*	$p < 0.0001$
M4	$F(2,119) = 6.34$*	$p < 0.01$

TABLE 1. The aperture-context interaction term of a two-factor (aperture × context) ANOVA, with F_0 as the dependent factor.



MEAN DURATION (ms)
MEAN F_0 (Hz)

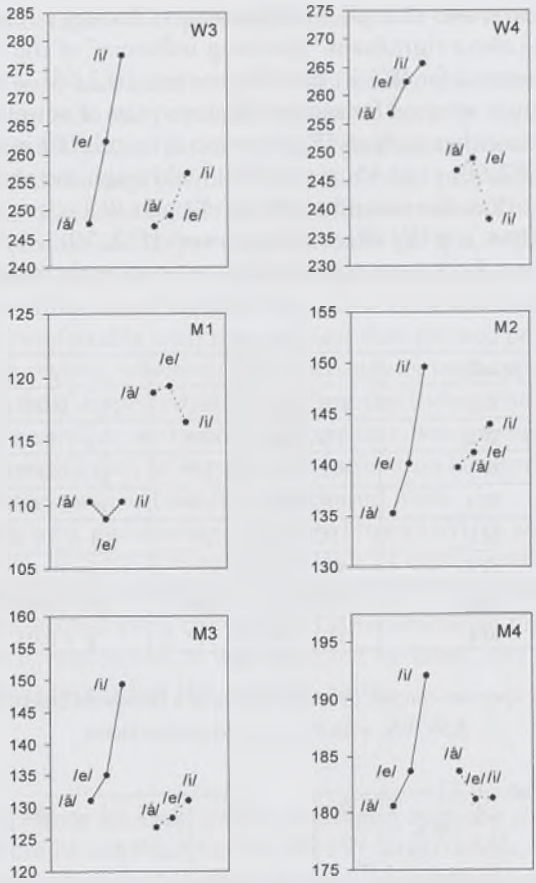


FIG. 1. Mean F_0 for the three vowels /ɛ/, /e/ and /i/, in /d/-context (plain line) and /a/-context (dotted line). Number of tokens per point goes from 18 to 21.

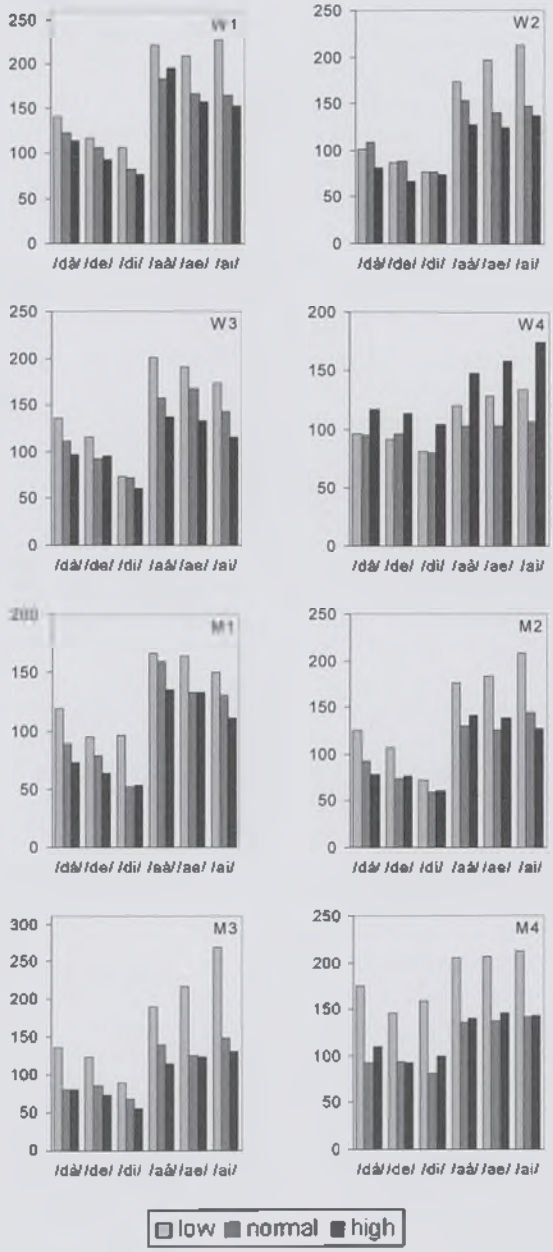


FIG. 2. Mean vowel duration for the three vowels /ɛ/, /e/ and /i/, in /d/- and /a/-contexts. Number of tokens per point goes from 4 to 6.

For subjects M2, M3, M4 and W2 the "classical" IF_0 hierarchy found in /d/-context does not prevail (it is "neutralized") in /a/-context [$F(2,60) = 3.11$, $0.05 < p < 0.1$; $F(2, 57) = 2.27$, $0.2 > p > 0.1$; $F(2, 58) = 0.50$, $0.6 < p < 0.7$; $F(2, 60) = 1.46$, $0.2 < p < 0.3$ (resp.)]. It only prevails for W3 [$F(2,58) = 8.38$, $p < 0.001$], so the classical IF0 effect is only *attenuated* for this particular speaker. All speakers therefore have a common tendency toward inversion of the IF_0 hierarchy in /a/-context, even if an actual inversion is found only for speakers W1 and (except for the /e/-/ɛ/ pair) W4, and to a lesser extent for M1.

The influence of phonemic context on the IF_0 hierarchy thus appears to be notable and robust. Whether this effect can be considered, or not, as related to the amplitude of the articulatory gestures involved in the realization of vowels, will be discussed in the following section. However, another potential influence on the F_0 of vowels is the speech rate. The mean vowel durations are shown on Fig. 2.

It can be seen that while mean vowel duration is noticeably increased under the low speech rate condition (except for speaker W4), the distinction between normal and high speech rates is far less clear-cut. For this reason speech rate here cannot be simply thought of as divided into three categories, and should rather be regarded as gradient. Therefore, linear regressions are to be preferred to ANOVAs. However, just by looking at figures 1 and 2, it is clear that vowel duration alone does not have any systematic correlation with F0, even when considering the /aV/ context alone (the differences found in the /dV/ context come - at least partially - from the aspiration noise at the start of higher vowels).

An important point to consider is that, for the higher speech rates, there is a well-known effect of target "undershoot", making the articulation of an /ai/ sequence, for instance, sound rather like a transition between an /ɛ/ and an /e/ (when considering each vowel in isolation). Therefore, there is a correlation between speech rate (modeled through vowel/syllable duration) and articulatory amplitude, both of them being introduced as potential factors of IF0 in section 1. The precise assessment of the influence of these interwoven factors goes beyond the scope of this paper. However, a multiple linear regression (with syllable duration and vowel central F1 as independent variables) was performed for 3 speakers within each speech rate category. It showed no systematic influence of either factor in the whole /aV/ and /dV/ contexts, except that each time a significant influence of syllable duration (it occurred in four cases) was found, it was positive.

4. DISCUSSION

The two major explanatory hypotheses for IF0, namely the tongue-pull hypothesis and the "deliberate planning" hypothesis, are both contradicted by some well-established facts. The results of the present study add to this divergence between facts and theory. The tongue-pull hypothesis cannot explain the reversing effect of the /aV/-context on the IF0 hierarchy, since it considers IF0 as the mere consequence of the tongue height. The "deliberate planning" hypothesis is perception-oriented, but its premises are

largely independent of the surrounding phonemes: the acoustic difference between vowels is claimed to be enhanced by IF0 (this is a theory which we will not discuss here; for a critical review see for instance [13]), but the direction of this enhancement only depends on the vowels themselves. Therefore, the total inversion of the classical IF0 hierarchy in the /aV/-context, which is obtained with two subjects, runs contrary to this principle.

The "reversing effect" of the /aV/-context on IF0 is very robust: it is significant for 7 out of 8 speakers (see table I), and the speaker who does not have this effect has actually no IF0 differences between vowels, except in the /aV/-context where F0(/i/) < F0(/e/) and F0(/i/) < F0(/ɛ/) ($p < 0.01$ in both cases). This robustness is quite interesting, because it allows to state that there is an effect of phonemic context on IF0, which is a new fact in this domain of research. In particular, this calls into question the conception of IF0 as being a purely segmental phenomenon, by showing that the phonemic context influences it in a constant direction.

It also calls for an explanation, which presently does not exist. The recent experimental research on the IF0 of vowels has shown that at least some major refinements are needed for the current explanatory hypotheses to subsist. But even the tongue-pull hypothesis, which appears to be most compatible with the previous data, is totally unable to explain the actual inversion of the IF0 hierarchy which is reliably produced by speakers W1 and W4. Nor is it able to explain the more general reversing effect of the /aV/-context on the IF0 hierarchy.

In this quite puzzling situation, we are led to consider other possible hypotheses to account for the data on IF0. In section 1 we introduced the hypothesis of a trade-off relationship between the motor activities responsible for the articulatory movements, on the one side, and the production of F0 on the other side. At this point the hypothesis cannot be stated in a clearer way, but however this makes it possible to discuss it with respect to the data. Regarding the results of the present study, there may be a reversed connection between the amplitude of articulatory movements and F0, but not a negative correlation between them, since the IF0 hierarchy is far from being reversed for all speakers. However, it is a priori possible that articulatory movements of a higher amplitude be achieved through less articulatory motor activity: in particular, the speed of the movement can be lower, and then involve less activity. In other words, the transition between two phonemes can have different speeds, which hence need not be proportional to the articulatory distance to cover. Overall, then, the "trade-off" hypothesis cannot be properly assessed from the results of the present study.

However, this hypothesis is not incompatible with these results, while both the other discussed hypotheses are. Moreover, it could also provide an explanation for other results, which the latter hypotheses currently cannot account for. To begin with, the lack of IF0 in the lower part of the F0 range could come from the lack of a trade-off there, resulting from a lesser use of the available motor activity when F0 is low. And the fact that "tense" and "lax" vowels have similar IF_0 values, despite their quite different tongue heights, could possibly be related to a smaller motor activity in some part of the tongue for the "lax" variants, which would compensate for a higher activity in other

tongue muscles, due to the higher amplitude of the articulatory movements involved. This could then contribute to justify the somewhat "impressionistic" and controversial terminology used to describe these variants. However, these considerations are presently quite speculative, and we prefer to leave them in the state of a heuristic for further experimental research.

5. CONCLUSION

This study shows that the intrinsic F0 (IF0) hierarchy is reliably and remarkably influenced by the neighboring phonemes. When the latter are set to be the vowel /a/, a reversing effect is observed on the classical hierarchy, to the point that it collapses ; it is even reversed for a few speakers, resulting for instance in their F0(/i/) being lower than their F0(/ɛ/) in the /aV/ context. These results raise the question of the intrinsic nature of the so-called "intrinsic F0" of vowels, and contributes, with other studies, to call into question the explanatory hypotheses which have been proposed for this phenomenon. The design of this experiment comes from a heuristic approach, which consists in the search for manifestations of hypothetical trade-off relationships between simultaneous motor activities. This approach is now to be continued through further experimental studies. In particular, this will involve a parallel study for back vowels.

BIBLIOGRAPHY

[1] WHALEN, D., H., LEVITt, A., G. The universality of intrinsic F_0 of vowels. *Journal of Phonetics* 23, 349-366 (1995).

[2] FISCHER-JœRGENSEN, E. Intrinsic F0 in tense and lax vowels with special reference to German, *Phonetica* 47, 99-140 (1990).

[3] AUTESSERRE, D., ROUBEAU, B., DI CRISTO, A., CHEVRIE-MULLER, C., HIRST, D., J., LACAU, J. & MATON, B. 1987: Contribution du cricothyroïdien et des muscles sous-hyoïdiens aux variations de la fréquence fondamentale en français: approche électromyographique. *Actes du XIe Congrès International de Phonétique*, (Tallinn, URSS).

[4] DYHR, N. The activity of the cricothyroid muscle, and the intrinsic fundamental frequency in Danish vowels. *Phonetica* 47, 141-154 (1990).

[5] VILKMAN, E., AALTONEN, O., LAINE U. & ILKKA, R. (1991) Intrinsic Pitch of Vowels - a complicated Problem with an obvious Solution? *Sixth Vocal Fold Physiology Conference*, Stockholm July 30 to August 3, *Vocal Fold Fysiology, Acoustic, Perceptual, and Physiological Aspects of Voice Mechanisms*, (Gauffin J. & Hammarberg B., eds.), pp. 159-166.

[6] HIROSE, H., GAY, T. The activity of the intrinsic laryngeal muscles in voicing control. An electromyographic study. *Phonetica* 25 (3), 140-64 (1972).

[7] ATKINSON, J., E. Correlation analysis of the physiological factors controlling fundamental voice frequency. *J.A.S.A.* 63(1), 211-22 (1978).

[8] DIEHL, R., L., KLUENDER, K., R. "On the objects of speech perception." *Ecological Psychology* 1(2), 121-144 (1989).

[9] KINGSTON, J. The phonetics and phonology of perceptually motivated articulatory covariation. *Language and Speech* 35, 99-113 (1992).

[10] FAHEY, R., P., DIEHL, R., L. The missing fundamental in vowel height perception. *Perception and Psychophysics* 58 (5), 725-33 (1996)

[11] WHALEN, D. H., ANDREA G. LEVITT, PAI-LING HSIAO, and IRIS SMORODINSKY. Intrinsic F0 of vowels in the babbling of 6-, 9- and 12-month-old French- and English-learning infants. *Journal of the Acoustical Society of America*, 97, 2533-2539 (1995).

[12] C.H. SHADLE: Intrinsic fundamental frequency of vowels in sentence context. *J. Acoust. Soc. Am.* 78, 1562-1567 (1985).

[13] WHALEN, D. H., GICK, B., KUMADA, M., & HONDA, K: Cricothyroid activity in high and low vowels: Exploring the automaticity of intrinsic F0. *Journal of Phonetics* 27, 125-142 (1998).

[14] OHALA, J., J., EUKEL, B., W: Explaining the intrinsic pitch of vowels. In: R. Channon & L. Shockey (eds.), *In honor of Ilse Lehiste.* Ilse Lehiste Pühendusteos. Dordrecht: Foris. 207-215 (1987).

[15] PIOT, O.: *Vers une théorie unifiée de la prosodie du français et de l'anglais: des émotions à la phonologie.* Ph. D. thesis (Université Paris III, Paris 2002).

[16] SCHERER, K., R.: "Vocal affect expression: a review and a model for future research", *Psychol. Bulletin* 99, 141-165 (1986).

[17] BOERSMA, PAUL & WEENINK, DAVID (2005). *Praat: doing phonetics by computer* (Version 4.3.21) [Computer program]. Retrieved September 1, 2005, from http://www.praat.org/

PROTOCOLO DE ANÁLISIS FONÉTICO-ACÚSTICO PARA LA PREVENCIÓN Y EVALUACIÓN DE LA DISFONÍA EN DOCENTES*

SILVIA PLANAS MORALES
Grup de Recerca en Lingüística Aplicada (Universitat Rovira i Virgili)
XAVIER VILLALBA
Grup de Gramàtica Teorica (Universitat Autonoma de Barcelona)

1. INTRODUCCIÓN

Este protocolo tiene dos objetivos bien definidos:

1. Diagnosticar fonética y acústicamente el estado de la voz (*Voice Quality*) de los docentes.
2. Obtener la información necesaria para asesorar en el mantenimiento y la mejora de la voz.

Se ha de tener en cuenta que la consecuencia inmediata de la mala técnica vocal es la disfonía funcional y que ésta puede desembocar en una disfonía orgánica y, por lo tanto, en una patología de la voz.

Un estudio sobre los resultados obtenidos en la aplicación piloto del protocolo de análisis en una muestra de 30 profesores de distintas facultades de la Universitat Rovira i Virgili (Planas-Morales y Villalba, 2003) permitió observar que un 67% del profesorado evaluado presentaba síntomas de abuso vocal y que en el 90% de los casos desconocía o no había practicado nunca la técnica vocal. También se pudo constatar que los parámetros utilizados permitían realizar un asesoramiento personalizado.

2. METODOLOGÍA Y TÉCNICAS DE ANÁLISIS

Los parámetros de análisis son, básicamente, los parámetros fonéticos y acústicos que se utilizan para analizar la voz patológica. Pero el protocolo también se ha adaptado

* Agradecemos la colaboración de todos los docentes que han prestado sus datos a la investigación.

a las características específicas del colectivo docente con el fin de obtener mejores resultados en el asesoramiento.

Para los análisis se utilizan los programas *Macquirer* (*Scicon*) y *SoundScope/16* (*GWInstruments*). Con *Macquirer* se obtienen los espectros promediados con LPC, los espectrogramas y las curvas melódicas (*Pitch*) y de intensidad. Con *SoundScope*, las estadísticas de *jitter* y *shimmer*.

Las muestras de voz parar el análisis se obtienen del siguiente corpus:

- [s] sostenida después de tres inspiraciones profundas para obtener la máxima duración.
- [i], [a] y [u] sostenidas y emitidas en tono modal (TM) o habitual, en tono agudo (TA) y en tono grave (TG). Para evitar el golpe de glotis inicial se articulan en sílaba con [b]. La muestra para el análisis es de unos 2 segundos. Para ser utilizadas en el análisis con *SoundScope*, se convierten los archivos de *Macquirer* en archivos AIFF.
- [ĩ] y [ã] sostenidas y emitidas en TM, formando sílaba con [n].
- [i] y [a] sostenidas en TM, formando sílaba con [b] y después de tres inspiraciones profundas para obtener la máxima duración.
- Series numéricas.
- Frases controladas de distinta longitud, del tipo "El coche de Helena no funciona desde el lunes día uno", "El coche de Helena no funciona desde el lunes día uno, cuando vino a Vinerò".
- Lectura de un texto seleccionado.
- Conversación espontánea.
- Tonos naturales: "uhum", tos, carraspeo, risa. Las muestras que se recogen son los F0 producidos por fonación espontánea o neutra.

En cada subapartado del protocolo se especifica la técnica de análisis, así como las muestras de voz que se utilizan para saber el grado de la calidad de la voz (*Voice Quality*) del docente.

Las grabaciones se realizan en casetes *Sony* y *Marantz* de alta resolución con micrófono externo unidireccional o bien directamente en el programa *Macquirer* o *PCquirer*[1] con el mismo tipo de micrófono.

3. PROTOCOLO DE ANÁLISIS

El protocolo se divide en dos apartados:

a) La calidad de voz o modo de fonación del hablante.

b) La calidad del timbre o modo de articulación de los sonidos del habla; es decir, la impostación de la voz.

1 Versión del *Macquirer* para PC.

3.1. La Calidad de Voz

El diagnóstico de la calidad de voz depende de los resultados de los parámetros de análisis que se refieren a:

a) El control fono-respiratorio: Coeficiente s/z(a), Tiempo Máximo de Fonación (TMF) en vocal, coeficiente del TMF, alteraciones en la curva melódica y observaciones complementarias sobre el control fono-respiratorio (CFR)
b) La vibración de las cuerdas vocales: *jitter* y *shimmer* de vocales sostenidas, componente inarmónico en vocales TM, tipo de voz ronca (*hoarseness*), extensión de la voz y observaciones complementarias sobre la vibración de las cuerdas vocales (VCV).

3.1.1. Coeficiente s/z (a).

Coeficiente entre el tiempo máximo de emisión con fuente sorda y sonora. Los hablantes sin patología vocal deben mantener la misma duración en ambas fuentes, por lo que el coeficiente óptimo es de 1; pero se considera trastorno a partir de 1,4 (Colton, 1996; Gamboa, 2001; Jackson-Menaldi, 1992 y 2002; Planas-Morales y Villalba, 2003).

- Muestras: [s] y [a] TM en máxima duración.
- Técnica: Cronómetro.

3.1.2. Duración fuente sorda.

Se tiene en cuenta el resultado de este parámetro por si el coeficiente s/z(a) es menor de 1,4, pero el máximo de emisión con fuente sorda no llega a los 20s (Jackson-Menaldi, 1992; Planas-Morales y Villalba, 2003), lo que indicaría que la capacidad pulmonar es bastante limitada. Estos resultados se dan en personas con respiración clavicular, por ejemplo.

- Muestra: [s] en máxima duración.
- Técnica: Cronómetro.

3.1.3. Tiempo Máximo de Fonación (TMF)

Consideramos que los valores normales del TMF[2] de una vocal sostenida en tono modal o habitual son de 15s para las mujeres y de 20s para los hombres. Considerando como 'excelentes' los valores superiores a 25s para las mujeres y 35s para los hombres (Colton, 1996; Gamboa, 2001; Hirano, 1981; Jackson-Menaldi, 2002; Morrison, 1996; Planas-Morales y Villalba, 2003).

- Muestras: [i] y [a] TM en máxima duración.
- Técnica: Cronómetro.

---

2 En inglés *Maximum Phonation Time* (TPM).

3.1.4. Coeficiente del TMF

En este apartado aplicamos la siguiente fórmula:

TMF del hablante / TMF mínimo normal > 0,70

Según Jackson-Menaldi (1996) y (2002), la duración mínima normal de TMF en vocal sostenida es de 15s en los Hombres y 14,3s en las mujeres. Si el TMF del hablante, dividido por el TMF mínimo normal correspondiente es mayor de 0,70, el cierre glótico funciona perfectamente. En caso contrario, puede indicar insuficiencia del control glótico, insuficiencia respiratoria, problemas neurológicos o patrones respiratorios erróneos (Jackson-Menaldi, 1996 y 2002; Planas-Morales y Villalba, 2003).

- Muestras: [i] y [a] TM en máxima duración.
- Técnica: Cronómetro.

3.1.5. Alteraciones en la curva melódica

En este parámetro se constatan los aumentos bruscos de F0 y los finales de oración u otros segmentos fónicos sin vibración (Fig.1) (Planas-Morales y Villalba 2003).

- Muestras: Segmentos de la lectura de un texto seleccionado, series numéricas, frases controladas de distinta longitud y segmentos de conversación espontánea.
- Técnica: Gráfico del *pitch* del programa *Macquirer*.

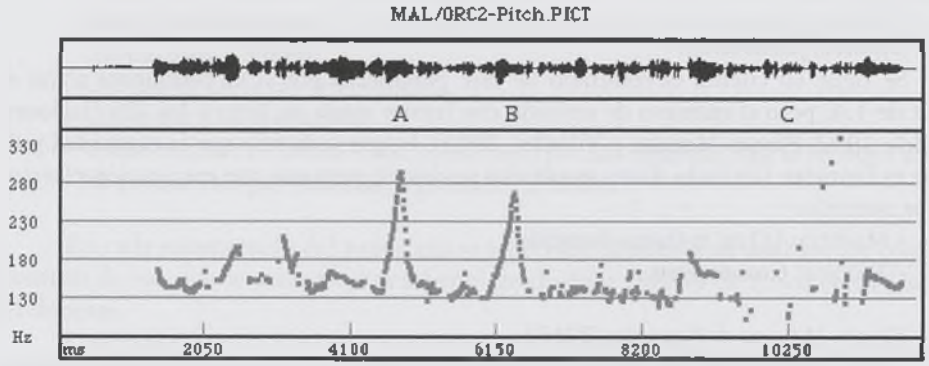


Figura 1. Aumentos bruscos de F0 (A y B) y segmento fónico sin vibración (C) en la curva melódica de una oración en conversación espontánea.

3.1.6. Observaciones complementarias del control fono-respiratorio (CFC).

En este parámetro se constatan si hay alteraciones en los niveles de intensidad (Jackson-Menaldi, 2002), si las oraciones se inician con aspiraciones bucales y cualquier otro dato relacionado con la potencia del flujo del aire que tiene el hablante y el control que ejerce sobre él (Planas-Morales y Villalba, 2003).

- Muestras: Segmentos de la lectura de un texto seleccionado, series numéricas, frases controladas de distinta longitud y segmentos de conversación espontánea.
- Técnica: Espectrograma y gráfico de intensidad del programa *Macquirer.*

3.1.7. Coeficiente de jitter (%) en vocales sostenidas.

El *jitter* es el porcentaje de irregularidad en la frecuencia de vibración de las cuerdas vocales.

Consideramos que los valores normales de *jitter* en tono modal o habitual han de ser menores del 1% (Arauz, 2002; Casado, 2001; Gamboa, 2001; Horii, 1981; Jackson-Menaldi, 2001; Planas-Morales y Villalba, 2003).

- Muestras: [i] y [a] sostenidas en TM.
- Técnica: *SoundStadistic Select* con *SoundScope.*

3.1.8. Coeficiente de shimmer (%) en vocales sostenidas.

El *shimmer* es el porcentaje de irregularidad en la amplitud de vibración de las cuerdas vocales.

Consideramos que los valores normales de *shimmer* en tono modal o habitual han de ser menores del 2,5% (Arauz, 2002; Casado, 2001; Jackson-Menaldi, 2002; Planas-Morales y Villalba, 2003).

- Muestras: [i] y [a] sostenidas en TM.
- Técnica: *SoundStadistic Select* con *SoundScope.* Esta opción del SoundScope permite obtener al mismo tiempo el *jitter* y el *shimmer* de la muestra.

3.1.9. Componente inarmónico.

En este parámetro constamos si en el espectrograma hay presencia de *breathiness* (Laver, 1991); es decir, si aparece señal inarmónica en F2 y F3 inferior al *hoarseness* de tipo I de Yanagihara (1967) (Planas-Morales y Villalba, 2003).

- Muestras: [i], [a] y [u] sostenidas en TM.
- Técnica: Espectrograma de banda estrecha con *Macquirer.*

3.1.10. Tipo de voz ronca o Hoarseness.

Según la clasificación de Yanagihara (1967) en el espectrograma se pueden visualizar cuatro tipos de voz ronca o *hoarseness*, dependiendo del predominio de señal inarmónica sobre la armónica en los formantes de las vocales. La clasificación va desde el tipo I, que es el más leve, al tipo IV, en el cual todos los formantes pasan a tener señal inarmónica.

- Muestras: [i], [a] y [u] sostenidas en TM (Planas-Morales y Villalba 2003).
- Técnica: Espectrograma de banda estrecha con *Macquirer.*

3.1.11. Extensión de la voz.

En este parámetro se registra el tono de voz -grave, modal o agudo-, se compara la extensión de los tonos del habla con la extensión de los tonos naturales – la extensión de los tonos del habla que no son representativos de los tonos naturales suelen ser debi-

dos a un uso muscular inadecuado (Morrison, 1996) – y se constata la tesitura de la voz – entendida como la extensión de F0 que podemos considerar más cómoda a partir del *jitter* y el *shimmer* de las vocales sostenidas realizadas en TG, TM y TA (Planas-Morales y Villalba, 2003).

- Muestras: [i], [a] y [u] sostenidas en TM, TA i TG, los tonos naturales y segmentos de conversación espontánea.
- Técnicas: Espectro con LPC de *Macquirer* y *SoundStadistic Select* con *SoundScope*.

3.1.12. Otras irregularidades en la vibración de las cuerdas vocales (VCV).

En este parámetro se constata la presencia de temblor, subarmónicos (Fig.2) y diplofonía (Fig.3) en la emisión de las vocales, o cualquier otro dato relacionado, como pueden ser oraciones con componente inarmónico o con golpe de glotis inicial (Planas-Morales y Villalba, 2003).

- Muestras: [i], [a] y [u] sostenidas en TM y segmentos de conversación espontánea.
- Técnicas: Espectrograma de banda estrecha, para la observación del temblor y los subarmónicos, y gráfico del *pitch*, para la diplofonía, ambas con *Macquirer*.

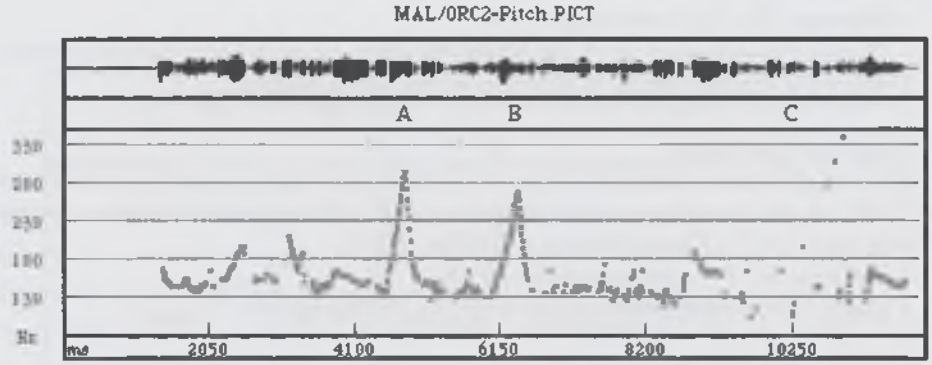


Figura 2. Subarmónicos en el espectrograma de vocal sostenida en TM.

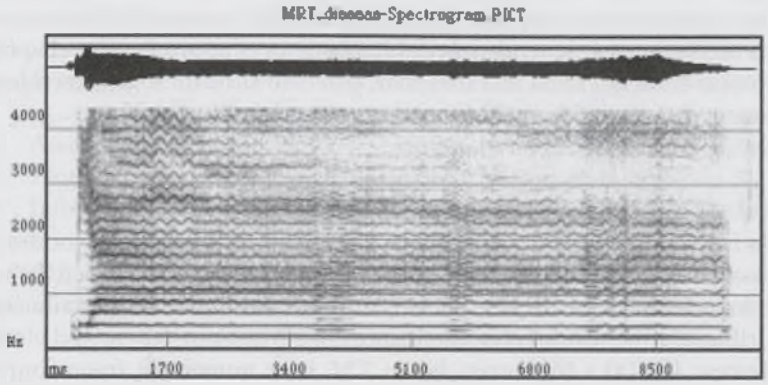


Figura 3. Diplofonía en la curva melódica de vocal sostenida en TM.

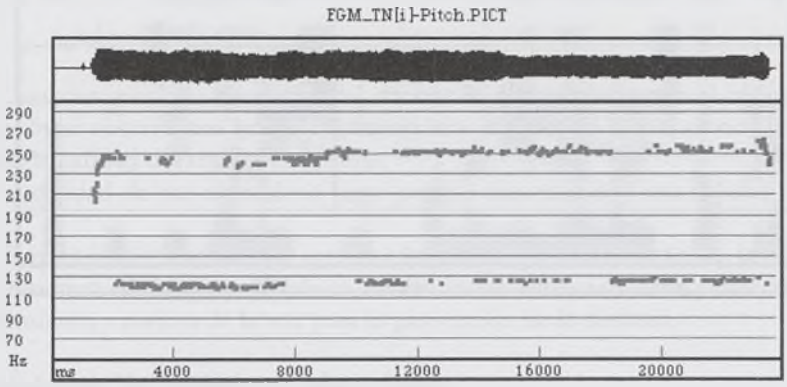


Figura 3. Diplofonía en la curva melódica de vocal sostenida en TM.

3.2. La calidad del Timbre.

En el diagnóstico sobre la calidad del timbre se han escogido parámetros de análisis que den información sobre la impostación y el color de la voz.

3.2.1. Oscilograma estridente en TN {u}.

Por las características articulatorias de la vocal [u], es muy fácil constatar visualmente en un oscilograma si la tendencia del hablante es el apoyo facial o el apoyo laríngeo, en cuyo caso el oscilograma será irregular, parecido al de un sonido estridente.

- Muestra: [u] sostenida en TM.
- Técnica: Oscilograma con *Macquirer*.

3.2.2. Armónicos {-AMP} > 2500Hz.

En las frecuencia superiores a 2500Hz, si el espectrograma de las vocales presenta formantes poco o nada amplificados – [-AMP] – el color de voz que se percibe es tierno, sordo y velado, mientras que si aparecen muy amplificados, la voz es descrita como rica, coloreada, brillante e incluso, a veces, estridente.[3]

- Muestras: [i], [a] y [u] sostenidas en TM, serie numéricas, frases controladas y segmentos de conversación espontánea.
- Técnicas: Espectrogramas de banda estrecha con *Macquirer*.

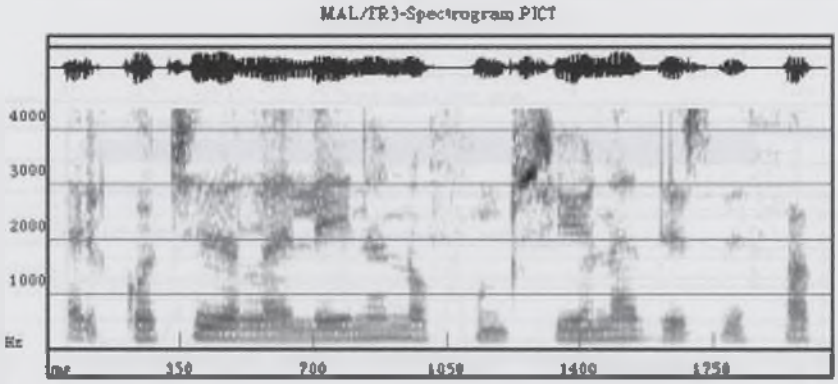


Figura 4. Armónicos [-AMP] > 2500Hz.

3.2.3. Nasalización.

En este apartado comparamos muestras de vocales con muestras de vocales nasalizadas, para descartar cualquier irregularidad del velo del paladar.

- Muestras: [i] y [a] TM, en sílaba con [b], y [ĩ] y [ã] TM, en sílaba con [n].
- Técnicas: Espectrogramas de banda estrecha y espectros con *Macquirer*.

3 También es fácilmente visualizable en el espectro de las vocales.

4. RESULTADOS DEL PROTOCOLO DE ANÁLISIS

El protocolo de análisis que presentamos forma parte del proceso de asesoramiento de voz que se realiza en la Unitat de Fonètica de la Universitat Rovira i Virgili. El proceso completo es el siguiente:

a) Anamnesis.
b) Protocolo de análisis fonético-acústico (descrito en el apartado 3),
c) Informe fonético de la calidad de voz que incluye:
   - Resultados del protocolo de análisis.
   - Asesoramiento.

Atendiendo a los resultados de los análisis del control fono-respiratorio y de la vibración de las cuerdas vocales, diagnosticamos la calidad de voz como:

- Óptima.
- Normal.
- Indicios de Abuso Vocal.
- Abuso Vocal.

Atendiendo a los resultados de los análisis sobre la impostación y el color de la voz, diagnosticamos la calidad del timbre como:

- Óptima.
- Normal.
- Deficiente.
- Nasalizada.

Consideramos que la calidad del timbre es deficiente cuando el apoyo de voz es laríngeo. Si se observa una buena impostación y el color de la voz es rico, coloreado y brillante, la calidad del timbre es óptima.

La calidad de voz normal no tiene porqué coincidir con una calidad normal del timbre.

El resultado de los análisis -gráficos inclusive- y el diagnóstico de la calidad de la voz se presentan en el informe fonético, el cual incluye asesoramiento personalizado en el mantenimiento y mejora de la voz para la prevención de la disfonía.

4.2. Calidad de Voz: Indicios de abuso vocal.

Ejemplo de diagnóstico de calidad de voz con indicios de abuso vocal:

a) Control fono-respiratorio:
   - Coeficiente s/z(a) < 1,4
   - Duración fuente sorda <20s (14s).
   - TMF en vocal <15s (13s).
   - Coeficiente de TMF > 0,70.
   - Alteraciones curva melódica:
     - ✓ Aumentos bruscos de F0.
     - ✓ Finales OR y series sin vibración.
   - Observaciones complementarias CFR: Ninguna.

Aún cuando los coeficientes s/z(a) y TMF sean correctos, la duración de la fuente sorda y el TMF en vocal nos indican que hay poca capacidad pulmonar, como ocurre, por ejemplo, en el tipo de respiración clavicular.

b) Vibración de las cuerdas vocales:
   - jitter < 1%.
   - shimmer <2,5%.
   - Componente inarmónico en vocales TM: Sí.
   - Tipo de voz ronca o *hoarseness*: Ninguna.
   - Extensión de la voz:
     ✓ Vocales TM = vocales TG.
   - Otras irregularidades en VCV: No.

4.3. *Calidad del Timbre: Deficiente.*

Un ejemplo de diagnóstico de calidad del timbre deficiente es el siguiente:

- Oscilograma estridente en [u] TM: Sí.
- Armónicos [-AMP] > 2500Hz: Sí.

La impostación de voz es muy mala, porque el primer parámetro indica que la voz tiene apoyo laríngeo y el segundo, que la resonancia es muy limitada.

5. CONCLUSIONES

El protocolo permite:

a) crear un historial sobre la calidad de la voz de los docentes;
b) diagnosticar el abuso vocal;
c) diseñar medidas preventivas o correctoras personalizadas.

El protocolo contribuye a:

a) mejorar los hábitos de higiene vocal;
b) mejorar las condiciones laborales de los docentes;
c) prevenir las disfonías;
d) diseñar medidas preventivas o correctoras personalizadas.

BIBLIOGRAFÍA

ARAUZ, J.C., ANDONEGUI, L. y ORTEGA, M. *Aplicación del laboratorio computerizado de la voz.* [17/01/02]. Disponible en: http://www.sinfomed.org.ar/Mains/casos/lavoz.htm.

CASADO, J.C., ADRIÁN, J.A., CONDE, M., PIÉDROLA, D., POVEDANO, V., MUÑOZ, E. CANTILLO, E. y JURADO, A. "Estudio objetivo de la voz en población normal y en disfonía por nódulos y pólipos vocales. *Actas de Otorrinolaringología Española, 52, 476-582.* 2001.

COLTON, R.H. y CASPER, J.K. *Understanding Voice Problems: A Physiological Perspective for Diagnosis and Treatment.* Baltimore: Lippincott Williams & Wilkins. 1996 (2ª ed.)

GAMBOA, J., JIMÉNEZ-JIMÉNEZ, F.J., MATE, M.A. y COBETA, I. "Alteraciones de la voz causadas por enfermedades neurológicas". *Revista de Neurología, 33 (2), 154-161*. 2001.

HIRANO, M. *Clinical examination of voice*. Viena: Springer-Verlag. 1981.

HORII, Y. "Vocal shimmer in sustained phonation". *Journal of Speech and Hearing Research, 23, 202-209*. 1980.

JACKSON-MENALDI, M.C. *La voz normal*. Buenos Aires: Editorial Panamericana. 1992.

JACKSON-MENALDI, M.C. *La voz patológica*. Buenos Aires: Editorial Panamericana. 2002.

LAVER, J. *The Gift of Speech*. Edinburgh University Press. 1991.

MORRISON, M. y RAMAGE, L. *The management of voice dissorders*. California: Singular Publishing. 1996.

PLANAS-MORALES, S. y VILLALBA, X. "Assessing in teaching practice: results from the application of acoustic parameters". *Proceeding of the 15th International Congress of Phonetic Sciences*. Barcelona. 2003.

YANAGIHARA, N. "Significance of harmonic changes and noise compnents in hoarseness". *Journal of Speech and Hearing Research, 10, 531-541*. 1967.

LA ENERGÍA Y LA DURACIÓN COMO VARIABLES QUE AFECTAN EN LA PERCEPCIÓN DE LA FRECUENCIA FUNDAMENTAL

José Ignacio Puebla Gutiérrez, David Escudero Mancebo
Departamento de Informática
Alfonso Gordaliza Ramos
Departamento de Estadística e Innovación Operativa
Universidad de Valladolid.

1. INTRODUCCIÓN

1.2 Acerca de la naturaleza del acento en español.

Se han hecho numerosos estudios acerca de la naturaleza del acento. Una de las primeras referencias al acento en el español se la debemos a Navarro Tomás (1948:24): "En español es el acento de intensidad el que mantiene la estructura prosódica de las palabras".

Bolinger y Hodapp (1961) se inclinan a creer que el tono es el elemento más importante. Para Contreras (1963) también el tono es el más poderoso indicio del acento. Podría actuar reforzado o en conflicto con la duración y la intensidad. Quilis (1981:331) cree también que es el tono la referencia más clara del acento. La segunda, sería la duración.

Canellada y Madsen (1987:77) decidieron que la variable más importante es la duración y en segundo lugar, la intensidad. Figueras y Santiago (1993:99) en un estudio de habla sintética consideraron la frecuencia como la variable más clara. En segundo lugar estaría la duración.

Romera (1999:315) considera que el elemento más importante es la duración. Ante tanta disparidad de resultados no extraña que Demers (1994:20) considere que para conocer la existencia o no del acento el único criterio es la opinión mayoritaria de esa evidencia al margen de los valores contrastivos que se pueden obtener instrumentalmente. Esto se consigue a través de test que se presentan a informantes.

Consideramos importante la aportación de Pamies (1997:27). Para este investigador el factor fundamental es la duración pero para los casos en que no esté claro, propone la utilización de unos coeficientes relativos de cada una de las tres variables.

Finalmente, Fabio Tamburini (2005) considera que además de las tres variables citadas, hay que tener en cuenta el "énfasis espectral". Habría que estudiar la energía computada en la banda que va de 500 a 4000 Hz.

1.2 Umbrales de F0, Duración y Energía

No son muchos los estudios que se han realizado sobre umbrales, a pesar de la importancia que parecen tener. Desconocemos la existencia de estudios que traten los umbrales de estas variables de dos en dos o, las tres de forma conjunta.

1.2.1 Umbral de F0

Pamies y otros (2002:272) concluyen que "aunque hay indicios de que el umbral perceptivo sea de 1 semitono, el umbral funcional se sitúa a medio camino entre el mínimo perceptible (1 st) y el máximo imperceptible (2 st), lo cual, interpretado de forma categorial y discreta, confirma que 1,5 st, sería la unidad mínima potencialmente relevante desde el punto de vista prosódico".

Toledo (2003:156) acepta también estos valores y, así, espacia la escala vertical en sus gráficos en umbrales perceptivos de 1,5 st.

1.2.2 Umbral durativo

Pamies y Fernández (2002) abordaron este problema y concluyeron que "el umbral está colocado en un tercio por encima o por debajo, es decir, no se percibirán diferencias de duración que no sobrepasen dicho intervalo".

Es lógico pensar que cuando dos variables se combinan, o lo hacen las tres a la vez, el cerebro humano necesite menos incremento de cada una de ellas para que el efecto aditivo sea suficiente para decidir que una sílaba está acentuada prosódicamente. Un incremento de F0 puede ser tan pequeño que sea en la práctica inaudible. Pero sumado a otros incrementos de Energía y de Duración (también inaudibles por separado) puede producir un conjunto que sí que sea audible.

Estamos muy lejos de cuantificar científicamente estos umbrales cuando actúan de forma conjunta. Por sentido común los umbrales deben disminuir cuando se combinan dos variables y lo harán aún más si se combinan las tres a la vez.

Recientemente hicimos un estudio (Puebla: 2004) sobre umbrales de variables prosódicas tomadas de dos en dos. En aquel momento llegué, entre otras, a estas conclusiones:

- Consideradas las variables de forma aislada, F0 tiene los mayores valores, casi duplicando los de las otras dos, que a su vez, tienen valores semejantes.
- Consideradas las variables de dos en dos, la combinación F0 – I es la que tiene valores mayores. Como en el caso anterior, casi duplica las otras dos combinaciones: I-D, F0-D.
- F0 e Intensidad son las variables que más ligadas están.

3. HIPÓTESIS DE PARTIDA Y OBJETIVOS

Con la presente investigación pretendemos comprobar hasta qué punto es cierta la siguiente hipótesis intuitiva que teníamos. Según ella la percepción de F0 en el habla espontánea para el oído humano vendría condicionada por la energía y por la duración.

Dado que las tres variables prosódicas se combinan siempre en el habla, el oído humano está acostumbrado a oír combinaciones de las tres a la vez. Pensábamos que incluso para un informante con "buen oído musical", las modificaciones en Duración y Energía afectarían a su apreciación de F0.

3.1 Elaboración de material

La secuencia [po] se grabó con voz natural de hombre a 121.16 Hz (la propia voz de uno de los autores). A continuación fue duplicada por copia para formar secuencias silábicas [popo]. Se alteró mediante resíntesis la frecuencia fundamental de las vocales exactamente en un semitono (128.63 Hz).

Como resultado disponemos de 36 secuencias bisílabas que se muestran en la tabla del anexo.

3.2 Test de prueba con personas con un "oído adiestrado"

Antes de pasar el Test a grandes grupos hicimos una prueba con dos niños Ana y Juan que habían cursado al menos hasta 3° de piano. El objetivo era comprobar que éstos eran capaces de detectar el cambio tonal de un semitono[1] en diferentes situaciones de Duración y Energía en la cadena hablada. Si conseguía un éxito cercano al 100 % habríamos demostrado que partíamos de un corpus asequible, en cuanto a lo que se quiere demostrar, para un oído adiestrado.

La pregunta que se les hace es "De la primera a la segunda sílaba, ¿Nota usted subida, bajada, o mantenimiento del tono?". El oyente puede oír dos veces las 36 secuencias que se ordenan al azar. La pregunta principal fue convenientemente explicada. Se dejaba oír la primera secuencia. A continuación se ponía un ejemplo de gran subida o gran bajada entre las secuencias [popo] para asegurarme que los informantes entendían exactamente la pregunta.

En cuanto a los resultados Ana tuvo dos errores y Juan, ninguno.

Ahora bien, lo que puede parecer fácil de discernir en el oído adiestrado, puede ser completamente diferente para el común de la gente. Se trata de un test largo, de 36 secuencias y que puede dar lugar a múltiples errores.

3.3 Pasando el test

Primeramente pasamos el test en una clase de Musicología (Facultad de Filosofía y Letras de la Universidad de Valladolid). Los universitarios reaccionaron positivamente

1 Ya Navarro Tomás (1948) consideraba el semitono como la unidad en los estudios entonativos del español.

a la prueba y posteriormente me hicieron preguntas relativas a los objetivos de la investigación.

Posteriormente pasamos el test en el Instituto Zorrilla y en el Conservatorio en una misma semana de junio de 2005.

El test fue grabado en cintas DAT y y las secuencias fueron oídas gracias a unos altavoces acoplados al DAT.

En total pasé el test a 343 informantes de tres centros:

1. Instituto Zorrilla: 274
2. Conservatorio: 49
3. Musicología: 16
4. Otros: 4

4. RESULTADOS

4.1 Resultados de todos los informantes

Se quiere probar la hipótesis de que en estas casillas las respuestas no se producen al azar (probabilidad 1/3 para cada una de las tres respuestas posibles) sino que las variables energía y duración influyen facilitando la elección correcta.

Planteamos un contraste de hipótesis sobre la proporción de aciertos.

Hipótesis nula: $p = 1/3$ (elección al azar)

Hipótesis alternativa: $p > 1/3$ (las otras variables ayudan a elegir correctamente).

Realizamos un test de proporciones con las respuestas de los 343 informantes en cada una de las seis casillas elegidas. En todos los casos el estadístico de prueba z es muy significativo:

Casilla	Descripción 2ª sílaba	Aciertos	Proporción	p-valor
19	F0 (+1st) D (+100%)	238	0,6939	0,000
31	F0 (+1st) I (+100%)	269	0,7843	0,000
32	F0 (-1st) I (-50%)	246	0,7172	0,000

Véase el Gráfico 1

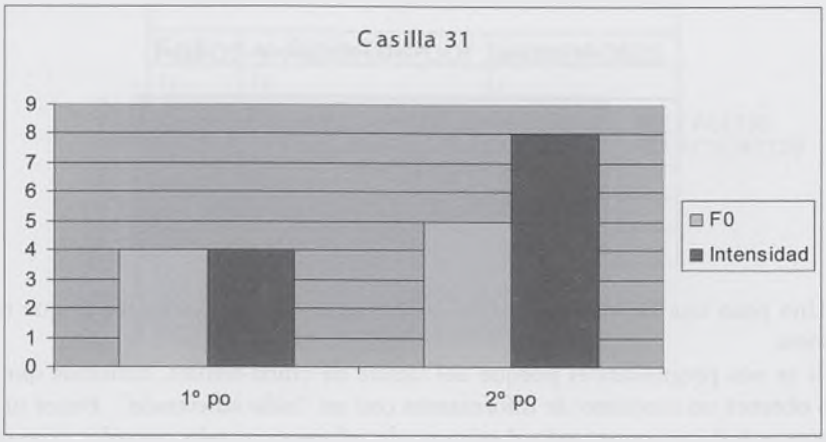


Gráfico 1

Hay otras tres casillas con porcentajes de aciertos especialmente bajos:

Casilla	Descripción 2ª sílaba	Aciertos	Proporción
4	F0 (+1st) I (-50%)	125	0,3644
6	F0 (+1st) D (-50%)	131	0,3819
10	F0 (=) I (-50%)	134	0,3907

En esta última tabla, aunque los porcentajes superan 1/3, si quitáramos los datos pertenecientes a los estudiantes de música (Conservatorio y Musicología), entonces serían inferiores a 1/3.

En las casillas 19, 31y 32, F0 se asocia en el mismo sentido a la Energía (I) o a la Duración (D). Esto ayuda a los informantes a conseguir un alto número de aciertos.

En las casillas 4, 6 y 10, F0 se asocia en sentido contrario a la Energía (I) o a la Duración (D). Esto hace que los informantes tengan un alto número de errores.

4.2 Resultados de los informantes con buen oído musical

Hemos rechazado los test que contienen 6 errores o más. En total han sido rechazados los test de 266 informantes (77,55 %). 36 informantes han tenido 0 errores (10,5%). 41 informantes han tenido 5 errores o menos (11,95 %). En total el número de errores de estos 41 informantes suman 102 que se reparten de esta manera:

Nº de errores	Nº de informantes	Total
1	15	15
2	11	22
3	3	9
4	4	16
5	8	40
		102

Uno puso una "x" en una casilla de "No sabe / no contesta" por lo que tenemos 101 errores.

Si se nos preguntara el porqué del límite de cinco errores, diríamos que hemos querido obtener un conjunto de informantes con un "oído adiestrado". Poner un límite de 10 errores hubiera aumentado el número de informantes seleccionados pero consideramos prioritario la "calidad" de oído. Mirando el límite bajo (cinco errores), no consideramos que sea un abuso teniendo en cuenta la dificultad del test.

Si nos atenemos al límite de cinco errores o menos para calificar un oído de adiestrado, entonces, obtenemos solo 77 informantes (22,45% de los encuestados). Si solo nos fijamos en los encuestados del Instituto Zorrilla, obtenemos el 14,55 %

Observando los test se ve que por una parte, muchos informantes tienen bastantes errores (más de 15, y hasta más de 20); por otra parte está el grupo de selectos con "buen oído" (22,45 %). El nº de informantes que tienen entre 5 y 15 errores, es relativamente bajo. Esto nos lleva a pensar fundamentalmente en dos tipos de informantes: aquellos que tienen un "buen oído" y tienen pocos errores y, aquellos que carecen de esta cualidad y cometen demasiados.

4.2.1 Valoración de los datos obtenidos

Tenemos 36 secuencias y en cada una de ellas, dos posibles errores; en total 72 casillas para los errores.

Dado que tenemos un total de 101 errores, un reparto equitativo daría dos casos, como mucho, por cada una de las 72 casillas. Sin embargo no he tenido una distribución equitativa ya que hay una casilla con 7 errores, otra con 8 y otras dos con 9.

En las secuencias de mantenimiento de F0 es donde más desequilibrio ha habido entre fallos y aciertos:

```
Contingency Tables (TIPO="IGUAL")

Chi-Square Test
-----------------------------------------
 Chi-Square Df P-Value
-----------------------------------------
 47,65 11 0,0000
-----------------------------------------
```

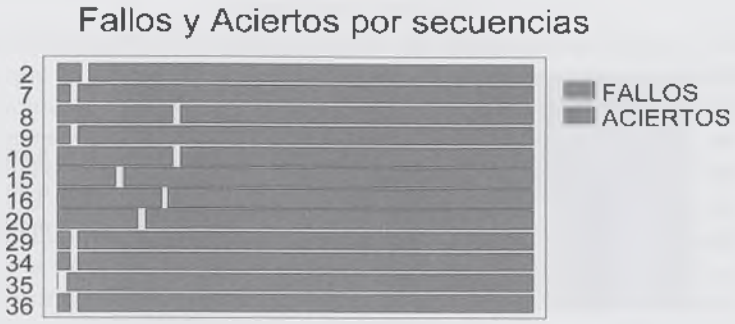


Gráfico 2

Casilla	Descripción	Errores
10	F0 (=) I 2ª sil(-50%)	9
20	F0 (=) I 1ª síl (+100%)	7
8	F0 (=) D2ª síl (+100%)	9
16	F0 (=) D1ª síl (+100%)	8

Las casillas 10 y 20 relacionan Energía y F0. La 1ª sílaba tiene mayor intensidad que la segunda. En ambos casos F0 se mantiene pero 9 y 7 informantes, respectivamente han considerado que F0 bajaba.

Las casillas 8 y 16 relacionan Duración y F0. F0 se mantiene. La Duración de la 1ª sílaba es mayor en la secuencia 16, y menor en la 8. Sin embargo 9 y 8 informantes, respectivamente, afirmaron que el tono bajaba en la 2ª sílaba.

4.3 El factor "edad" en la apreciación de F0

Observando los datos obtenidos en el Conservatorio, y, como se puede ver en el gráfico 3, con el aumento de la edad, aumenta el oído musical. Aquí se nos ocurre la explicación de que con la edad, el oído va consiguiendo mayor precisión, aunque no en todas las personas, en la apreciación de diferencias de F0. Otra explicación quizá sea que algunos alumnos con peor oído abandonen el Conservatorio con el paso del tiempo.

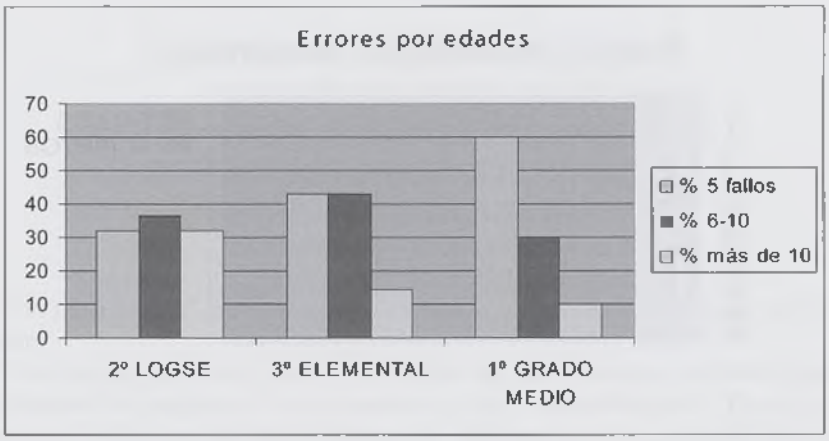


Gráfico 3

4.4 El factor "Centro" en la apreciación de F0

Como era de esperar (Gráfico 4), los informantes de Musicología han obtenido los mejores resultados, seguidos de los del Conservatorio. Los del Instituto Zorrilla, que pueden representar el común de la gente, han obtenido resultados mucho peores.

Gráfico 4

5. CONCLUSIONES

- La percepción de F0 no es independiente de variaciones en Energía y Duración. Dado que en la práctica, en el habla espontánea, siempre las tres variables van modificándose en su transcurso, es necesario seguir haciendo estudios de acentuación, contando con las tres variables.
- Cuando F0 varía en el mismo sentido que la Energía y/o la Duración, la apreciación de la variación de la primera, se facilita.
- Cuando F0 varía en sentido contrario que la Energía y/o la Duración, la apreciación de la variación de la primera, se confunde con facilidad.
- Los informantes aciertan algo mejor la variación de F0 cuando va asociada a la Energía que cuando lo hace con la Duración.

BIBLIOGRAFÍA

BOLINGER, D; HODAPP, M (1961), "Acento Melódico. Acento de intensidad", *BFUCH*, XIII, 33-48.

CANELLADA, M y J MADSEN, (1987) *Pronunciación del español*, Madrid, Castalia.

CONTRERAS, H (1963) "Sobre el acento en español", *BFUCH, XV*, 223-37.

DEMERS, M (1994) "A propos de l'analyse intonative. L'oreille ou la machine", *Langues et Linguistique*, 20, Québec, págs. 1-27.

FIGUERAS C y M SANTIAGO «Investigaciones sobre la naturaleza del acento a través del VISI-PITCH", *Estudios de Fonética Experimental*, V, 81-112.

NAVARRO TOMÁS, T. (1974 [1948]): *Manual de entonación española*, Madrid, Guadarrama.

PAMIES, A (1997) "Consideraciones sobre la marca acústica del acento fonológico" *Estudios de Fonética Experimental*, VIII, 13-49.

PAMIES, A; A. M. FERNÁNDEZ; E. MARTÍNEZ CELDRÁN; A. ORTEGA y M. C. AMORÓS (2002) "Umbrales tonales en español peninsular", *Actas del II Congreso de Fonética Experimental*. Sevilla, Universidad de Sevilla, 272-8.

PAMIES, A; A. M. FERNÁNDEZ (en prensa) "La percepción de la duración vocálica en español", *Actas del V Congreso de Lingüística General*, León, 2002.

PUEBLA, J.I. "Las variables del acento en el español estándar peninsular". *VI Congreso de Lingüística general*, Santiago de Compostela, 2004 (en prensa).

QUILIS, A (1981) *Fonética acústica de la lengua española*, Madrid, Gredos.

ROMERA, L (1999), "La interacción del F0 y la duración en el reconocimiento automático del acento en español", *Actas del I Congreso de Fonética Experimental. Tarragona.*

TAMBURINI, F. "Automatic Prominence Identification and Prosodic Typology" *INTERSPEECH*, Lisboa, 2005, págs 1813-6.

TOLEDO, G (2003): Modelo autosegmental y entonación: los corpus DIES-RTVP", *Estudios de fonética Experimental* XII,142-163.

ANEXO

TEST PERCEPTUAL
PROTOCOLO GRABACIÓN.

	Secuencia Energías						Secuencia Duraciones					
	Sil1			Sil2			Sil1			Sil2		
Id	F0	E	D	F0	E	D	F0	E	D	F0	E	D
1	=	+100%	=	=	=	=	=	=	+100%	=	=	=
2	=	+50%	=	=	=	=	=	=	+50%	=	=	=
3	=	- 50%	=	=	=	=	=	=	- 50%	=	=	=
4	=	=	=	=	+100%	=	=	=	=	=	=	+100%
5	=	=	=	=	+50%	=	=	=	=	=	=	+50%
6	=	=	=	=	- 50%	=	=	=	=	=	=	- 50%
7	=	+100%	=	$+1^{St}$	=	=	=	=	+100%	$+1^{St}$	=	=
8	=	+50%	=	$+1^{St}$	=	=	=	=	+50%	$+1^{St}$	=	=
9	=	- 50%	=	$+1^{St}$	=	=	=	=	- 50%	$+1^{St}$	=	=
10	=	=	=	$+1^{St}$	+100%	=	=	=	=	$+1^{St}$	=	+100%
11	=	=	=	$+1^{St}$	+50%	=	=	=	=	$+1^{St}$	=	+50%
12	=	=	=	$+1^{St}$	- 50%	=	=	=	=	$+1^{St}$	=	- 50%
13	$+1^{St}$	+100%	=	=	=	=	$+1^{St}$	=	+100%	=	=	=
14	$+1^{St}$	+50%	=	=	=	=	$+1^{St}$	=	+50%	=	=	=
15	$+1^{St}$	- 50%	=	=	=	=	$+1^{St}$	=	- 50%	=	=	=
16	$+1^{St}$	=	=	=	+100%	=	$+1^{St}$	=	=	=	=	+100%
17	$+1^{St}$	=	=	=	+50%	=	$+1^{St}$	=	=	=	=	+50%
18	$+1^{St}$	=	=	=	- 50%	=	$+1^{St}$	=	=	=	=	- 50%

Id: Identificador. F0: F recuencia Fundamental. E Energ ía. D Duración. Sil1, Sil2 sílaba de la secuencia. = No se altera. +X% incremento de un X por ciento. –X% decremento de un X por ciento.

TEST PERCEPTUAL

Edad:	Sexo:
¿Tiene conocimientos de Solfeo?:	

A continuación oirá una serie de 36 secuencias bisílabas. Complete la tabla respondiendo a la siguiente pregunta:

De la primera a la segunda sílaba, ¿Nota usted subida, bajada, o mantenimiento del tono?

	SUBIDA	BAJADA	IGUAL	NS/NC
1				
2				
3				
4				
5				

DICIONARIO FONÉTICO GALEGO

Xosé L. Regueira
ILG-USC

1. PRESENTACIÓN

O proxecto que vou presentar e no que veño traballando nos últimos tempos é un *Dicionario de pronunciación da lingua galega.* Este dicionario consta de 46773 entradas, en que a carón da forma ortográfica e da categoría gramatical se ofrece unha transcrición fonética que se propón como modelo de pronunciación para esa forma.

Este proxecto inscríbese dentro dunha liña de traballo que pretende elaborar modelos de pronunciación para o uso público da lingua, en particular nos medios de comunicación e no ensino. Neste sentido xa existían traballos previos sobre o estándar oral galego, así como unha descrición da fonética estándar da lingua (Regueira 1998). E tamén o coñecemento da lingua usada nos ámbitos en que se espera unha variedade estándar (actos públicos, actividades culturais, docencia, medios de comunicación...), así como o da lingua falada tradicional.

As primeiras fases deste traballo foron realizadas en equipo, pero o carácter do traballo, a súa prolongación no tempo e os diferentes intereses dos membros do grupo fixeron que continuase como un traballo individual, mais sempre contando co apoio e a colaboración de dos membros do equipo e de diferentes colegas que xenerosamente me dedicaron unha parte do seu tempo.

Tratábase, pois, de elaborar un dicionario de pronuncia, do estilo do de Daniel Jones (2003[16]) ou o de Upton, Kretzschmar & Konopka (2001) para o inglés, e o do *Duden* para o alemán, coa diferenza de que polo de agora non incluímos nomes propios.

A información necesaria para a toma de decisións sobre a pronuncia foi obtida por dúas vías diferentes. Por unha parte, empreguei a información fonética disponible nas bases de datos do Instituto da Lingua Galega, e pola outra formei unha rede de informantes que responderon a diversas enquisas fonéticas. Naturalmente, tívose en conta a pronunciación do portugués estándar, e para os estranxeirismos a pronunciación da lingua fonte.

As tres liñas de traballo principais que desenvolvemos durante a elaboración do dicionario foron a sociolingüística, a fonética e a fonolóxica.

2. ASPECTOS SOCIOLINGÜÍSTICOS: LINGUA ESTÁNDAR E PLANIFICACIÓN

Dado que este traballo ten unha dimensión sociolingüística obvia, unha das direccións en que se desenvolveu a investigación foi o estudo da variación e dos estándares lingüísticos, tanto no tocante á súa formación e difusión coma no seu funcionamento nas sociedades actuais. Tratábase de facer explícita a natureza do noso traballo, discutir os fundamentes e os criterios, e tamén de ser conscientes das súas repercusións sociais potenciais.

En particular as preguntas que quixemos responder foron:

a) Qué é un estándar lingüístico?
b) É necesario un estándar, e en particular un estándar oral para o galego?
c) Debe planificarse de maneira deliberada e orientada un estándar oral?
d) De ser así, con qué criterios?
e) Qué factores poden determinar a recepción e a eficacia da proposta?

Permítaseme, por razóns de espazo e de oportunidade, omitir aquí a discusión destas cuestións, ás que lles demos a maior importancia, xa que supoñían a xustificación e, en todo caso, determinaban a orientación do noso traballo.

En calquera caso, entendemos que na situación actual do galego a existencia dunha variedade estándar séntese como necesaria para que a lingua poida asumir determinadas funcións. Tocante á lingua oral, ademais, hai que ter en conta que a forte presenza social e o prestixio do español favorece a aparición de modelos urbanos de galego que converxen con el en aspectos fonéticos e prosódicos.

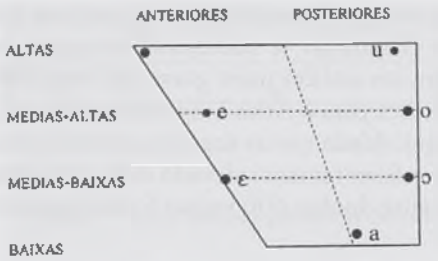
De acordo cos criterios con que se estableceu o estándar escrito e coas directrices que viñan presidindo os traballos sobre o estándar oral, os criterios que seguimos son:

a) continuidade coa lingua falada pola comunidade;
b) actitude purista fronte á influencia do español (manter a identidade fónica do galego fronte á converxencia con esta lingua);
c) atención á extensión xeográfica e demográfica das formas consideradas;
d) preferencia polas formas conservadoras fronte ás innovadoras;
e) harmonía coas demais linguas romances e co portugués en particular, de maneira especial na adaptación de palabras novas.

Este último criterio implica seguir o comportamento xeral dos estándares europeos, entre eles a dependencia respecto da lingua escrita. Por outra parte, os estándares orais destas linguas están baseados nas variedades faladas por certas elites sociais. No caso do galego, este grupo de referencia é o formado polos galegofalantes de formación universitaria e relacionados co mundo da cultura, usuarios do galego escrito e en contacto directo con discursos públicos en galego.

3. ASPECTOS FONÉTICOS: OS SONS DA VARIEDADE ESTÁNDAR

O sistema fonemático do galego e as realizacións consideradas estándar están xa recollidas en obras anteriores e descritas nas gramáticas da lingua. Non obstante, algunhas cuestións foron revisadas e, en ocasións, modificadas. En particular foi revisada e discutida a transcrición das vogais átonas. En xeral ás vogais átonas galegas dedícaselles pouca atención nas descricións fonéticas, e adoitan cos mesmos signos das vogais tónicas:



Mais unha descrición máis atenta e precisa destas vogais ten que sinalar que a diferenza existente entre as dúas vogais de *lobo* ou de *bolo* vai máis alá ca a debida á tonicidade da primeira respecto da segunda. Algúns traballos xa mostraron que as vogais postónicas presentan unha marcada elevación respecto das tónicas, así como unha notable centralización (Regueira, no prelo; V. tm. Molinos Castro 2002):

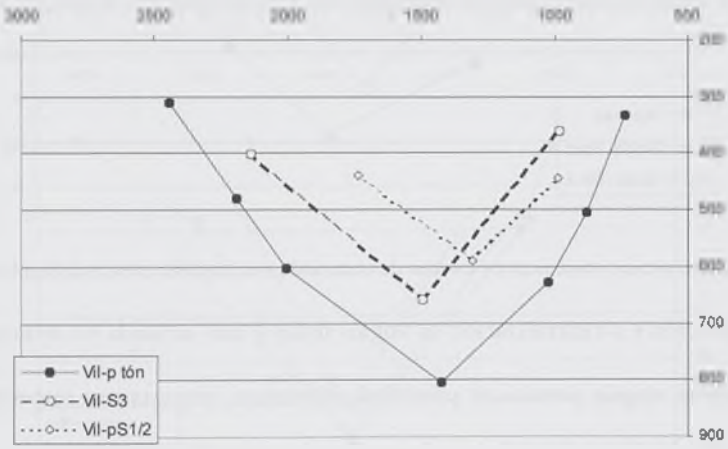


Gráfico 1. Triángulo das vocais finais en posición interior de enunciado (S1/2) e final de enunciado (S3) no galego de Vilalba respecto das vocais tónicas (contexto bilabial)

A elevación e a centralización son menos acusadas en posición final de enunciado, onde a duración é notablemente maior.

De acordo co Alfabeto Fonético Internacional, estas vogais poden representarse como [ɪ], [ɐ], [ʊ]. Neste dicionario, dado que se procura, na medida do posible, unha transcrición accesible e clara para non especialistas, decidiuse representar estas vogais cos signos [e] [a] [o]: [ʁloβo], [ʁbaβa], [ʁseβe]. Estes signos son algo menos precisos, pero teñen a vantaxe de coincidiren coa representación gráfica destas vogais e tamén coas transcricións habituais noutras gramáticas e manuais.

Tamén as vogais pretónicas presentan diferenzas coas tónicas, ás veces moi acusada. Isto pode percibirse claramente se analizamos as vogais pronunciadas nun modelo de lingua tradicional. Para iso utilicei unha gravación realizada en Moeche (A Coruña) en 1975 por Manuel González para o Atlas Lingüístico Galego (gravación depositada no Instituto da Lingua Galega). Aínda que as vogais aparecen en contextos non controlados, se analizamos un número suficientemente elevado delas, podemos obter resultados significativos. Neste caso, a análise dunhas 600 vogais fornece estes resultados:

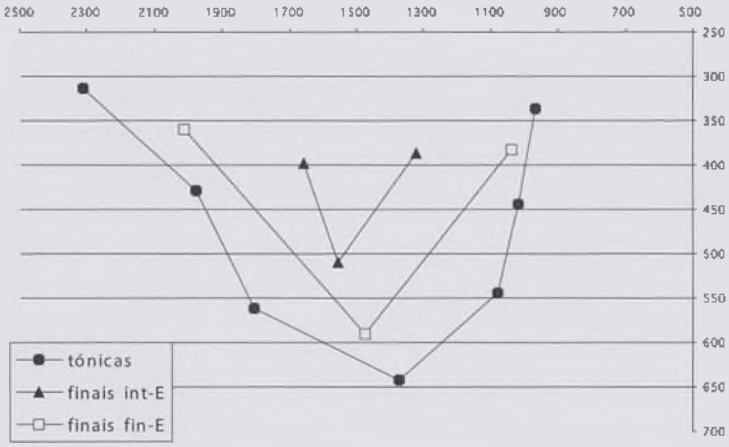


Gráfico 2. Vogais tónicas e postónicas en interior de enunciado e en final de enunciado fonolóxico (Moeche)

A elevación e a centralización as vogais finais é moi acusada en interior de enunciado.

Tamén as vogais pretónicas presentan diferenzas importantes respecto das tónicas:

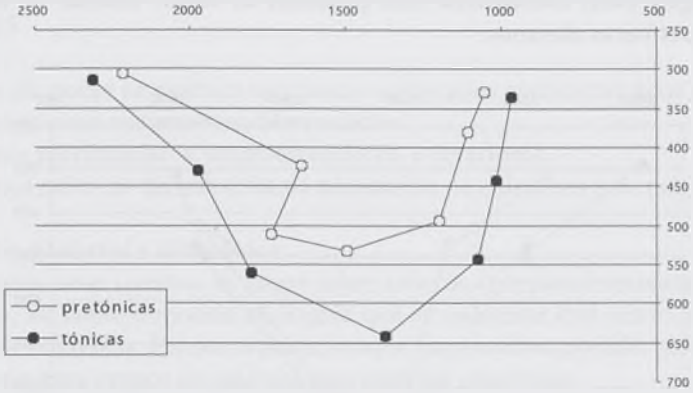


Gráfico 3. Vogais tónicas e pretónicas (Moeche)

Cando menos as vogais medias pechadas e a vogal aberta deberían ter unha transcrición diferenciada da das tónicas, que podería ser [ɛ] [ɐ] [ʊ]. Nos falantes "cultos" con fonética galega encontramos situacións semellantes, con algúns matices.

Se comparamos este caso con outros modelos converxentes co español, que se sitúan no outro extremo do rango de variación fónica do galego, podemos observar diferenzas notables. Tomemos como exemplo unha gravación en que unha escritora galega le un fragmento dunha novela súa (gravación dispoñible en Álvarez Cáccamo et al. 2001: CD 2, pistas 25 e 26). Despois de analizar varios centos de vogais, obtense como resultado para as vogais finais:

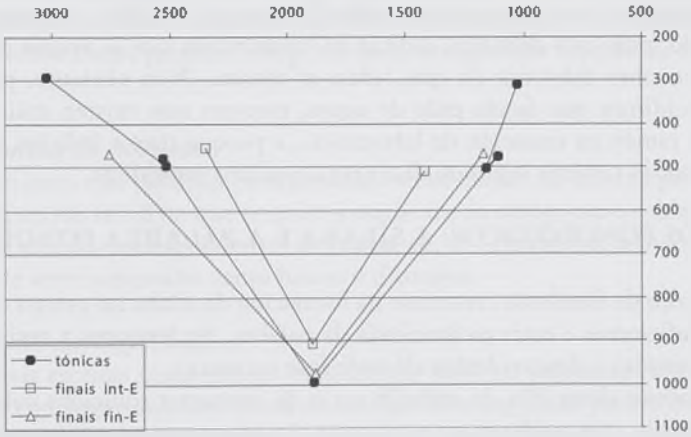


Gráfico 4. Vogais tónicas e finais (en interior e en final de enunciado fonolóxico) na lectura dun fragmento narrativo pola súa autora

As vogais finais encóntranse moi próximas ás vogais tónicas. Se analizamos as tónicas e as pretónicas obtemos:

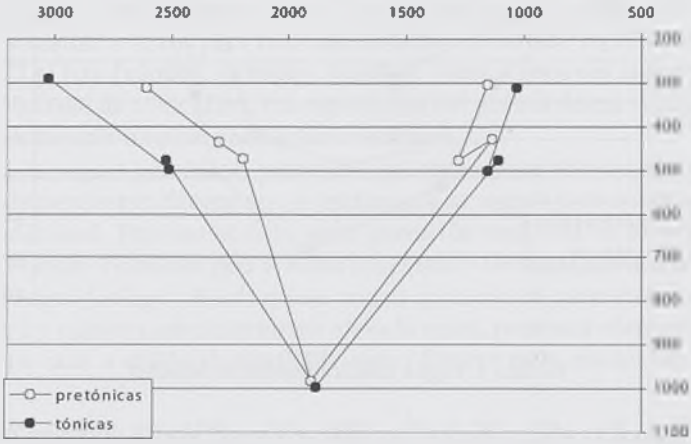


Gráfico 5. Vogais tónicas e pretónicas na lectura dun fragmento narrativo pola súa autora

Como se ve, só se distinguen cinco vogais tónicas e átonas, e estas teñen valores moi próximos aos das tónicas. Isto coincide co que dicía Navarro Tomás (1985[22]: 46) para o vocalismo átono do español ("La distancia entre la vocal relajada y el tipo tenso y normal que cada vocal corresponde no pasa nunca de límites relativamente reducidos"), así coma cos resultados obtidos por Martínez Celdrán (1984: 293-294, 299).

Naturalmente, os criterios manexados lévannos a describir o primeiro modelo e non o segundo, polo que debemos indicar na transcrición que as vogais átonas se pronuncian cun timbre diferente do que teñen as tónicas. Non obstante, para as vogais pretónicas decidimos non facelo polo de agora, mentres non existen máis estudos que o corroboren, tamén en situación de laboratorio, e porque temos indicios de que poden aparecer diferentes timbres segundo distintas situacións prosódicas.

4. ASPECTOS FONOLÓXICOS: A SÍLABA E A PALABRA FONOLÓXICA

O traballo de fonoloxía centrouse na formación da sílaba no galego estándar e nas variedades tradicionais e mais na fonoloxía da palabra. Iso levounos a revisar os estudos sobre estas cuestións e desenvolvelos alí onde fose necesario.

O propósito desta liña de traballo era o de encontrar solucións coherentes e, na medida do posible, que puidesen ser explicadas con argumentos lingüísticos, a diferentes problemas da fonoloxía da palabra, como a elisión e a coalescencia de vogais en locucións e palabras compostas ou a adaptación de cultismos e palabras novas. A falta de espazo non nos permite desenvolver agora este aspecto.

5. A APLICACIÓN DOS CRITERIOS SELECCIONADOS: PROBLEMAS E PROPOSTAS

Á hora de poñer en práctica os criterios adoptados apareceron dous tipos de problemas que requiriron tratamentos diferenciados:

a) Léxico patrimonial: a variación dialectal e xeracional.

b) Léxico culto: as diverxencias na adaptación de cultismos e de palabras novas.

5.1. Variación dialectal e diacrónica

Para estes casos tivemos en conta sobre todo os criterios de extensión xeográfica e demográfica, así como o modelo de lingua que actualmente está sendo utilizado polo grupo social de referencia. Por outra parte, sempre que resultou posible, procurouse unha solución común para grupos de palabras coas mesmas secuencias.

Sen entrarmos en detalles da variación fonética na xeografía galega, procurouse recoller pronuncias moi estendidas, deixando fóra variantes menos representadas. Por esa razón, en casos en que existen dúas pronuncias moi extensas aparecen dobres solucións, en igualdade ([ʁoʎo] [ʁøʎo]) ou establecendo unha forma preferente e outra minoritaria ([moɹmento̝] > [moɹmento̝]). Non obstante, procuramos que o número de formas dobres fose reducido.

O criterio de extensión xeográfica non se aplica cando unha determinada pronuncia non se admite por razóns sociolingüísticas. O caso máis evidente é o da *gheada*, un cambio fonético demográfica e xeograficamente maioritario, pero que non é aceptado no galego estándar e contra o que existen fortes prexuízos. O feito de que non se recollan as pronuncias con gheada non quere dicir, por tanto, que non as consideremos válidas; só se está constatando que, na actualidade, a gheada non é admitida nos usos formais da lingua estándar.

Existen tamén discrepancias entre os datos máis conservadores (gravacións e transcricións de persoas maiores, realizadas hai trinta ou máis anos) e os usos das xeracións máis novas. Nestes casos, procurei dar prioridade, sempre que foi posible á forma tradicional.

5.2. A adaptación de cultismos

Os cultismos e as palabras incorporadas recentemente ao idioma presentan certas pautas de adaptación tanto no que se refire á estrutura da sílaba como sobre todo ao timbre das vogais medias, correspondentes ás grafías <e, o>, e tamén ás secuencias de vogais susceptibles de seren adaptadas como hiatos o ditongos.

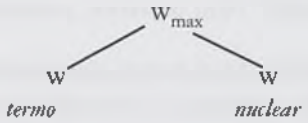
5.2.1. Adaptación das vogais medias

As vogais escritas como <e, o> en posición tónica o pretónica en palabras novas son adaptadas con vogais medias-baixas ou medias-altas seguindo pautas complexas, que só en parte somos capaces de enxergar.

Algúns comportamentos descritibles, e mencionados xa en traballos anteriores, inclúen a presenza de vogais medias-baixas en sílaba tónica de palabras esdrúxulas (*bólido,*

bélico), en sílabas con oclusivas en coda (*directo, copto*), en posición inicial de palabra (*elástico, obeso*), nos ditongos [jɛ] [wɛ] (*quiniela, secuela*). Algunhas terminacións e sufixos parecen seguir comportamentos fixados segundo patróns xa existentes no léxico patrimonial, e así as palabras en *–eta* teñen a mesma vogal do sufixo *–eta* [ʁeta]_ (*asceta, raqueta*), as palabras en *–ón* teñen todas vogal [o], igual que as acabadas en *–ono* (*emoción, carbono*), e as acabadas en *–eno* levan todas [e] (*benceno*). En realidade non temos unha explicación para a maioría destes comportamentos, amplamente compartidos entre galego e portugués (e estamos a falar da adaptación de cultismos, non de léxico patrimonial).

Nalgunhas ocasións, por contra, cremos poder apuntar algún dos factores que están incidindo no comportamento destas vogais. Por unha parte, comprobase o abaixamento das vogais en posición tónica en esdrúxulas: *ecónomo* [ɛɹkɔ̝nomo] - *economía* [ɛkonoɹmia] – *económico* [ɛkoɹnɔ̝miko]. Por outra banda, algúns "prefixos" cultos (o raíces) de tipo *rino-, termo-, poli-, retro-* etc., que constitúen un dominio prosódico diferenciado (enténdase como unha palabra fonolóxica ou como un pé métrico que se integra na palabra fonolóxica seguinte), presentan un comportamento semellante ao das palabras compostas, tanto na vogal tónica como na postónica:



[ʁtɛˈmo]_> [ɪtɛˈmo̞nukleɹa] (o portugués, en cambio, ten un tratamento diferenciado para as vogais en límite de palabra fonolóxica en compostos coma este: [ʁtɛˈmu] > [ɪtɛˈmɔ̝nuɹkljaˈ]). Por contra, estes segmentos teñen un comportamento diferente en voces con estrutura prosódica diferente: [teˈɹmɔ̝metˈo̞].

De acordo con este comportamento (moi estendido), esperariamos que *biblioteca, discoteca, filmoteca,* etc., presentasen unha vogal [o̞] ao final da raíz inicial [ɪbiβljo̞ɹtɛka], etc., e así aparece nunha parte dos falantes, pero outros informantes, por contra, pronuncian [biβljɔ̝ɹtɛka]. Non obstante, todos teñen [o] en *bibliografía, bibliofilia*, etc.

Estas adaptacións fan que debamos establecer hipóteses sobre a estrutura prosódicas destas voces e de cales son os patróns que rexen a asignación de timbres vocálicos en situación pretónica.

No caso das adaptacións discrepantes entre uns falantes e outros, pode supoñerse que se deben a diferenzas na priorización dos criterios de adaptación, coma no caso de *tolueno,* que para uns falantes ten vogal tónica media-baixa e para outros media-alta. Algúns falantes seguen o criterio xeral de adaptar os ditongos escritos <ie> <ue> con vogal media-baixa [jɛ] [wɛ], e polo tanto [to̞ɹlwɛno], pero para a maioría é a vogal xeral da terminación *–eno* (cfr. *rabeno, cacheno*) a que prevalece: [to̞ɹlweno].

En moitos casos obsérvase que as diferentes adaptacións aparecen en grupos de idade diferentes, de maneira que parecen apuntar a cambios en marcha na fonoloxía da lingua.

5.2.2. Ditongos e hiatos

A adaptación de secuencias vogal alta + vogal, susceptibles de seren adaptadas como ditongos ou como hiatos, permite observar un destes cambios. O comportamento máis común do galego neste tipo de secuencias é o mantemento de hiato: *piano, tiara, triángulo, hiato; maniota, carriola; dual, puntual, suave* (pero *diante, piastra,* con ditongo, así como outros casos de ditongos [jɛ], [wɛ]: *ciencia, sueco*).

Non obstante, a miúdo se producen dúas adaptacións alternativas. Un exemplo pode ser o prefixo *dia-*. A grande maioría dos falantes pronuncia [diaɹɫɛktɔ], con hiato. A partir desta base, uns derivan [djalɛktolɔɹʃia], mentres que outros presentan [dialɛktolɔɹʃia]_ O mesmo sucede con [diɹalɔuɔ] > [djalɔɹua'] ~ [dialɔɹua']. Esta diferenza parece remitir a unha diferente ordenación das constricións prosódicas, e obsérvase que o mantemento de *dia-* como bisilábico en posición pretónica afastada da sílaba tónica aparece sobre todo nos grupos de menor idade.

6. PRESENTACIÓN FORMAL

O resultado deste proxecto é un dicionario que se abre cunha introdución en que se explica o carácter da proposta, fanse explícitos os criterios seguidos e descríbese o modelo fónico estándar. Inclúense, naturalmente, táboas aclaratorias dos signos fonéticos utilizados e das abreviaturas.

A pronuncia proposta preséntase en transcrición fonética, de acordo coas convencións da Asociación Fonética Internacional, con algunhas simplificacións que buscan facilitar a lectura das transcricións para lectores non expertos en fonética, pero mantendo sempre a información fonética relevante para a pronunciación da palabra.

As entradas organízanse do xeito seguinte:

entrada *categoría gramatical* [transcrición(s) fonética(s)] | *xénero complementario* [transcrición, se diverxe da da entrada], *sinónimo*, NOTA

En fases futuras prevese a posibilidade de ampliar as entradas con información sobre a distribución xeográfica dalgunhas pronuncias. Outra mellora consistirá na inclusión nomes propios.

5. CONCLUSIÓN

Este é un traballo de carácter moi distinto ás típicas investigacións sobre fonética, sexan traballos de laboratorio ou de campo. Aquí interrelaciónanse sociolingüística, fonética e fonoloxía de tal maneira que non podería asegurar cal delas é máis importante. Traballar nesta encrucillada de disciplinas permitiume aprender sobre planificación, estándar e variación, e a atender a aspectos da fonoloxía galega aínda pouco tratados, como a prosodia da palabra fonolóxica, entre outras. Quedan abertas varias cuestións relevantes da fonoloxía do galego que coinciden con algunhas liñas de

traballo en desenvolvemento sobre linguas como o portugués, o catalán, ou o español (entre outras).

Nestes momentos, este dicionario está sendo sometido a unha última revisión pola miña parte. A continuación serán consultados os membros do Instituto da Lingua Galega para que acheguen suxestións, melloras, críticas e opinións. Logo será sometido a outros colegas doutros centros e universidades, antes da súa publicación. Estes filtros sen dúbida permitirán mellorar e enriquecer o resultado.

Teño expectativas favorables sobre a súa recepción. De feito xa foi utilizado para a transcrición das vogais tónicas no *Novo dicionario da lingua galega*, publicado recentemente por Obradoiro-Santillana (Xove Ferreiro 2005), e as transcricións están sendo incorporadas á nova edición do *Gran dicionario Xerais da lingua galega*.

En todo caso, ben sabemos que a existencia dun modelo de pronunciación non vai supoñer unha inflexión na situación do galego falado. A evolución de esta situación está determinada por factores sociais e políticos, non lingüísticos. Por tanto, espero que este dicionario sexa un instrumento útil para certos ámbitos (ensino, dobraxe, medios de comunicación en xeral), pero os seus efectos sobre a consolidación ou non de certos modelos de galego serán limitados. É dicir, a variación fónica do galego seguirá determinada pola forte (e crecente) presenza do español na sociedade galega, polas valoracións sociais desiguais do galego e o español, pola diferenciación social entre os grupos urbanos de clase media que falan galego respecto dos galegofalantes de clases baixas o de procedencia rural, e polo descoñecemento e o baixo prestixio do portugués. Modificar esa situación non está ao alcance dun dicionario. Mais un gran tamén fai morea.

REFERENCIAS BIBLIOGRÁFICAS

Álvarez Cáccamo, Xosé María, et al. (2001): *Poetas e narradores nas súas voces.* Vol. 1. Santiago de Compostela: Consello da Cultura Galega.

Duden Aussprachewörterbuch. Wörterbuch der deutschen Standardaussprache. Bearbeitet von Max Mangold in Zusammenarbeit mit der Dudenredaktion. Mannheim: Dudenverlag, 1990[3].

Jones, Daniel (2003[16]): *English pronouncing dictionary.* Ed. by Peter Roach, James Hartman & Jane Setter. Cambridge: Cambridge University Press [1ª ed. 1917].

Martínez Celdrán, Eugenio (1984): *Fonética.* Barcelona: Teide.

Molinos Castro, Rita (2002): "As vocais átonas finais en galego: estudio acústico", *Cadernos de Lingua* 24, 55-91.

Navarro Tomás, T. (1985[22]): *Manual de pronunciación española.* Madrid: CSIC.

Regueira, Xosé L. (no prelo): "Vocais finais en galego e en portugués: un estudio acústico", in *Actas do VII Congreso da Asociación Internacional de Estudios Galegos.* Barcelona (en publicación).

Regueira, Xosé Luís (coord.) (1998): *Os sons da lingua.* Vigo: Xerais.

Upton, Clive; William A. Kretzschmar Jr. & Rafal Konopka (2001): *The Oxford dictionary of pronunciation for current English*. Oxford: Oxford University Press.

Xove Ferreiro, Xosé (dir.) (2005): *Novo dicionario da lingua galega*. Santiago de Compostela: Obradoiro – Santillana.

FLUENCY ARTICULATION AND SPEECH RATE AS NEW PARAMETERS IN THE SPEAKER RECOGNITION

Luciano Romito
Rosita Lio
Vincenzo Galatà

INTRODUCTION

Over the years, many and various solutions have been proposed to solve the problem of the identification of an unknown speaker.

The human ability to recognize, in daily life, the voice of a known person[1], has led to the belief that the identification by means of voice can have a possible application in judicial enquiries.

On account of its variable nature, the voice would not appear to be suitable for such a complex field. However, scientific knowledge of phonetic parameters depending on the speaker (such as the fundamental frequency F_0, the vowel formants' frequencies[2], the resonance of nasal consonants FN1, FN2, FN3, the Articulation Rate and the Speech Rate[3]), allows the use of the voice for speaker recognition in forensics.

The ascription of a voice to a speaker requires analysis, observation, evaluation and comparison of parameters.

Since this study focuses on the search for new parameters to be used in Speaker Recognition, the typical procedures of the semi-automatic method have been adopted:

- choice of acoustic parameters to be analysed;
- evaluation of the statistical relevance of the considered parameters;

1 The identification procedures are however different in forensic speaker recognition because the voices are always unknown.

2 These parameters are approved and used in judicial enquiries.

3 These parameters are for the moment only used in laboratory for experimentation.

- estimation and comparison of samples by means of a statistical test (using the results of previous measurements).

The goal of this study is to verify the reliability of the fluency indexes of Articulation Rate and Speech Rate, through the verification of the identity of a speaker in a closed-set of known voices.

The audio-verbal communication is realized through the production and reception of phonic blocks of varying size (sentence, word groups, single words or fragments of a word). These blocks are internally modeled by a certain pitch, by speed variation and by stress position. The features of the fluency indexes are prosodic phenomena or, more frequently, over-segmental ones[4]. The over-segmental features of phones differ from the segmental ones because of their intrinsic nature in relation to the surrounding phones: a phone is considered in terms of more or less duration, intensity, quality considered not in absolute terms but always in reference to the context in which it finds itself.

The Articulation Rate depends on the intrinsic duration of the various phones, on the rate of articulation movements and on the rules of co-articulation.

The Speech Rate depends on speaker-specific features, on the communicative situation (diatopic and diastratic variables), but it varies also in a sentence with accelerations and contractions indicating, among other things, the degree of attention that the speaker asks the listener (speed for less attention, slowness for more attention). Furthermore, the rhythm of speech in a similar situation depends on the phonetics alphabet used[5].

The study that follows is divided into three steps:

1. Study of reliability of the fluency index of Articulation Rate;
2. Study of reliability of the fluency index of Speech Rate;
3. re-evaluation of the perceptual method using a group of *naïf* listeners.

ARTICULATION RATE AND SPEECH RATE: SOME DEFINITIONS

According to the definition of the indexes, it is important to establish what parameters it is necessary to consider and how to calculate them.

By means of a brief review of some relevant definitions in linguistics literature, it is possible to observe how, over the years, there have been various attempts to define the fluency indexes of Articulation Rate and Speech Rate.

For example, for the Articulation Rate, there is:

4 This means above the segments because they refer to the entire sequence.

5 Studies on three different varieties (Bari, Napoli and Pisa) demonstrate that the production of syllables in less time is not a voluntary strategy, but it depends on the different phonetics alphabets used (Pettorino 2003).

- "*La vitesse d'articulation. Véritable vitesse de phonation, quisqu'on lui retrace le temps de pauses. elle est exprimée en nombre de syllabe/sec et s'obtient en divisant le nombre de syllabes émises par le temps d'articulation du locuteur*" (Grosjaen and Deschamps 1975: 148) "Articulation Rate. Real speech rate, since we count the time of pauses, it is expressed in number of syllables per second and can be calculated by dividing the number of syllables uttered by the speaker's articulation time";

- "The articulation rate of each group ('utterance') was computed by dividing the total number of syllables in the group by the cumulative duration of the runs comprising the group (excluding any pause time)", where run is "the stretch of speech that contains no pauses, with a pause defined as a silent interval of 250 ms or greater" (the calculus of Articulation Rate is made on a group of runs with a minimum of 30 syllables, (Miller *et al.* 1984: 218-219);

- "*il rapporto tra il numero delle sillabe e la durata della catena fonica*[6]" (Sorianello 1996:95) "The ratio between the number of syllables and the duration of the speech chain";

- "*il numero delle sillabe diviso per la durata della catena fonica*[7]" (Magno Caldognetto and Vagges 1993: 101, recalling Duez 1982) "The number of syllables divided by the duration of the speech chain";

- "*espresso come il numero di sillabe al secondo della sequenza articolata*[8]*, risultante dalla formula: numero delle sillabe della sequenza articolata/durata della sequenza articolata*" (Zmarich *et al.* 1996:120) "Expressed as the number of syllables per second of the speech chain resulting from the formula: number of syllables of the uttered sequence/duration of the uttered sequence";

- "average number of (phonetic) syllables per second of the articulation phase of speech[9] (number of syllable/[duration-combined duration of all pauses])" (Künzel 1997: 1358);

- "*la velocità di articolazione è data dal rapporto tra il numero delle sillabe e la durata delle catene foniche. Nel computo delle sillabe vengono di*

6 Here the writer uses the definition of speech chain by Pettorino and Giannini 1994: "*La catena fonica è la porzione di un enunciato compresa tra due pause vuote*".

7 The speech chain is for the authors: "*Total articulation time: durata globale della produzione verbale del soggetto, costituita da catene foniche e dai silenzi*".

8 That is the speech chain without the non silent pauses.

9 In this case it is the duration of each speaker's task, equivalent to half a minute.

norma inclusi tutti quei fenomeni udibili quali pause piene e prolungamenti vocalici." (Giannini 2000: 253) "The articulation rate is given by dividing the number of syllables by the duration of the speech chains. Any audible phenomenon, such as full pauses and lengthened vowels, are normally included in the number of syllables";

- "*equivale al rapporto tra il numero delle sillabe realmente pronunciate e il tempo impiegato per realizzarle*" (Pettorino 2003: 228) "equals the ratio between the number of syllables really pronounced and the time used to realize them".

The first difficulties that we run into after this brief but important review of definitions are those related to the different ways to consider what parts of the signal to examine: "*temps d'articulation du locuteur*" (articulation time of the speaker), run, speech chain, uttered sequence, the articulation phase of speech, the time used to realize the syllable and the different definitions of uttered sequence and speech chain. Moreover, the different definitions of silent pauses[10] and the various limits given when silence occurs followed by stop consonants [11].

Some definitions of Speech Rate are:

- "*il rapporto tra il numero delle sillabe e la durata dell'enunciato*" (Sorianello 1996: 95) "The ratio between the number of syllables and the duration of the enunciation";

- "*Speech Rate: espresso come numero di sillabe al secondo della sequenza articolata, in relazione alla durata della catena fonica comprensiva di esitazioni e disfluenze, risultante dal rapporto tra il numero delle sillabe della sequenza articolata e la durata dell'intera catena fonica*" (Zmarich *et al.* 1996: 120) "Speech rate: expressed as the number of syllables per second of the uttered sequence in relation to the duration of the phonic chain including hesitations and disfluencies resulting from the ratio between the number of syllables in the uttered sequence and the duration of the entire phonic chain";

- "*rapporto tra il numero di sillabe e il tempo totale dell'enunciato*" (Pettorino 2003: 228) "Ratio between the number of syllables and the total time of enunciation".

10 Miller *et al* 1984: "a silent interval of 250 ms or greater"; Sorianello 1996 "*una momentanea sospensione dell'attività fonatoria e conseguente assenza di rumore spettrografico, di almeno 100 ms nel parlato spontaneo*"; Künzel 1997 gives the limit of 100 ms.

11 Magno Caldognetto and Vagges 1993 give 100 ms to the phase of consonantic occlusion, whilst Zmarich *et al* 1997, give 50 ms.

In general, the result is a certain degree of confusion.

In this study, we will refer to Zmarich *et al.* (1996: 120) to calculate the Articulation Rate while to calculate the Speech Rate we used the definition of these writers considering the formula: number of syllables of the uttered sequence/duration of the phonic chain, expressed in syll/sec.

MATERIALS AND METHODS

The materials used in this study are recordings of read speech taken from four female speakers having the same diastratic and diatopic features (same peer-group, same culture degree etc). The four speakers, indicated as D1, D2, D3, D4, have also parental relationships: D1 (age 25) is sister of D4 (age 24); D2 and D3 are twins (age 23) and are aunts of D1 and D4. The recordings consist of a list of ten sentences and a list of ten repetitions of a same sentence. The sentences have been constructed trying to avoid the beginning of the words by stop consonants: in this way it was possible to detect the presence of possible silent pauses the speaker used within a sentence. The segmentation[12] of the sentences was hand made by means of the visualization of three windows: waveform, spectrogram, energy curve and with the help of auditory feed-back. The syllable computation was carried out by identifying the phonetic syllable[13] according to the rules of syllabification[14].

We considered only uttered sequences and phonetic chains with a number of syllables superior or equal to eight, to remove the possibility of misleading data. We did not establish limits to the duration of pauses. The material as described was used for the calculation of articulation rate and speech rate.

Considering the parental situation of our speakers, we decided to apply the perceptual method to verify if the voices would be recognized in the right way although their similarity and although the different recording channels. For this task we made two types of recording for each one of the speakers: one in a silent room and the other one through the telephone line. To the four speakers used for the preceding task we added here another female speaker, D5 (age 45) mother of D1 and D4 and sister of D2 and D3 to complicate the listeners' recognition task. The speakers were asked to read two passages from a newspaper of no less than 120-130 seconds. The reading of the two passages was repeated more times to achieve more spontaneity in the reading. For the experiment we extracted only 6 sentences for each speaker (3 ones for the silent room

12 A phonetic transcription with informations on the temporal setting of unit boundaries (see Salza 1990: 24-25).

13 The notion of syllable has been object of study for a long time because of its phonetic and phonological "double nature"; in this study we refer respectively to: "*la struttura elementare che sta alla base d'ogni raggruppamento di fonemi*" (see Jakobson 1974: 94) and "*l'unita prosodica costituita da uno o piu foni, agglomerati intorno ad un picco d'intensità*"(see Leoni and Maturi 1998: 74-5).

14 Differences between phonetic and phonological syllabification are not so great, only in the case of group /s/ and /z/ + consonant (for example "*festa*" is phonetically divided into [ˈfɛs-ta] and phonologically in [ˈfɛ-sta]), and some connections, not autochthonous, such as /tm/ and /tl/ (for example '*atmosfera*, *atleta*'), see Muljačić 1969: 471.

recording and 3 ones for the telephone line) of minimum 2,7 seconds of length[15]. For the test we used 60 *naïf* listeners (university students of both sexes) who were asked to answer "yes" if the voice compared belonged to the same speaker and "no" in the case it did not. In doubtful cases the listeners could answer "perhaps" thus expressing that it could belong to the same speaker. To verify the validity of the listeners' answers we used 26 control tests excluding those listeners who did not answer them correctly.

DATA ANALYSIS

General Consideration

The first thing to keep in mind is that we are dealing here read speech.

In the case of repetitions of a same sentence it is possible to see how each speaker plans approximately the same duration for the production and how it differs among the four speakers.

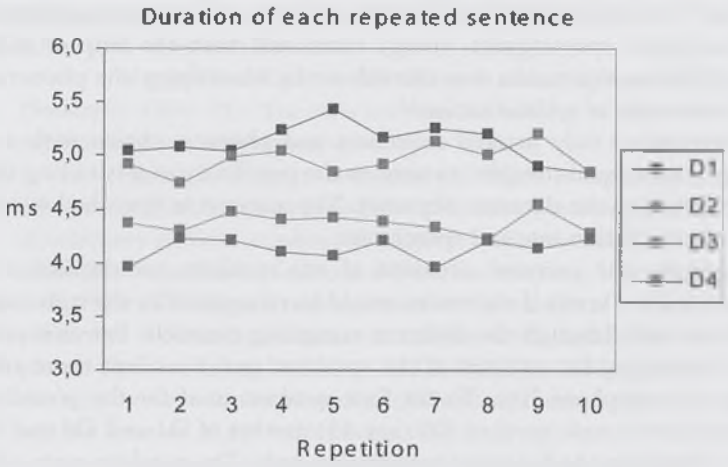


Figure 1 Duration of each repeated sentence considering all kind of pauses.

Articulation Rate

We remember that for the Articulation Rate we used here the formula of Zmarich *et al.* (1996: 120) resulting from: number of syllables of the uttered sequence/duration of the uttered sequence.

15 According to Romito L., 2000, pp.199-203.

The indexes of Articulation Rate calculated for the four speakers are presented in Table 1 and Table 2 for the two sentences' lists.

	10 repetitions of the same sentence																																	μ	σ	n
speaker	1			2			3			4			5			6			7			8			9			10								
D1	6,2	-		6,0	-		6,0	-	-	6,2	6,0	-	6,3	5,7	-	6,4	5,8	-	6,6	5,5	-	6,7	6,0	-	6,6	6,1	7,1	6,3	-				6,2	0,4	17
D2	6,9	6,1		6,9	6,6		7,1	6,4	6,3	6,8	5,9	-	7,2	6,4	6,0	7,3	6,0	-	6,8	3,6	5,7	6,8	6,4	5,9	6,8	6,1	7,1	6,4	6,2				6,4	0,7	25
D3	6,9	-		7,0	-		6,7	-	-	6,8	-	-	6,8	-	-	6,9	-	-	6,9	-	-	7,1	-	-	6,5	7,0	7,0	-	-				6,9	0,2	11
D4	7,6	-		7,5	-		7,8	-	-	7,6	7,9	-	7,9	7,9	-	7,9	-	-	7,9	7,7	-	7,4	-	-	7,5	7,7	7,5	-	-				7,7	0,2	14
μ: Mean Articulation Rate; σ: Articulation Rate's standard deviation; n: number of values																																				

Table 1 Articulation Rate index calculated for each uttered sequence of the 10 repetitions of the same sentence.

	single sentences																																	μ	σ	n
speaker	1			2			3			4			5			6			7			8			9			10								
D1	5,1	-		5,1	-		5,0	5,0	-	5,7	4,9	-	5,3	5,3	-	5,3	5,0	-	5,3	6,0	5,1	5,3	4,8	-	6,0	-	5,5	5,7	-				5,3	0,4	18
D2	5,4	-		7,0	-		6,7	6,4	-	7,3	5,8	-	7,1	5,7	6,1	6,6	5,5	6,8	6,8	7,4	6,2	6,6	5,8	-	6,9	-	7,2	7,1	-				6,5	0,6	20
D3	6,0	-		6,6	-		6,9	-	-	7,4	6,1	-	6,6	-	-	6,1	-	-	7,0	6,0	-	6,2	-	-	6,8	-	6,8	-	-				6,5	0,4	12
D4	5,6	-		6,2	-		6,2	-	-	6,4	6,9	6,3	6,6	-	-	6,7	6,0	6,8	7,2	5,6	-	6,8	-	-	8,1	6,7	7,8	-	-				6,6	0,7	16
μ: Mean Articulation Rate; σ: Articulation Rate's standard deviation; n: number of values																																				

Table 2 Articulation Rate index calculated for each uttered sequence of the 10 single sentences.

At a first look we can see that within the same speaker there is already great difference if we consider the mean articulation rate and the standard deviation for the two sentences' lists.

10 repetitions of the same sentence					single sentences			
D1	D2	D3	D4		D1	D2	D3	D4
6,2	6,4	6,9	7,7	μ	5,3	6,5	6,5	6,6
0,4	0,7	0,2	0,2	σ	0,4	0,6	0,4	0,7
17	25	11	14	n	18	20	12	16
μ: Mean Articulation Rate; σ: Articulation Rate's standard deviation; n: number of values								

We used here the T-test to calculate the dissimilarity percentages, which provide the results in Table 3 for each comparison.

Intra-speaker repetitions vs. single	p	% dissimilarity
D1 - D1	0,0000000598	99,9999940152
D2 - D2	0,4748176757	52,5182324259
D3 - D3	0,0311056454	96,8894354560
D4 - D4	0,0000150707	99,9984929281

Table 3 T-test results for the Articulation Rate.

The difference between repetitions of a same sentence and single sentences within the same speaker is probably due to the increasing confidence with the repeated sentence and according to us it does not represent a problem in this case considering it quite normal.

Inter-speaker repetitions	p	% dissimilarity
D1 - D2	0,86573179349	67,1623751786
D1 - D3	0,00000342724	99,9996572758
D1 - D4	0,00000000001	99,9999999998
D2 - D3	0,00006998279	99,6105945181
D2 - D4	0,00000000009	99,9999996662
D3 - D4	0,00000000073	99,9999999922

Table 4 T-test results for the Articulation Rate.

According to what we said before for the repetitions the test seems to work well for all the speakers giving results higher than 99% except for the comparison between D1 and D2 with 67,2%.

Inter-speaker single sentences	p	% dissimilarity
D1 - D2	0,00000002956	99,9999982412
D1 - D3	0,00000009297	99,9999873364
D1 - D4	0,00000113624	99,9999240993
D2 - D3	0,78793137618	5,0165460178
D2 - D4	0,56213540878	32,0611884799
D3 - D4	0,71477394077	28,7261784387

Table 5 T-test results for the Articulation Rate.

On the other hand, considering single sentences, all the speakers seem to plan differently the single productions with a larger range of variability. This situation has its repercussions on the statistical results where we find only D1 well differentiated from all the other speakers.

Speech Rate

If the goal of this study is to find out a temporal parameter that characterizes the speaker and that describes his production task using the syllable number and the temporal aspect, it is necessary to consider all of the phonic production, including the syllables of non-silent pauses.

This choice is also suggested by results in Duez (1982)[16]: for this author, the phenomenon of hesitation is a speaker-specific feature, above all in the tendency to use or not to use full pauses.

To calculate the Speech Rate we used here the formula: number of syllables of the uttered sequence/duration of the phonic chain, expressed in syll/sec.

The indexes of Speech Rate calculated for the four speakers are presented in Table 3 and Table 4 for the two sentences' lists.

	10 repetitions of the same sentence																											μ	σ	n
speaker	1		2		3			4			5			6			7			8			9		10					
D1	6,2	.	6,0	.	6,0	-	.	6,2	6,0	-	6,3	5,7	-	6.4	5,8	-	6,6	5,5	.	6,7	6,0	.	6.6	6,1	7,1	6,3		6,2	0.4	18
D2	6,6	6,0	6,5	6,4	6,8	6,1	6,3	6,5	5,9	.	7,0	6,4	6.0	7,0	6,0	-	6,5	3,6	5,7	6,5	6,4	5,9	6,5	6,1	6,6	6,4	6,2	6,2	0,6	25
D3	6,9	-	7,0	.	6,7	-	-	6,8	-	.	6,8	-	-	6,9	.	-	6,9	-	-	7,1	.	-	6.5	7,0	7.0	-	-	6.9	0.2	11
D4	7,6	-	7,5	.	7,6	.	.	7,4	7,9	.	7,7	7,8	.	7,5	-	-	7,7	7,7	.	7,4	-	-	7,5	7,7	7,3	-		7,6	0,2	14
μ: Mean Speech Rate; σ: Speech Rate's standard deviation; n: number of values																														

Table 6 Speech Rate index calculated for each uttered sequence of the 10 repetitions of the same sentence.

	single sentences																											μ	σ	n
speaker	1		2		3			4			5			6			7			8			9		10					
D1	5,1	.	5,1	-	5,0	5,0	-	5,7	4,9	.	5,3	5,3	-	5,3	5,0	-	5,3	6,0	5,1	5,3	4,8	-	6,0	.	5,5	5,7	.	5,3	0,4	18
D2	5,3	-	7,0	-	6,7	6,4	.	7,3	5,8	.	7,1	5,7	6,1	6,6	5,5	6,8	6,8	7,0	5,9	6,6	5,8	.	6.9	-	6,9	7,1	-	6,5	0,6	20
D3	6,0	.	6,6	-	6,9		.	7,3	6,1	-	6,6	-	.	6,1		-	7,0	6,0		6,2	.	-	6,8	-	6,8	-	-	6,5	0,4	12
D4	5,6	-	6,2	-	6,2	-	.	6,1	6,9	[illegible]	6,6	.	-	6,7	6,0	6,8	7,2	5,6	-	6,8	-	.	8,1	6,7	7,8	.	.	6,6	0.7	16
μ: Mean Speech Rate; σ: Speech Rate's standard deviation; n: number of values																														

Table 7 Speech Rate index calculated for each uttered sequence of the of the 10 single sentences.

10 repetitions of the same sentence					single sentences			
D1	D2	D3	D4		D1	D2	D3	D4
6,2	6,2	6,9	7,6	μ	5,3	6,5	6,5	6,6
0,4	0,6	0,2	0,2	σ	0,4	0,6	0,4	0,7
17	25	11	14	n	18	20	12	16
μ: Mean Speech Rate; σ: Speech Rate's standard deviation; n: number of values								

Also in this case we have the same results as those presented for the Articulation Rate. This is due to the fact that we are here considering read speech.

---

16 Duez D. (1982): in a study on three styles of speech in French.

Inter-speaker repetitions	p	% dissimilarity
D1 - D2	0,86573179	13,4268206514
D1 - D3	0,00000343	99,9996572758
D1 - D4	0,00000000	99,9999999988
D2 - D3	0,00006998	99,9930017206
D2 - D4	0,00000000	99,9999999910
D3 - D4	0,00000000	99,9999999273

Table 8 T-test results for the Speech Rate.

Inter-speaker repetitions	p	% dissimilarity
D1 - D2	0,86573179	13,4268206514
D1 - D3	0,00000343	99,9996572758
D1 - D4	0,00000000	99,9999999988
D2 - D3	0,00006998	99,9930017206
D2 - D4	0,00000000	99,9999999910
D3 - D4	0,00000000	99,9999999273

Inter-speaker single sentences	p	% dissimilarity
D1 - D2	0,00000003	99,9999970441
D1 - D3	0,00000009	99,9999907030
D1 - D4	0,00000114	99,9998863761
D2 - D3	0,78793138	21,2068623821
D2 - D4	0,56213541	43,7864591221
D3 - D4	0,71477394	28,5226059233

Table 9 T-test results for the Speech Rate.

Perceptual method

According to the control test we considered only 40 listeners as reliable. From the results presented in Table 10 and Table 11 it is possible to see how the greatest part of the identification percentages are over the 90 %. The tables do not need further comments.

silent room recordings								
Comparison			Yes	No	Perhaps	Yes	No	Perhaps
1	D4	D5	9	105	6	7.5%	87.5%	5%
2	D4	D1	16	101	3	13.33%	84.16%	5%
3	D4	D2	4	115	1	3.33%	95.83%	0.83%
4	D4	D3	9	111	0	7.5%	92.5%	0%
5	D5	D1	2	115	3	1.66%	95.83%	2.5%
6	D5	D2	5	112	3	4.16%	93.33%	2.5%
7	D5	D3	7	105	8	5.83%	87.5%	6.66%
8	D1	D2	12	104	4	10%	86.66%	3.33%
9	D1	D3	11	103	6	9.16%	85.83%	5%
10	D2	D3	11	104	5	9.16%	86.66%	4.16%

Table 10 Results for the perceptual test.

telephone line recordings								
Comparison			Yes	No	Perhaps	Yes	No	Perhaps
1	D4	D5	7	112	1	5.83%	93.33%	0.83%
2	D4	D1	16	97	7	13.33%	80.83%	5.83%
3	D4	D2	6	112	2	5%	93.33%	1.66%
4	D4	D3	9	106	5	7.5%	88.33%	4.16%
5	D5	D1	8	109	3	6.66%	90.83%	2.5%
6	D5	D2	2	117	1	1.66%	97.5%	0.83%
7	D5	D3	9	104	7	7.5%	86.66%	5.83%
8	D1	D2	14	103	3	11.66%	85.83%	2.5%
9	D1	D3	17	91	12	14.16%	75.83%	10%
10	D2	D3	13	105	2	10.83%	87.5%	1.66%

Table 11 Results for the perceptual test.

CONCLUSIONS

On the basis of a careful analysis of the used methods and of the obtained results, we can conclude as follows:

- We have statistically significant results for Articulation Rate and Speech Rate through the objective method when the temporal planning of the speaker is characteristic. On the other hand, considering speakers with a similar fluency rate we reach low recognition percentages;
- For Speech Rate we indicate here the mean production of syllables for each phonetic chain, where for phonetic chain we mean the sequence of phonetic

segments (including the non-silent pauses) between two silent pauses, and/or portions of signal included between one inspiration and another.

- To analyse the importance of the fluency index of Speech Rate, as a parameter in Speaker Recognition, it is necessary to consider the whole phonic realization of a speaker and making statistical comparisons with samples of voice of the same, and/or of different speakers, to support the used method considering a common segmentation protocol. The need to consider the speaker's entire phonic realization means in fact identifying the phonetic syllables that build the uttered sequence (including the non-silent pauses).
- Considering the perceptual method as we considered it, the test seems to give good results even with similar sounding voices (even if we have percentages that never reach the 99%).

Our purpose for the future is to calculate the indexes of Articulation Rate and Speech Rate using the same speakers (for their parental situation) analysing read passages to verify if it is possible to obtain more homogeneous data. On the other hand spontaneous speech will be taken into account because of the characteristic fragmentation of on-line planning[17] considering those features strictly bound to the realisation of uttered sequences and/or phonic chains of differing length caused by the syllables of the so-called non-silent pauses[18], the lengthening of vowels[19], the reduction and the diphthongization[20] and the fall of syllables or parts of them.

Once obtained and evaluated the results we will compare the voices of the above said speakers using the parametric approach used in Speaker Recognition (weighting in statistical terms the Articulation and Speech Rate indexes as information to be used in SR).

Moreover it is very important for this experiment to create a corpus containing spontaneous speech (with pragmatic variables) to be used as training and test set.

Last but not least, we will try to put down some guidelines for the tasks of segmentation, calculus and analysis of the various indexes defining some rules and finding homogeneous and non-ambiguous definitions to be used to achieve the same results on the same material in different laboratories.

---

17 Spontaneous speech is characterized by a tendency towards the reduction of articulatory effort, causing phenomena of assimilations, centralizations, de-accentuations, elisions and, from the acoustic point of view, medium shortening of all the phones, fall of phones and/or syllable tendency to omit segments of the sequence, see Kohler, 1995, op. cit. in Zmarich *et al.* 1997.

18 "*Pause piene: esitazioni, interiezioni, allungamenti di vocale, disfluenze, ecc.*": Zmarich *et al.* 1997.

19 The spectro-acoustical analysis of some vowels' "length" at the end of a word or in hesitations, suggests the production of two or more phonetic syllables.

20 Because of the combination and/or meeting of vowels in adjacent words.

BIBLIOGRAPHY

DUEZ D. (1982), Silent and non-silent pauses in three speech styles, *Language and Speech*, 3: 179-192.

GROSJAEN F., DESCHAMPS A. (1975), Analyse contrastée des variables temporelles des l'anglais et du français: vitesse de parole et variables composantes, phénomènes d'hésitation, in *Phonetica*, 31: 144-184.

JAKOBSON R. (1974), *Saggi di linguistica generale*, a cura di L.Helmann, Milano, Feltrinelli.

KÜNZEL H. J. (1997), Some general phonetic and forensic aspect of speaking tempo, in *Forensic Linguistics*, 4(1): 48-83.

LEONI F.A., MATURI P. (1998), *Manuale di fonetica*, Roma, Carocci.

MAGNO CALDOGNETTO E., VAGGES K. (1993), Le pause quali indici diagnostici per lo stile del parlato spontaneo, in *Atti delle 2ᵉ Giornate di Studio del G.F.S.*, Calabria, 28-29 novembre 1991, 19: 97-106.

MILLER J.L., GROSJAEN F., LOMANTO C. (1984), Articulation Rate and its variability in spontaneous speech: a reanalysis and some implication, in *Phonetica*, 41: 215-225.

MORI L., PAOLONI A., (2004), Sulla sociolinguistica forense:la costituzione di corpora vocali per l'analisi della velocità di articolazione in italiano, in *Atti delle XIV Giornate del GFS, Viterbo, 4-6 Dicembre 2003.*

MULJACIC Z. (1969), *Fonologia generale e fonologia della lingua italiana*, Bologna, il Mulino.

PETTORINO M. (2003), Caratteristiche prosodiche dell'italiano dialogico in *Voce Canto Parlato, studi in onore di F.Ferrero*, a cura di P. Cosi, E. Magno Caldognetto, A. Zamboni, Unipress, Padova, pp. 227-230.

ROMITO L. (2000), *Manuale di fonetica articolatoria, acustica e forense*, Centro Editoriale e Librario Università della Calabria, Cosenza.

ROMITO L. BLEFARI M. (2003), *Verso un nuovo parametro nel riconoscimento del parlatore* unpublished thesis, University of Calabria.

SALZA P. L. (1990), La problematica della segmentazione del segnale vocale, in *Atti delle 1ª Giornata di Studio del G.F.S*, Padova, 3-6-novembre.

SORIANELLO P. (1996), Dal parlato letto al parlato spontaneo: indici prosodici a confronto, in *Atti delle 7ᵉ Giornate di Studio del G.F.S.* Napoli, 14-15 novembre, 7: 89-110.

ZMARICH C., MAGNO CALDOGNETTO E., FERRERO F. (1997), Analisi confrontativa di parlato spontaneo e letto: fenomeni macroprosodici e indici di fluenza, in *Quaderni del CNR*, 16: 266-290.

LA PALATALIZACIÓN DE LA CONSONANTE NASAL: COMPORTAMIENTO ANTE DIPTONGO

ASSUMPCIÓ ROST BAGUDANCH
Universitat de Girona

1. INTRODUCCIÓN

En el estudio de la lengua oral desde un punto de vista sincrónico se ha observado que ciertos procesos reseñados en obras de tipo historicista parecen reproducirse en habla no cuidada en nuestros días. Un caso paradigmático susceptible de ser investigado desde este punto de vista sería el de la yod en sus distintos tipos. Así, siguiendo estos criterios, se ha creído interesante analizar el contexto de nasal ante diptongo (equiparable a uno de los supuestos de yod segunda) ya que, perceptivamente, en habla espontánea, se puede advertir cierta tendencia a la palatalización. De hecho, existen investigaciones, si bien de carácter dialectal, que parecerían confirmar dicha tendencia al presentar ejemplos en los que se ha llegado a una solución palatal partiendo del contexto antes mencionado.

En general, la bibliografía coincide en señalar que, en español, la nasal asimila el punto de articulación de la consonante siguiente pero no se ve influida, en principio, por semiconsonantes o vocales

Las consonantes nasales, acústicamente, se caracterizan por presentar estructura formántica y una intensidad menor que la de otros sonidos con esta misma estructura (especialmente las vocales), amén de la aparición de antirresonancias; es decir, de zonas de muy baja intensidad en el espectro. Además, pueden distinguirse uno o más formantes típicamente nasales y muy débiles; generalmente se detecta uno hacia los 900-1000 Hz que denominaremos FN. Los dos primeros formantes (F1 y FN) tienden a ser comunes a todas las consonantes de este tipo mientras que serían los dos siguientes (F2 y F3) los que las distinguirían. Cabe señalar que el primero y el cuarto dependen de la cavidad oral (ya que, por una parte, sus frecuencias no varían de modo importante y, por otra, se corresponden con el F1 y el F3 vocálicos), el segundo lo hace de modo exclusivo de la nasal (sus valores no varían en función del contexto) y el tercero de la oral, visto que su frecuencia se modifica considerablemente dependiendo del contexto (equivale al F2 de las vocales).

Siguiendo a varios autores[1], parece ser que las distintas nasales del español se podrían diferenciar a través de una serie de parámetros, entre ellos, la duración del segmento (mayor en la palatal), las transiciones o, como ya se había adelantado, la altura de F2 y F3 Así, tal como describe Fant (1970), una realización palatalizada se caracterizaría por tener el F2 hacia los 2000 Hz (la lengua en posición de [i]), aproximándose a los valores de la nasal palatal.

El presente trabajo pretende ser una aproximación al comportamiento de la nasal ante semiconsonante palatal. Su objetivo es doble: por una parte, tratar de comprobar si, en el contexto [nasal+diptongo], la consonante solamente se articula como alveolar y, por otra, analizar su manifestación acústica, partiendo de la hipótesis de que en este contexto sí se da variación y de que una de las posibles realizaciones correspondería a una nasal palatalizada

2. DISEÑO EXPERIMENTAL

2.1. Corpus e informantes

Para este estudio se ha partido de un corpus de lectura de párrafos en el que se ha equilibrado el número de casos de las distintas variantes del contexto [nasal+diptongo] (a saber, [ṇie], [ṇia] y [ṇio]), con lo cual se ha tenido en cuenta, además de la nasal seguida de semiconsonante palatal, la vocal siguiente en función del punto de articulación (palatal, central y velar). Por otra parte, puesto que se trata de comprobar la existencia de variación con respecto a la realización alveolar, se ha creído necesario incorporar al corpus un número de casos equivalente de la misma para poder establecer una comparación entre los dos contextos: [nasal+diptongo] y [nasal+vocal]. Se han recogido veinte casos de cada variable recurriendo a dos informantes varones castellanohablantes.

Asimismo, se ha tomado en consideración la tonicidad de la sílaba y, a pesar de que en un principio se pensó tener en cuenta la posición (interior o final), en vista de que algunos análisis preliminares parecen apuntar a su poca relevancia, se ha decidido desestimar este parámetro. De este modo, la distribución de los casos es como se aprecia en la tabla 1:

		Informante 1	Informante 2
[ṇie]	sílaba tónica	20	20
	sílaba átona	20	20
[ṇia]	sílaba tónica	20	20
	sílaba átona	20	20
[ṇio]	sílaba tónica	20	20
	sílaba átona	20	20
[ni]	sílaba tónica	20	20
	sílaba átona	20	20
TOTAL CASOS		160	160

Tabla 1: Número y distribución de los casos analizados.

[1] Albalá (1992) o Machuca (91).

2.2. Procedimiento de recogida de datos y método de análisis.

Como ya se ha indicado, se ha recurrido a un corpus de lectura de párrafos, advirtiendo a los dos informantes que intentaran leer del modo más natural posible, sin cuidar la elocución. Se ha desechado la posibilidad de recurrir a un corpus de habla espontánea debido a los problemas para obtener un número de casos suficiente para su tratamiento estadístico.

Las grabaciones se han realizado en una cámara anecoica utilizando un minidisc Sony MZ-R700PC y un micrófono Shure Unydne 5155D. Una vez realizadas, se han digitalizado para proceder a su análisis acústico con el programa Praat 4.3.11.

Para las mediciones acústicas, se ha tenido en cuenta la duración de la nasal, la frecuencia de sus formantes (puesto que se consideran buenos indicadores de su punto de articulación, especialmente F2) y la duración de las transiciones de la consonante.

Se ha efectuado el tratamiento estadístico de los datos con el programa SPSS 12.0, estableciéndose como variables la duración de cada uno de los segmentos del contexto, la de las transiciones, la frecuencia de cada uno de los formantes de la nasal, el acento y el punto de articulación de la vocal (anterior, central o posterior), así como la estructura silábica: consonante seguida de vocal (CV) o consonante seguida de semiconsonante y vocal (CGV).

3. RESULTADOS

Un análisis preliminar del contexto [nasal+diptongo] desde el punto de vista perceptivo, indica que los informantes producen diferentes realizaciones para la consonante, algo que el análisis acústico parece confirmar. Concretamente, se detectan cuatro: como nasal alveolar (esperable), como nasal palatal, como vibrante simple (correspondiente a segmentos con estructura formántica pero de una duración que no sobrepasa los 34 ms) y algunos casos de elisión (tanto del elemento palatal como del consonántico. La proporción en la que se encuentran en el corpus queda ilustrada en las figuras 1 y 2.

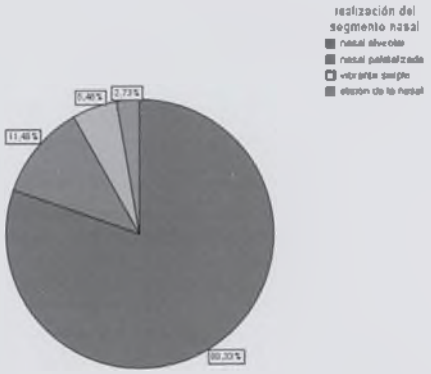


Figura 1: Realizaciones de la nasal en contexto [ni] en el informante 1.

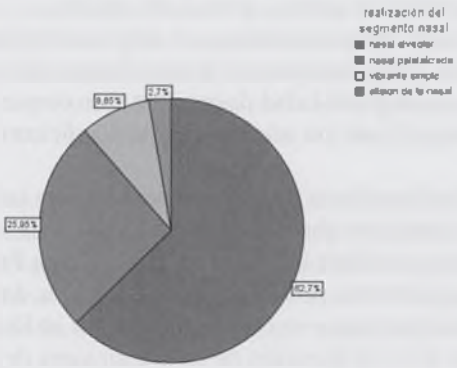


Figura 2: Realizaciones de la nasal en contexto [ni] en el informante 2.

En base a la información obtenida del análisis estadístico descriptivo, se ha procedido a la comparación entre los dos contextos tomados en consideración con el objetivo de determinar si existen diferencias significativas en la producción de la nasal. Para ello, se ha recurrido a la prueba estadística del anova, consistente en la comparación de dos o más grupos de datos, de modo que se pueda establecer si existe o no la suficiente variabilidad como para distribuirlos en grupos separados o si pueden constituir un mismo grupo.

Para mayor comodidad, se van a organizar los resultados obtenidos en el análisis comparativo de [ni] y [ni] presentando primero los relativos a la duración y, en segundo lugar, los referentes a la frecuencia de los formantes del segmento nasal.

3.1. Duración

En cuanto a la duración de la nasal, en el primer informante el valor medio obtenido para [ni] es de 56,70 ms mientras que para [ni] es de 56,03. El análisis de varianza revela que las divergencias no son apreciables, tendencia que se repite en el segundo locutor, en el que los valores medios de duración son menores (50,36 ms para [ni] y 50,73 para [ni]).

Lo mismo ocurre al estudiar los valores medios de las transiciones. Para el primer informante, la diferencia entre las transiciones desde el sonido precedente en los dos contextos tampoco son significativas, algo que se repite en las que llevan al sonido siguiente. Ello se da también en el locutor 2, en el que encontramos un valor medio de duración de la transición desde el segmento precedente de 14,87 ms en el caso de [ni] (frente a 19,70 en el locutor 1) y de 16,58 ms en el de [ni] (19,15 en el anterior informante), valores que tampoco conllevan diferencias representativas. En cuanto a las transiciones al sonido siguiente, se reproduce la misma situación: 23,58 ms de media en el contexto [nasal+vocal] y 21,26 en el de [nasal+diptongo] para el primer hablante mientras que, de nuevo, la duración es menor en el segundo: 18,99 ms en [ni] y 19,26 en [ni]. El anova arroja, otra vez, datos negativos

A priori, parece que la duración no es un elemento sobre el que pueda basarse la separación en dos grupos de la nasal dependiendo de su estructura silábica (CV o CGV). Sin embargo, se ha considerado apropiado analizar este parámetro en función de dos variables: el acento y el punto de articulación de la vocal que sigue a la nasal.

3.1.1. Influencia del acento.

Al poner en relación los datos de duración con la tonicidad de la sílaba, los resultados del análisis de varianza sí permiten distinguir dos realizaciones diferentes al evaluar la duración de la consonante nasal (F=9,049, p<0,003). Sin embargo, ello ocurre solamente en el primer informante, en el que, en sílaba átona, la secuencia [ni] es significativamente más breve que [ni] (49,59 ms y 59,31 ms respectivamente) mientras que, en sílaba tónica, la tendencia es a la inversa: [ni] dura 52,36 ms de media frente a 63,81 ms de [ni]. Como ya se ha indicado, los valores del segundo locutor no muestran divergencias lo suficientemente importantes como para advertir dos grupos distintos.

En cuanto a las transiciones, en parte se repite lo ya observado. En el segundo locutor, nada indica que sea posible la separación en dos realizaciones diferentes dependiendo del contexto; no obstante, en el primero, en sílaba acentuada, el anova sí permite realizarla (F=7,462, p<0,008): la duración de las transiciones al sonido siguiente son significativamente más largas en la secuencia [ni] que en [ni]

De modo general, se podría aventurar que el acento influye en la duración pero dependiendo del locutor.

3.1.2. Influencia del punto de articulación de la vocal.

Al establecer la comparación entre la secuencia [ni] y los distintos casos de [ni], se observa que ni el punto de articulación palatal ni el central parecen influir en la duración mientras que el velar sí lo haría en la de la transición al sonido siguiente (sensiblemente más breve en el caso de [ni]) aunque, de nuevo, ello se da solamente en el primer informante (F=9,781, p<0,003), tal como se aprecia en la figura 3. Parece ser, pues, que los valores relativos a la duración dependen, en gran medida, del locutor.

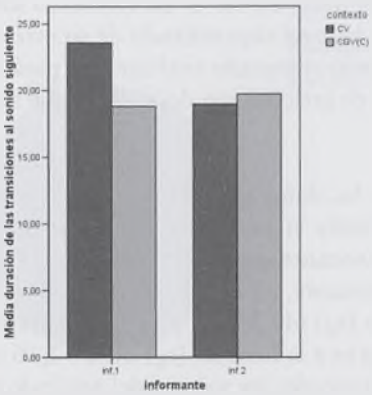


Figura 3: Valores medios de duración de la nasal en función del punto de articulación para los dos informantes.

3.2. Frecuencias

Ya desde una perspectiva general, se observan diferencias lo suficientemente representativas como para pensar en la existencia de dos grupos distintos en la realización de la nasal, algo que se advierte en las frecuencias de F2 en ambos locutores (F=4,858, p<0,029 para el primero y F=3,769, p<0,046 para el segundo), y en la de FN en el primero (F=4,305, p<0,040). En el caso de F2, las frecuencias son más altas en el contexto [nasal+diptongo] que en [ni], algo lógico puesto que este formante está vinculado con el punto de articulación del sonido en cuestión: habitualmente, los segmentos palatales presentan un F2 más alto que otros sonidos. Sin embargo, en FN, la relación es a la inversa.

3.2.1. Influencia de la tonicidad de la sílaba.

Al analizar la frecuencia en función del acento, se ha puesto de manifiesto que éste no parece influir en la altura de los formantes de la nasal, tal como se desprende de los datos de las tablas 2 y 3.

	Sílaba átona			Sílaba tónica		
	Valor medio de [ni]	Valor medio de [ni]	Significación	Valor medio de [ni]	Valor medio de [ni]	Significación
F1	418,94	426,80	0,218	407,86	410,17	0,753
FN	1037,35	1023,17	0,179	1037,61	1024,42	0,136
F2	1486,17	1594,18	0,054	1548,66	1586,57	0,308
F3	2459,17	2499,83	0,478	2506,32	2476,50	0,528

Tabla 2: Valores de significación resultantes de la comparación de los dos contextos en sílaba átona y en sílaba tónica en el informante 1.

	Sílaba átona			Sílaba tónica		
	Valor medio de [ni]	Valor medio de [ni]	Significación	Valor medio de [ni]	Valor medio de [ni]	Significación
F1	366,21	360,10	0,382	346,33	354,60	0,200
FN	1011,86	1000,12	0,379	1001,82	1000,23	0,880
F2	1723,45	1767,52	0,344	1672,25	1767,02	0,054
F3	2517,29	2532,18	0,740	2475,08	2490,76	0,622

Tabla 3: Valores de significación resultantes de la comparación de los dos contextos en sílaba átona y en sílaba tónica en el informante 2.

3.2.2. Influencia del punto de articulación de la vocal.

La tendencia común parece apuntar a la influencia del punto de articulación vocálico sobre la frecuencia del tercer y cuarto formantes de la nasal, si bien el alcance de esta relación varía dependiendo del locutor. Los datos obtenidos en un análisis más exhaustivo son, cuando menos, interesantes. Efectivamente, los resultados del anova permiten separar la realización nasal de [ni] de la de [nie] en lo que se refiere a la frecuencia de F2 tanto en el locutor 1 como en el 2 (F=19,024, p<0,000 y F=12,747, p<0,001 respectivamente). En ambos casos, la altura de este formante es significativamente mayor en el contexto [nasal+diptongo] que en el de [nasal+vocal].

La distinción no puede llevarse a cabo en el caso de [nia] puesto que la diferencia entre los valores de frecuencia de los formantes en los dos contextos no es lo suficientemente representativa, atendiendo a las pruebas estadísticas. Sin embargo, la comparación entre [ni] y [nio] muestra cierta variación en los resultados ya que parece posible distinguir la nasal de ambos contextos en base a los valores del formante nasal, significativamente más bajo ante diptongo que ante vocal (F=11,177, p<0,001), aunque ello sólo se da en el primer locutor (vid. figuras 4, 5 y 6).

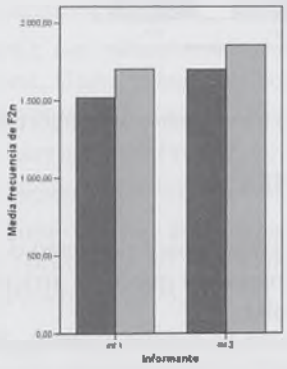


Figura 4: Valores medios de F2 en los contextos [ni] y [nie] para cada informante.

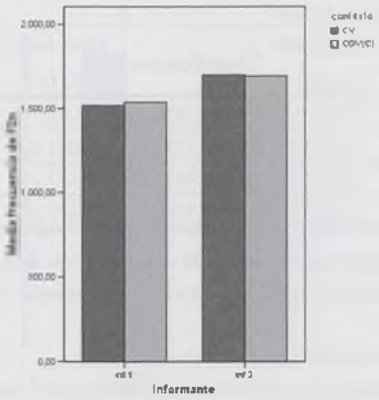


Figura 5: Valores medios de F2 en los contextos [ni] y [nja] para cada informante.

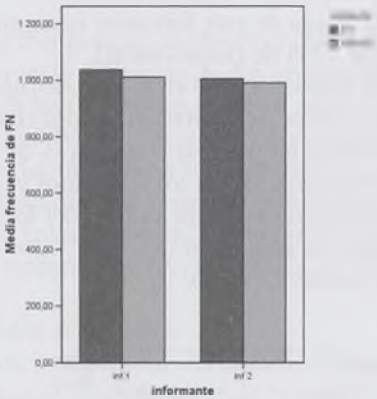


Figura 6: Valores medios de FN en los contextos [ni] y [nio] para los dos informantes.

4. DISCUSIÓN Y CONCLUSIONES

Como ya se ha indicado, este trabajo nace del deseo de comprobar si existía o no variación en el contexto [nasal+diptongo], el punto de articulación de cuya consonante ha venido considerándose como alveolar.

Vistos los resultados, se puede afirmar que, efectivamente, la realización de la nasal en este contexto no es unívoca. Así, se ha constatado la presencia de tres alternativas: o la elisión del segmento o su realización como vibrante simple o su manifestación como nasal palatalizada.

En este caso, se ha apreciado cómo los elementos que permiten distinguirla de una realización alveolar son, básicamente, una mayor altura en la frecuencia de F2 y F3. También se ha podido observar que la palatalización de la nasal ante diptongo se veía favorecida si la vocal del mismo era palatal, mientras que no se ha podido distinguir entre [ni] y [ni] en el caso de que la vocal fuera central, existiendo cierta ambigüidad, debida al factor "locutor", en el caso de que fuera velar.

Por otra parte, la tonicidad de la sílaba es un elemento que, *a priori*, no afecta la palatalización de la nasal; no obstante, para ser precisos, habría que poner, una vez más, esta variable en relación con el informante puesto que en uno de los dos casos el acento se relaciona con la duración del segmento y la de la transición al sonido siguiente a un nivel lo suficientemente representativo como para marcar diferencias entre una realización y otra.

Por último, cabe añadir, a modo de nota suplementaria, que, en un 11,15% de los casos de [nasal+diptongo], además de la palatalización de la nasal se ha producido la elisión de la semiconsonante.

BIBLIOGRAFÍA

AGUILAR, Lourdes y María Jesús MACHUCA: "Procesos fonéticos en el español actual y su relación con fenómenos diacrónicos: debilitamiento de obstruyentes, asimilación de consonantes contiguas y procesos de monoptongación" en *Actas del III Simposium de Hispanistas Polacos*, Polonia, 1993, pp. 71-86

AGUILAR, Lourdes: "Hiatus and diphtong: Acoustic cues and speech situation differences" en *Speech Communication*, 28, 1999, pp. 57-74.

ALBALÁ, Mª José: "Análisis y síntesis de las consonantes nasales" en *Revista de Filología Española*, Madrid, LXXII, 1992, pp. 37-55.

ALCOBA, Santiago: *La expresión oral*, Barcelona, Ariel, 2000.

ALVAR, Manuel (dir.): *Manual de dialectología española. El Español de España*, Barcelona, Ariel, 1996, Ariel Lingüística.

BLECUA, Beatriz [en línea]: *Las vibrantes del español: manifestaciones acústicas y procesos fonéticos*, Tesis Doctoral, Universidad Autónoma de Barcelona, 2001 [consulta: 24 de abril de 2005], www.tdcat.cesca.es/TDCat-0111102-110913.

FANT, Gunnar: *Acoustic Theory of Speech Production*, The Hague, Mouton, 1970.

GIL, Juana: *Los sonidos del lenguaje*, Madrid, Síntesis, 1999.

HUFFMAN, Marie K.: "Time-Varying Properties of Contextually Nasalized Wovels: Acoustics and Perception", *Actes du XIIème Congrès International des Sciences Phonétiques*, Aix-en-Provence, 1995, vol. III, pp. 130-133.

MACHUCA, María Jesús: *Análisis acústico de las consonantes nasales del español en habla espontánea y en habla de laboratorio* (Trabajo de investigación de tercer ciclo), Cerdanyola, Departamento de Filología Española, Universidad Autónoma de Barcelona, 1991.

MARTÍNEZ CELDRÁN, Eugenio: *Fonética*, Barcelona, Teide, 1984.

MARTÍNEZ CELDRÁN, Eugenio: *Análisis espectrográfico de los sonidos del habla*, Barcelona, Ariel, 1998.

NAVARRO TOMÁS, Tomás: *Manual de pronunciación española*, Madrid, Consejo Superior de Investigaciones Científicas, 1977.

QUILIS, Antonio: *Fonética acústica de la lengua española*, Madrid, Gredos, 1981, Manuales: 49.

QUILIS, Antonio: *Tratado de fonología y fonética española*, Madrid, Gredos, 1993, Manuales: 74.

RAE, *Esbozo de una nueva gramática de la lengua española*, Madrid, Espasa-Calpe.

SOLÉ, M. J. y J.J. OHALA: "Differentiating Between Phonetic and Phonological Proceses: The Case of Nasalization", *Actes du XIIème Congrès International des Sciences Phonétiques*, Aix-en-Provence, 1995, vol. II, pp. 110-113.

ANÁLISIS DE LA ENTONACIÓN DE LOS ENUNCIADOS DECLARATIVOS E INTERROGATIVOS SIN EXPANSIÓN EN UN HABLANTE MASCULINO DE JACA (HUESCA)

FRANCISCO JAVIER SIMÓN CASAS
Universidad de Zaragoza

1. INTRODUCCIÓN

El trabajo que presentamos trata de caracterizar la prosodia de las modalidades declarativa e interrogativa del español hablado en Jaca (Huesca). Se trata de una modalidad del español estándar similar a lo que conocemos como acento aragonés, si bien con unas características menos marcadas para el oído que el acento propio de la ribera del Ebro, el conocido de un modo popular como habla baturra.

Jaca es un núcleo urbano que se encuentra al norte de la provincia de Huesca. Es una de las ciudades más importantes del Pirineo aragonés. Resulta interesante destacar que la lengua de esta zona no ha sido siempre el castellano, sino que, hasta seguramente la primera mitad del siglo pasado, se utilizó una lengua derivada del latín con unos rasgos diferentes del castellano. Es lo que se conoce como aragonés. Esta lengua fue incluso utilizada por la cancillería. Hoy quedan numerosos documentos oficiales escritos de lo que fue esta lengua. En la actualidad, ha desaparecido casi totalmente. Solamente quedan huellas en algunos valles aislados del Pirineo. Sin embargo, estos restos no son uniformes, sino que, como es esperable en una zona bastante incomunicada, el habla es diferente en los distintos lugares.

Esta desaparición paulatina del aragonés ha tenido lugar por la creciente presión del español. En la actualidad, sin embargo, parece que esa amenaza ya no es tal a juzgar por las opiniones de los hablantes de estas lenguas. Salvo casos en los que la población de un lugar termine por desaparecer, no parece que los hablantes tengan conciencia de que

peligre. Todo ello sucede a pesar de que consideran que su lengua es un mal español o un mal catalán, en las zonas colindantes con esta lengua.[1]

Es peculiar la entonación aragonesa. Esta característica hace que un hablante aragonés sea claramente reconocido en cualquier otro lado del país. Navarro Tomás era consciente de esta especificad y señalaba que

> en los pueblos de Aragón, uniformados lingüísticamente en su mayor parte desde hace siglos por la influencia del castellano, el acento con que éste se pronuncia es, probablemente, el mismo con que dichos pueblos hablaron su antiguo dialecto aragonés antes de adoptar la lengua de Castilla. El ansotano y el cheso, restos del dialecto aragonés conservados en escondidos valles del Pirineo, coinciden esencialmente por su carácter prosódico con el acento peculiar del resto de la región. El rasgo más característico de este acento consiste en el tono relativamente alto con que de ordinario terminan las frases, aunque no sean interrogativas. En circunstancias análogas la inflexión final de una aseveración corriente termina en aragonés en una nota seis u ocho semitonos más alta que en castellano. Fuera de Aragón esta forma de entonación se encuentra también en Vasconia y Navarra[2].

Así pues, señala como característica de la entonación aragonesa la existencia de un F_0 más alto en el tonema de las declarativas de lo que lo suele ser en el resto del español. Además, destaca que esta peculiaridad es debida, probablemente, a la influencia del antiguo dialecto aragonés. Por ello, ningún sitio mejor para comprobar esta hipótesis que la población de Jaca donde hasta el siglo XX se hablaba este dialecto.

En otro trabajo[3], Navarro Tomás también alude a la cantidad vocálica y señala que la vocal final es particularmente larga en aragonés. Así pues, parece que se pueden destacar dos rasgos de la prosodia aragonesa, según este autor: tono más alto en el final del tonema de las declarativas y mayor duración de la vocal final de estos mismos enunciados. Debemos destacar por otra parte que estas afirmaciones son notas sueltas que se encuentran en sus trabajos. Además, son muy pocos los trabajos que se han realizado sobre este tema por lo que los datos con los que contamos son pocos.

2. METODOLOGÍA

2.1. Informante

Se trata de un hablante que se ajusta a las exigencias planteadas por el grupo de Amper-España. Es un informante masculino de 43 años de edad. No tiene estudios su-

---

1 Véase a este respecto Martín Zorraquino et alii. (1995) donde se pueden encontrar datos acerca de la valoración que hacen los hablantes de sus respectivas variedades lingüísticas locales.

2 Navarro Tomás (1935: 44)

3 Navarro Tomás (1917: 384)

periores. Ha nacido en esta población y ha pasado allí la mayor parte de su vida (salvo el periodo del servicio militar).

2.2. Corpus

He utilizado el corpus propuesto por Amper-España. Más concretamente, solo utilizamos los enunciados sin expansión, tanto en su modalidad declarativa como interrogativa. Ello da un total de 18 enunciados, nueve interrogativos y nueve declarativos. Cada uno de ellos ha sido repetido tres veces por el hablante.

2.3. Grabación

La grabación ha sido realizada con tecnología analógica. Para ello, se ha utilizado una grabadora Marantz CP430 y un micrófono Shure SM50. El soporte magnético utilizado ha sido un casete TDK modelo D IEC I/TYPE I de 90 minutos.

2.4. Digitalización

Una vez obtenidas en cinta las grabaciones se procedió a su digitalización por medio del editor de sonido Goldwave en su versión 4.25. Más tarde, la división individual de los enunciados, su limpieza de ruidos y la maximización de la intensidad se realizó también con el programa Goldwave, pero en su versión 5.10. A continuación, se analizaron estas oraciones por medio del programa Matlab (v. 4.0) with Simulink (v. 1.2c) tal y como se hace de acuerdo con las directrices generales del proyecto AMPER.

3. RESULTADOS

3.1. Análisis del F_0

Vamos a presentar unas curvas de F_0 resumidas del modo que explicamos a continuación, para evitar una casuística excesiva. En principio hemos presentado una media de todas las repeticiones de los comienzos oxítonos, paroxítonos y proparoxítonos y lo mismo hemos hecho para el tonema. Por otra parte, el verbo solo se ha considerado en los casos en que coincidía entre una terminación y un final de la misma naturaleza (oxítono-oxítono, paroxítono-paroxítono y proparoxítono-proparoxítono). Además, se ha reflejado el valor de F_0 del centro de la vocal en las sílabas pretónica, tónica y postónica de cada uno de los grupos, salvo en todas las terminaciones donde se ha escogido como valor final el último del F_0 de la última vocal.

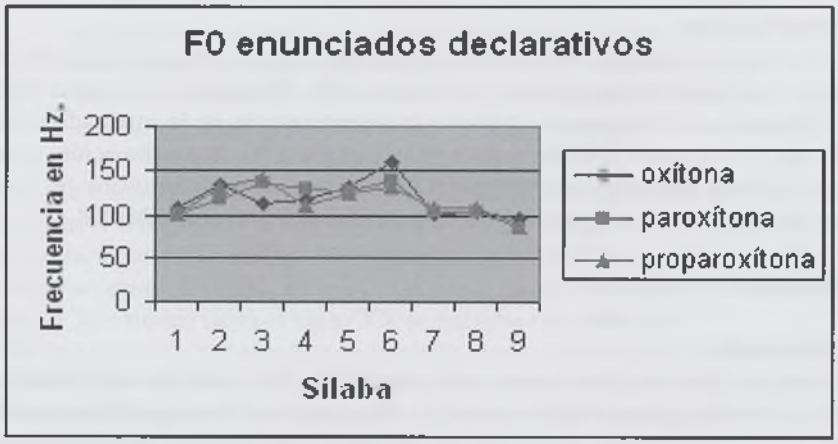
Al final, presentamos tres curvas diferentes. Cada una de ellas presenta un principio y un final de la misma naturaleza en cuanto a la posición del acento. Es decir, principio y final oxítono, principio y final paroxítono y principio y final proparoxítono. Además, en cada uno de los casos se ha puesto como valores de F_0 del verbo, como ya hemos indicado, los que correspondían exactamente al caso de la curva presentada. Por tanto, en cada ejemplo hemos presentado nueve posiciones de medición de F_0:

- 1. Valor F_0 en el centro de la vocal pretónica del sujeto
- 2. Valor F_0 en el centro de la vocal tónica del sujeto

- 3. Valor F_0 en el centro de la vocal postónica del sujeto
- 4. Valor F_0 en el centro de la vocal pretónica del verbo
- 5. Valor F_0 en el centro de la vocal tónica del verbo
- 6. Valor F_0 en el centro de la vocal postónica del verbo
- 7. Valor F_0 en el centro de la vocal pretónica del complemento de la preposición
- 8. Valor F_0 en el centro de la vocal tónica del complemento de la preposición
- 9. Valor F_0 en final del enunciado.

Y, todo ello, tanto para los enunciados declarativos como para los interrogativos.

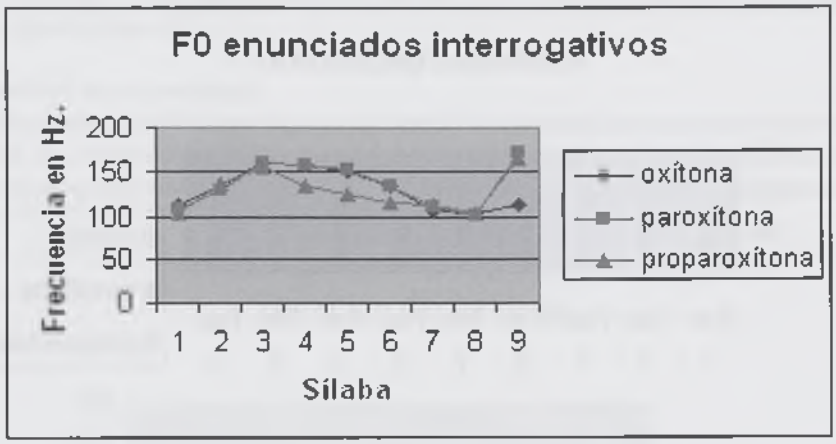
El cuadro resumen de estos valores para los enunciados declarativos es el siguiente:



Como puede observarse en el gráfico, en los inicios paroxítonos y proparoxítonos el pico tonal se produce en la postónica, frente al comienzo oxítono que tiene su pico tonal en la tónica para descender después en la postónica. Sin embargo, debe advertirse que en todos los casos el primer pico tonal coincide con la frontera del SN sujeto. En el caso del comienzo oxítono la sílaba postónica ya pertenece al verbo.

En cuanto al segundo acento tonal, puede verse un comportamiento similar en los tres tipos de estructuras acentuales. Así, desde la sílaba pretónica se ve un ascenso continuo, más acusado en la estructura oxítona, hasta el pico de la sílaba postónica. En el caso de la estructura oxítona, además, hemos de señalar que el ascenso ya comienza desde la sílaba postónica del primer acento tonal, sílaba en la que ya comenzaba el descenso en el primer acento tonal.

A continuación, se presenta el gráfico correspondiente a las estructuras interrogativas.



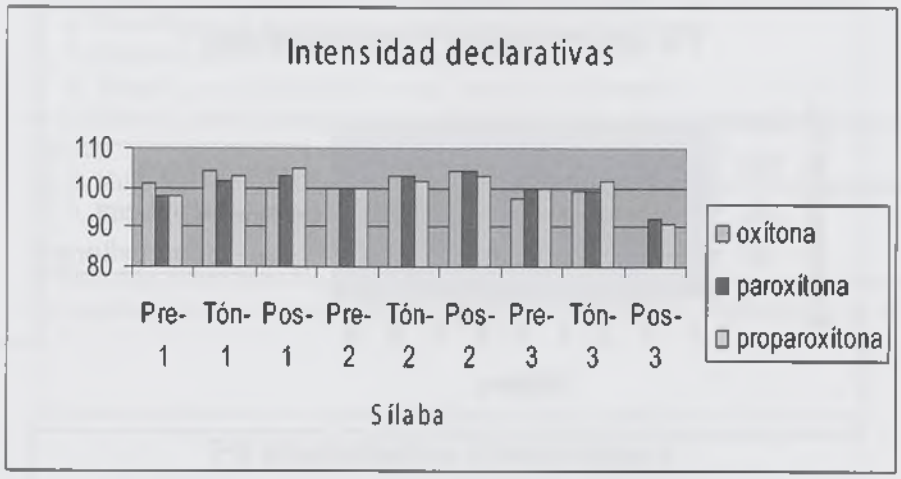
En las interrogativas, el comportamiento que observamos es similar al de las enunciativas al comienzo del enunciado. Puede verse también un primer pico tonal al final de la sílaba postónica del primer acento tonal. Sin embargo, podemos observar que en el caso de las oxítonas esta vez el pico coincide con la sílaba postónica y no con el final del sintagma, como sucedía en el caso de las enunciativas.

A partir de ahí, podemos ver un descenso paulatino hasta la sílaba tónica del tonema. En ese momento comienza un ascenso abrupto hasta el final del enunciado. Esta elevación no aparece tan marcada en el gráfico en el caso de las oxítonas pero las razones son metodológicas y no prosódicas. Sin embargo, sí que la subida real es equivalente. Recuérdese que el análisis de Amper se realiza exclusivamente sobre vocales por lo que, como sucede en este caso, si la sílaba termina en consonante sonora perdemos algo de elevación tonal.

3.2. Análisis de la intensidad

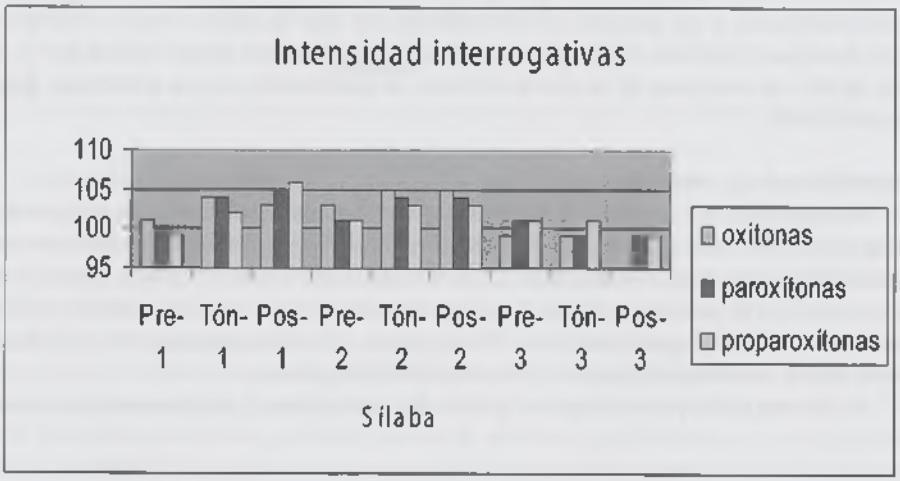
En este caso, las sílabas son las mismas que hemos considerado en el caso de los valores del F_0. Por tanto, no habrá intensidad en la sílaba postónica final de los enunciados oxítonos puesto que la sílaba final en estos casos es la tónica. Por otra parte, el valor de intensidad en la postónica del tonema corresponde al valor en intensidad de la última sílaba en el caso de las proparoxítonas. Y en cuanto a la cifra particular de la intensidad en cada sílaba, mostramos la que el programa MatLab genera.

A continuación presentamos el gráfico de intensidad en los enunciados declarativos:



Podemos ver cómo la intensidad crece significativamente en las sílabas tónicas y postónicas de cada grupo acentual excepto en el tonema. En este caso, lo que sucede es que hay una reducción muy significativa (incluso más de 10 dB) de la intensidad en la sílaba final. Esto no sucede en los finales oxítonos dado que la sílaba final es tónica.

El gráfico que corresponde a la intensidad en las interrogativas es el siguiente:

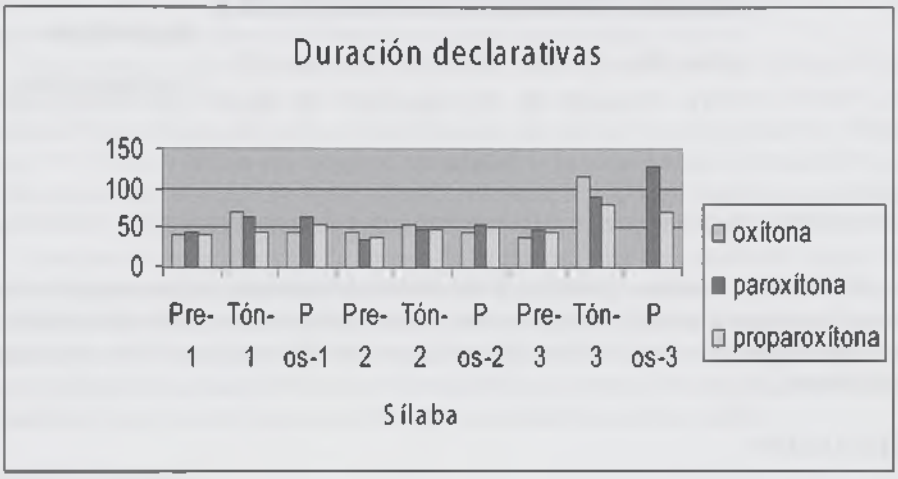


Los resultados en este caso, son similares en líneas generales a lo que sucedía con los enunciados declarativos. No obstante, llama la atención la reducción en intensidad de

todo el tonema, sin ser particularmente llamativo el caso de su postónica con respecto a todo su grupo acentual.

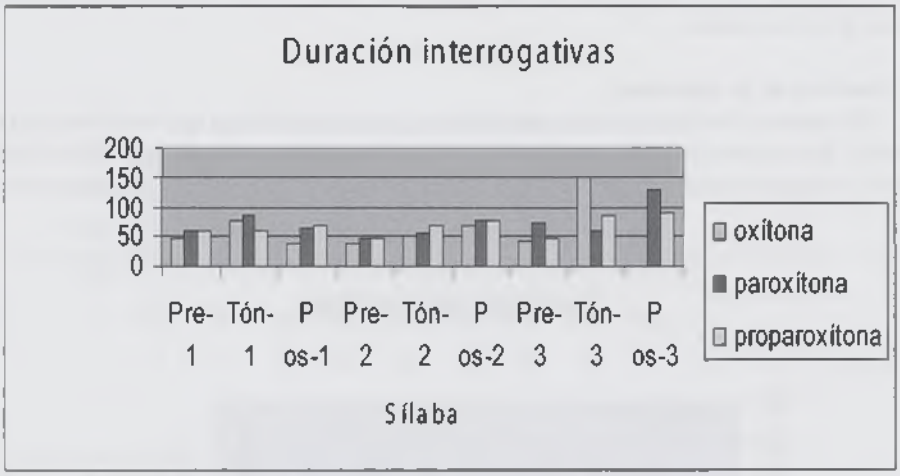
3.3. Análisis de la cantidad

De nuevo, el análisis en este caso muestra las mismas sílabas que en el caso de la intensidad. Sus valores también corresponden a lo comentado en el caso de la intensidad. El gráfico correspondiente a los enunciados declarativos es el que aparece a continuación:



Como puede verse en el gráfico, destacan claramente en cuanto a su duración las sílabas tónica y postónica del tonema con respecto a cualquier otra sílaba del enunciado. En algunos casos, se supera el doble de la duración con respecto incluso a la tónica del verbo. En la realización de las proparoxítonas no se ve un pico en la duración tan acusado. Esto se debe a que la duración postónica se debe repartir entre más sílabas que en cualquier otro caso. Si sumáramos las cantidades de las dos sílabas postónicas obtendríamos valores más acusados.

En cuanto a la duración de los enunciados interrogativos, presentamos el siguiente gráfico:



En él vemos la mayor duración de las sílabas del tonema con respecto a las demás sílabas y con respecto a la duración de estas mismas sílabas en los enunciados declarativos. En determinados casos, la duración es más del triple entre las sílabas más largas y las más breves.

4. DISCUSIÓN

En los resultados de este trabajo, observamos algunos hechos que parecen peculiares del español de Aragón. Comenzaremos con el comentario de los resultados del F_0 en los enunciados declarativos. En estos casos, podemos ver que el español de Aragón termina el tonema con una cadencia no demasiado abrupta y se podría hablar de semicadencia. Tomando como ejemplo el caso de las proparoxítonas, desde el último acento hasta el final del enunciado el descenso en semitonos es de 3,9 (109 Hz. hasta 87 Hz.). Esta cifra contrasta con los datos del español general. Según Navarro Tomás (1944), la caída del tono en español es de unos 8 semitonos desde el último acento. Esto nos da una primera diferencia entre nuestros datos del hablante de Jaca y el español general. Parece que los enunciados declarativos en Aragón (o, por lo menos, en Jaca) terminan con un tono más alto de lo que lo hacen en el español general. Si recordamos la cita de Navarro Tomás comentada anteriormente, vemos que nuestros datos coinciden esencialmente con esa aseveración. Este autor decía que en Aragón, la terminación de estas frases era de unos 6 a 8 semitonos más alta que en español. En nuestro caso, son 4 semitonos de diferencia, pero también hemos señalado que el acento de Jaca no resulta tan marcado como el de otras zonas de Aragón, principalmente el de la Ribera del Ebro.

En lo que respecta al comportamiento melódico de la frase completa, resultan visibles dos picos tonales. Cada uno coincide con la sílaba postónica de cada acento tonal

salvo el tonema ya comentado. En el caso de las oxítonas, el pico recae sobre la tónica puesto que la postónica ya forma parte del siguiente grupo acentual. Además, la subida se produce gradualmente en los dos casos desde la sílaba pretónica hasta el final.

En los enunciados interrogativos de Jaca encontramos exactamente el patrón del español general presentado por Navarro Tomás (1944). Hallamos un primer pico tonal en la postónica del primer grupo acentual. Esto se da en todos los casos, incluido el enunciado con comienzo oxítono. A partir de ahí viene un descenso gradual hasta la sílaba tónica del grupo acentual final. Este descenso oscila sobre los 7 semitonos. A partir de ahí, viene la elevación tonal característica de este tipo de enunciados de unos 8 semitonos. En este comportamiento no se observa diferencia alguna con el español general.

En cuanto a la duración, hemos señalado en la presentación de los gráficos la gran cantidad vocálica del final del enunciado, que, en algunos casos, superaba el triple de la cantidad de otras sílabas de la frase. Recordemos, de nuevo, la cita anterior de Navarro Tomás (1917) donde señala esa longitud particular de la vocal final en el español de Aragón. De nuevo, las notas de este autor sobre la entonación aragonesa vuelven a confirmarse en nuestros resultados. Los gráficos presentados señalan claramente esta característica.

Finalmente, el análisis de la intensidad presenta una característica clara y es el descenso notable de esta propiedad prosódica en las sílabas postónicas del último acento tonal. Esta característica es tan considerable que, en algunas frases particulares, llegaba a dificultar el análisis por parte del software informático. No conocemos ningún trabajo en el que se hable de esta característica de la prosodia en el español de Aragón, pero creemos que puede ser también un hecho peculiar de la entonación de esta región.

5. CONCLUSIONES

Los datos obtenidos en nuestras encuestas y su posterior análisis nos han permitido observar una serie de hechos peculiares en la entonación del español de Jaca, que es posible que se puedan extender al español que encontramos en otros lugares de Aragón. Los hechos relevantes son los que citamos a continuación:

1. Descenso final en los enunciados declarativos menos acusado que en el resto del español general. Concretamente, en nuestros ejemplos oscila sobre los 4 semitonos frente a los 8 semitonos señalados por Navarro Tomás para el español estándar.
2. Cantidad ostensiblemente más larga de las sílabas finales, especialmente la última, con respecto a cualquier otra sílaba del enunciado.
3. Intensidad reducida en las sílabas posteriores a la tónica en el último acento tonal.

Resulta razonable pensar que estas tres características en su conjunto sean las que determinen la peculiaridad del llamado acento aragonés. Es un tipo de entonación claramente perceptible por cualquier español y que adscribe geográficamente, sin ninguna duda, a los hablantes de esta variedad del castellano.

BIBLIOGRAFÍA

CASTAÑER, R. M.; M. P. GONZÁLEZ; J. SIMÓN (2005): "Aproximación al estudio de la entonación aragonesa", en *Estudios de Fonética Experimental*, XIV pp. 273-293.

CONTINI, M.; C. FRANCHON; A. RHARDISSE (1998): "Analyse comparée de l'intonation en castillan et en aragonais", en *Atti XXI Congresso Internazionale de Lingüistica e Filologia Romanza*, Tubinga: Max Niemeyer Verlag, vol. V, pp. 143-157.

FERNÁNDEZ PLANAS, A. M.; E. MARTÍNEZ CELDRÁN (2003): El tono fundamental y la duración: dos aspectos de la taxonomía prosódica en dos modalidades de habla (enunciativa e interrogativa) del español", *Estudios de Fonética Experimental*, XII pp.165-200.

MARTÍN ZORRAQUINO, M. A.; M. R. Fort; M. L. ARNAL; J. GIRALT (1995): *Estudio sociolingüístico de la Franja Oriental de Aragón*, Zaragoza, DGA-Universidad de Zaragoza.

NAVARRO TOMÁS, T. (1917): "Cantidad de las vocales inacentuadas", *RFE*, IV, pp. 371-388.

NAVARRO TOMÁS, T. (1935): *El acento castellano. Discurso leído por el autor en el acto de su recepción académica el día 19 de mayo de 1935*, Madrid: Tipografía de Archivos. Olózaga, I.

NAVARRO TOMÁS, T. (1944): *Manual de entonación española*, Madrid: Guadarrama. Se ha manejado la 3ª edición, México, Colección Málaga, 1966.

EXPERIMENTAL PHONETICS APPLIED TO PORTUGUESE AS SECOND LANGUAGE: REMAINING AND DETECTABILITY OF NATIVE LANGUAGE CHARACTERISTICS

António J. S. Teixeira
Departamento de Engenharia Electrónica e Telecomunicações/ IEETA, Universidade de Aveiro ajst@det.ua.pt

Helena Margarida Vaz Duarte
Centro de Línguas e Culturas & Departamento de Línguas e Culturas, Universidade de Aveiro hvazduarte@dlc.ua.pt

1. INTRODUCTION

The experience of one of the authors of Portuguese as Foreign Language (PFL) teaching motivated a reflection on the difficulties that foreign students, even of advanced level, feel in the accomplishment of certain sounds or classes of sounds.

It is a well-known fact that, during the process of teaching learning of a foreign language, it is considered fundamental, beyond the knowledge and domain of the written text, that the pronunciation of linguistic statements in a given context should be more or less spontaneous and fluent.

The PFL students are confronted with some problems in the production of the verbal text, due, obviously, to their level of knowledge of the language and the situational contexts where they produce them.

In a more advanced phase of the learning of the Portuguese language, where the students already have a good grasp of the language at the lexical and syntactic level, have emerged certain phonetic or phonological problems, which are apparently not connected to the linguistic competence that the students demonstrate at other levels.

The proximity among the languages facilitates the process of learning a second language, whenever the structures coincide, thus there is no apparent necessity to acquire new knowledge, but rather a transposition of structures, which will attenuate the error possibility. On the other hand, as far as language production is concerned the differences can cause major problems, contributing to an imperfect pronunciation. We believe that the situation is aggravated in the production by those who possess Romanic language as their mother tongue.

In view of this perception we intend, with this brief study, to analyze certain problems of the pronunciation of Portuguese statements by foreign students, who have Spanish, French, Italian and Rumanian as their mother tongues. In fact, this kind of students, possessing an advanced knowledge of the Portuguese language, still keep, in its verbal speech, phonetic characteristics of its mother tongue.

It is our intention to investigate whether the biggest problems in their production of Portuguese are more related to the production of the most distinct sounds and even indistinct sounds in the original language of the students, especially concerning the vowels.

Therefore, our objectives are the identification of the recurrent difficulties, in production terms, of Latin language students, the analysis of these problems, based on experimental documents, and the relation of these production problems with the mother tongue of the students.

2. METHODOLOGY

The work consisted in the recording of verbal output of advanced levels students, annotation of detected problems, as well as the analysis and characterization of the more common problems.

3. CORPUS

3.1. Corpus definition

Regarding corpus and due to our objective, our choice was to avoid isolated words or phrases and select a brief narrative text. We thus considered that the isolated recording of words or phrases would not allow us to obtain a continuous speech close to the spontaneous one. We found that that making students speak more spontaneously increases the occurrence of problems. We consequently opted for the type of above cited text, of easy understanding for the students. To reach our objectives and to select the text to record, we considered diverse criteria. Therefore the text would have to include:

- Several types of phrase (declarative, interrogative,...) that would allow a reading production concerning several intonation levels;
- Long words, previously considered more difficult in their pronunciation. As an example we have, "filosoficamente" (philosophically);
- Paroxytone words that are read by the foreigners as oxytones, as are verbal forms of the past perfect, in the 3rd person of the plural ("arrumaram");
- Several types of sounds (vowels and consonants) that, in accordance with our experience, would cause difficulties to the students. Examples are: taps and trills [R], [r], secondary stressed vowels [ɐ] ("mesa" [mezɐ]) and [i] ("dezanove" [dzɐnɔvi]), the raising diphthongs ("idiota" [idjɔtɐ]) and falling ones ("perdeu" [pirdew]).

In accordance with these established criteria, we selected, almost arbitrarily, an excerpt of a Portuguese contemporary novel (see below) which was unknown by the students (AFONSO, 2004).

> As dezanove horas e trinta minutos os formandos arrumaram o que lhes pertencia — alguns deles colocaram os trabalhos na mesa oval situada no centro da sala — e começaram a sair. Carolina, acompanhada por Aurélio, agitava distraída o cigarro e gesticulava quando uma luz forte lhe bateu contra o rosto e ela perdeu o fio à conversa. Atrás deles, Paulo, o formador, e Emília, professora de comunicação, trocavam impressões.
>
> Um novo flash, de luzes no máximo, incendiou o rosto de Carolina.
>
> — Estúpido!
>
> Aurélio olhou para o lado esquerdo e depois para trás. Estaria a falar com ele?
>
> — Como? — inquiriu filosoficamente.
>
> As luzes voltaram a acender e a apagar.
>
> — Desculpe. Refiro-me ao idiota que está a mandar os máximos para aqui. — esclareceu Carolina.

Table 1 – The excerpt select for recordings, extracted from (AFONSO, 2004).

3.2. Speakers

The speakers were all foreign students, who, at the time, attended the Advanced Level of the PFL course, in the academic year of 2004/2005. Recordings contemplate Latin and non-Latin languages and, if possible, more than one speaker for each language. Information regarding all the speakers recorded is presented in Table 2. For the present study only the material regarding four speakers having a Latin language as mother tongue was used. Analyses of the remaining languages will be presented elsewhere.

All four selected speakers were attending different undergraduate courses. They were aged between 23 and 33.

Mother Tong	Country	Gender	First Name
Rumanian	Rumania	1M	Adrian
French	France	2F	Céline, Eva
Italian	Italy	1F+1M	Enrica, Fabio
Spanish	Spain	1F	Raquel
(Bilingual) English	Canada	1F	Sandra
Russian	Ukraine	1F	Iana
Russian	Byelorussia	1M	Aliaksander
German	Germany	2F	Carla, Aimara

Table 2 – Information regarding speakers recorded.

3.3. Recording

The excerpt, which was previously distributed, was read and recorded in Portuguese, as well as in the mother tongue of the students. There is a file corresponding to each one. The students also translated the excerpt into their respective mother tongues.

The recording was carried out using SFS (Speech Filling System) (http://phon.ucl.ac.uk/sfs) and directly recorded to the hard disk computer using the equipment Kay CSL 4400. Recordings were made in the University of Aveiro Laboratory of Phonetics.

3.4. Annotation

Figure 1 – An annotation example. At the top, the speech waveform; next, the orthographic annotation (words); at bottom, the phones where problem were detected.

We started with the orthographic annotation, delimiting the beginning and the end of words. At this stage the annotator also kept a log of the detected problems. The existence of difficulties in the same word, despite the difference of the student's native language, was verified. Words such as "arrumaram", "pertencia, "Carolina", "acompanhada", "agitava", "distraída", "cigarro", "gesticulava", "perdeu", "Emília", "professora", "filosoficamente", "Estúpido", "Desculpe" and "esclareceu" suffered several phonetic alterations.

In the presence of the variety phonetic alterations and as a detailed analysis of all these situations would not fit within the scope of this article, we opted to analyze only the most recurrent problems. For this analysis, we proceeded to an annotation of the phones including both the correct/intended pronunciation and the actual pronunciation made by the speaker. For example, a phone, that if pronounced correctly would be an [ɐ] and is produced by the student as a more open [a] is annotated as [ɐa]. We used the annotator perceptual judgment for the attribution of the actual production label.

3.5. Parameter Extraction

After the annotation, values for the first two formant frequencies were obtained by using the SFS system. Using SFS language SML values of those formants and phones durations were automatically extracted and exported to a data file in a format suitable for statistical programs such as SPSS or R.

4. RESULTS

The analysis, made in SPSS and R language was focused on the study of the vowels, responsible for great part of the detected problems. We considered some aspects: the study of the occurrences presented in illustrative graphs, the observation of the duration and of the formants, as well as the study of F0. As far as consonants are concerned, in this first phase, we only intend to present an analysis of their occurrences.

4.1. Quantitative analysis of the detected problems

In Figure 2 we present the counting of the detected problems.

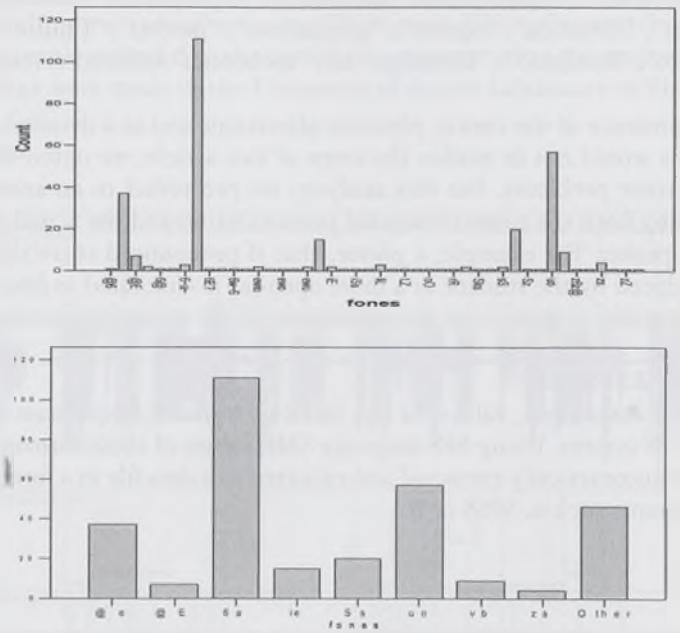


Figure 2 – Bar plot showing the number of occurrences of the several problems detected/annotated. At the top bar graph all problems are included; the bottom graph only includes the errors representing more than 1 % of the total. In de figure {6}>{a}, {@}>{i}, {O}>{ɔ}, {E}>{ɛ}, {w~}>{w̃}, {6~}>{ɐ̃}, {a~}>{ã}, {S}>{ʃ}, {Z}>{ʒ}, {L}>{ʎ}.

It is clear that some problems are much more common than others. In the first group we have, from left to right in the Figure: {i}>{e} (meaning an {i} pronounced as an {e}), {ɐ}>{a}, {i}>{e}, {ʃ}>{s}, {u}>{o}, and {v}>{b}.

Being vowels the most problematic area, in Figure 3 we present a pie chart with the distribution of the annotated problems regarding vowels, collapsing data from all the speakers.

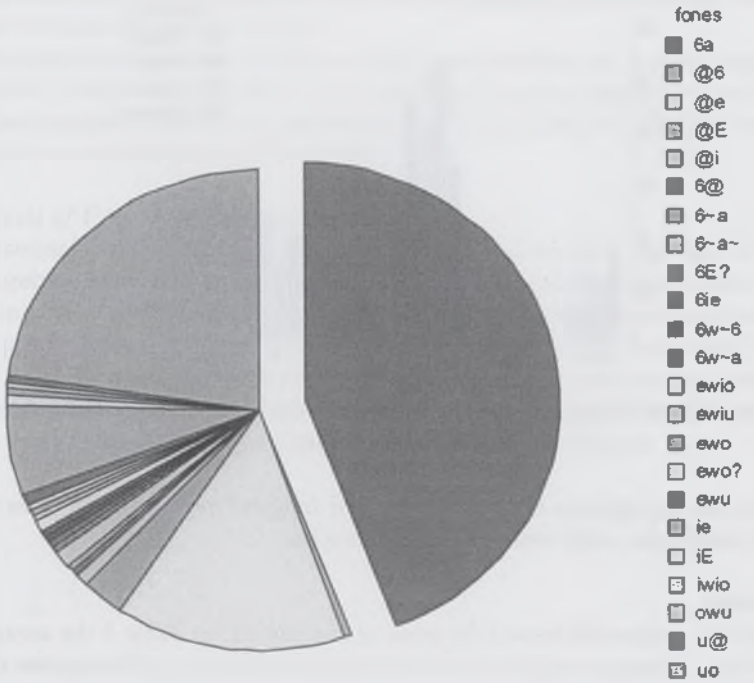


Figure 3 - General vision of the problems regarding vowels production detected. In de figure [6]>[a], [ɐ̃]>[ĩ], [E]>[ɛ], [w~]>[w̃], [6~]>[ɐ̃], [a~]>[ã]

It is significant that problems like [ɐ]>[a], [i̇]>[e], [i] > [e] and [u]>[o] are the most common.

Many incorrections were detected in the production of the sound [ɐ] that was mostly produced as [a]. We could verify it in words as "Carolina" (Caroline) that is pronounced [karulina], "acompanhada" (accompanied) pronounced [akõpaŋada], "trinta" (thirteen) pronounced [trĩta] and "dela" (her) produced as [dɛla].

There are three problems with the i̇. First it is produced as [e] in "professora" (teacher), pronounced [prufesorɐ]; "refiro-me" (I refer to) [Refirumi̇]; "perdeu" ([she] loose) [perdew], "gesticulava" ([she] gesticulated) [ʒeʃtɪkulavɐ]; "impressões" (impressions) [ĩpresõjʃ]. In a second version it is produced as [ɛ], eg. "pertencia" ([it] belonged) [pɛrtẽsiɐ], "forte" (strong) [fɔrtɛ], "luzes" (lights) [luzɛʃ]. The third changes [ɨ] to [i]. Examples: "dezanove" (nineteen) [dizanɔvi], "desculpe" (sorry) [diʃkulpi].

The sound [u] pronounced as [o] appears in "Carolina" (Caroline), pronounced [kɐrolinɐ]; "professora" (teacher) [profisorɔ], "Paulo" (Paul) [pawlo].

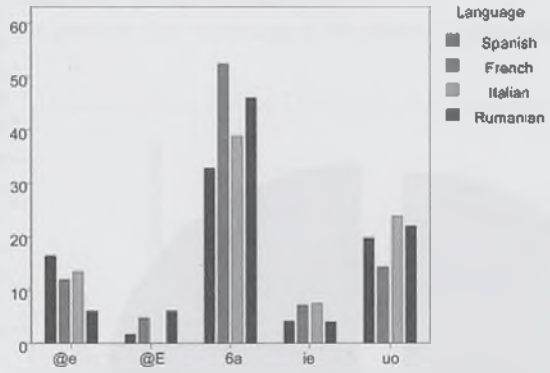


Figure 4 - Distribution by language of some of the most common problems. In the figure [@] = [i]; [E] = [ɛ]; [6] = [ɐ];

From the figure, it is clear that the most frequent problems are common to the four native languages, with some minor differences.

4.2. Duration

Our first analysis addressed duration of the vowels. In Table 3 the average for our corpus and of a previous study of our research group covering six Portuguese regions are presented (Teixeira et al. 2004).

Our *corpus*					EP (Teixeira et al. 2004)	
Vowel	Spanish	French	Italian	Rumanian	Vowel	Mean value (ms)
[ɐ] > [a]	77.62 (N=40)	82.24 (N=21)	68.04 (N=26)	75.54 (N=22)	[ɐ]	66
					[a]	122
[i] > [e]	76.61 (N=5)	105.7 (N=2)	112.7 (N=5)	162.2 (N=1)	[i]	104
					[e]	112
[u] > [o]	71.05 (N=24)	50.95 (N=4)	62.8 (N=15)	79.26 (N=11)	[u]	66 (unstressed) 126 (stressed)
					[o]	116

Table 3 - Duration information for 3 of the most common problems in vowels.

In a general analysis, it is evident that durations aren't always close to the average duration values of native speakers. For example the duration for the [i]>[e] substitution of the Spanish speaker is considerably shorter that average durations for both [i] and [e] of native EP speakers. Another general tendency is in the use by speakers of the 4 lan-

guages across several phones which were closer in duration to the average value for the unstressed vowel of EP native speakers.

The duration of vowel sounds emerged as a problem for correct pronunciation of EP vowels. Further analyses should contemplate the stress/unstressed position of the vowels. Such analyses couldn't be performed due to the lack of information regarding stress in the annotations presently available.

4.2. Analysis of First Two Formants

As second step in our analyses, we looked at the two first formants for the most common problems regarding vowels. As an example of the analyses performed, in Figure 5 we present the results regarding the incorrect pronunciation of the unstressed [ɐ], perceptually produced as a lower vowel (possibly [a]); the use of [o] in place of [u]; and [i] produced as [e]. Due to space limitations only results for the analyses were performed for the male Rumanian speaker. We include average values for an EP female native speaker (data recorded during an ongoing study of Portuguese diphthongs).

Figure 5 – F1/F2 values for 3 of the most common "mistakes" in the pronunciation of EP oral vowels by the 3 male speakers analyzed. For comparison purposes average values of a Beira Littoral region native is presented at bottom right. The ellipses enclose 1 standard deviation around the mean value. **In the figure [6] = [ɐ].**

Some problems in the determination of F1 in [ɐ]>[a] were hand corrected using EMU.

In general, there is some dispersion of the formant values and formants values indicate the use of tongue positions, different to those usually used by EP native speakers. Also there is a concentration of configurations used in F1/F2 space toward the neutral schwa like configuration.

5. DISCUSSION/CONCLUSIONS

In this brief study, we present a first analysis of a corpus obtained from speakers of Portuguese second language productions.

The adopted methodology allows us to identify the recurrent difficulties in production terms and can represent a valuable tool for the process of teaching learning Portuguese, as well as offering the possibility of a greater awareness of main problems and, therefore, a more effective correction and consequent progress on the part of the students.

As a first step, this analysis and annotation process has a qualitative facet and does not seem to offer unsurpassable difficulties to a user who does not have any experience of Experimental Phonetics. A PFL teacher will therefore be able to carry out this type of work, at least in the first two analyses presented (counting/inspection of problems and durations).

From the analysis, it emerges that the main problematic area is clearly the vowels, stressed or non-stressed and to a lesser degree, some fricatives. Thus, as we show in Figure 6, we are able to conclude that the **vowels, when in unstressed position, are systematically substituted by a higher vowel** ([ɐ] by [a]; [i] by [ɛ] and [u] by [o]) **or they stay as high** ([i] substituted by [i]).

It is interesting to notice that these students always substitute the only unstressed vowel in Portuguese [i], and we know that a native speaker never produces it. It is also curious that the vowel [ɐ] that occurs very rarely in stressed position (examples: "cama", "manha") is always substituted by [a], with occasional change of the stress position. It is possible to see these facts as an inversion of the process occurring in the secondary stressed vocalism of Portuguese, where rising and centralization occurs (example: [ɛ]>[i]).

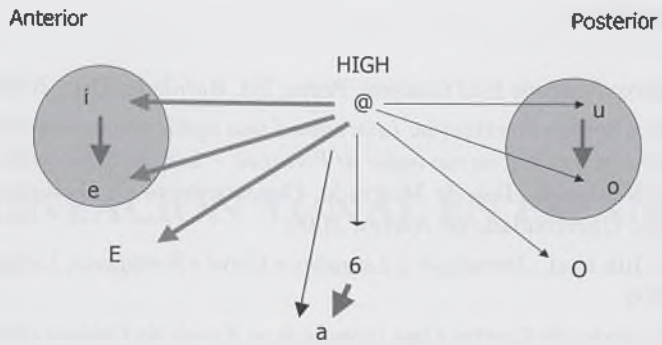


Figure 6 – Diagram showing the substitutions of vowels made by the students.

We have also verified that the duration issue also brings difficulties to these students. It is related to a vocalism problem connected to stress. In fact, a PFL student, in presence of the complexity of this aspect of the Portuguese language, has difficulties in distinguishing the primary stressed sounds from the secondary stressed ones, due to the existence, among others factors, of the different ways of a vowel pronunciation.

As far as the consonants are concerned, we confirmed that they don't cause so many problems.

5.1. Suggestions for future work

We intend to conduct a perceptual test, using both Portuguese and native language recordings, to access in what extent native language remains detectable in their recordings of Portuguese.

We consider that this study may be a contribution to a greater research, comparing, for example, the data presented here with phonetic and phonological information of the native languages of the speakers. Besides, we think that it would be equally interesting to extend to the study the languages of non-Latin origin, among students attending the same level of learning. We also believe that it will be possible and useful to observe the accomplishment of these sounds among students of other levels of learning.

The systematization of the difficulties detected in speakers of different languages and with different knowledge of the Portuguese language will constitute a useful tool for the PFL teacher.

Concerning to the prosodic analysis, that was already initiated, although not exhaustive due to the scarcity of examples, we consider that it constitutes a useful work, which should be continued.

6. BIBLIOGRAPHY

AFONSO, Antero: *Ninguém Está Contente*, Porto, Ed. Raridade, Dec. 2004.

BRINK, Antonia Sophia Geertruida: *Leite creme é uma espécie do vosso Vanillevla: Mobilização de conhecimentos prévios nas aulas de Português – Língua Estrangeira no ensino Universitário Neerlandês*, Tese de Mestrado, Departamento de Didáctica e Tecnologia Educativa, Universidade de Aveiro, 2000

FARIA, Isabel Hub et al.: *Introdução à Linguística Geral e Portuguesa*, Lisboa, Ed. Caminho, 1996

MOUTINHO, Lurdes de Castro: *Uma Introdução ao Estudo da Fonética e Fonologia do Português*, Lisboa, Plátano Editora,

TEIXEIRA, António; VAZ, Francisco; MOUTINHO, Lurdes Castro; COIMBRA, Rosa Lídia; LISBOA, Raquel Castro: «Para a melhoria da síntese articulatória das vogais nasais do Português Europeu: Estudo da duração e de características relacionadas com a fonte glotal», in *Revista de Estudos da Linguagem/ The Journal of Language Studies*, Faculdade de letras da UFMG, vol. 12-2, Jul./Dez. 2004, pp. 65-92.

VAZ, Margarida, et al.: «Quanto mais pior: considerações acerca da aquisição das vogais orais do Português Europeu por hispanófonos», in *Cadernos de PLE 3*, Centro de Línguas e Culturas, Universidade de Aveiro, 2003

ALINEACIÓN TONAL EN ESPAÑOL

GUILLERMO TOLEDO
Conicet, Argentina y Université Laval, Canadá

1. INTRODUCCIÓN

El modelo métrico y autosegmental (AM) es lineal, el contorno de entonación se genera por la simple unión de elementos tonales, los acentos. Estos acentos tonales pueden definirse como un rasgo tonal del contorno entonativo, son cambios tonales. Están asociados a un máximo tonal o a un mínimo tonal, pueden ser analizados medio de niveles tonales o estados estáticos, Alto (H) y Bajo (L) (Ladd, 1996: 43). Esto señala que la sílaba a la que están asociados es prominente, esto es métricamente fuerte (Ladd, 1996: 46-47). Así, los acentos tonales son el elemento constructor del contorno entonativo (Ladd, 1996: 286). No se interrelacionan con el contorno tonal para formar el patrón de entonación, son el contorno en sí mismo Ladd, 1996: 79-112).

La prominencia tonal de cada acento tiene diferentes realizaciones fonéticas que reflejan niveles fonológicos equivalentes, la alineación entre el tono y el texto. El pico alto de F0 no es una propiedad fonética de una sílaba prominente, sino un elemento de la estructura fonológica de la frase, en el mismo nivel que la sílaba prominente en sí misma. Por lo tanto, es coherente investigar los detalles fonéticos de la asociación entre el pico de F0 y la sílaba, es decir, la alineación temporal del pico con respecto a la sílaba (Ladd, 1996: 55). En la prominencia alineada con la sílaba acentuada, los tonos resultantes son H o L (Alto y Bajo según la fonología autosegmental). Estos dos tonos son monotonales. En esta taxonomía se indica el asterisco como grafema que señala el alineamiento del tono a la sílaba acentuada, entonces, H* y L*. En la prominencia H alineada con la sílaba inacentuada precedente se realiza un pico en esa sílaba precedente y un tono L en la sílaba acentuada. El efecto es de preshooting 'prerrealización del pico', es decir, el tono es H + L* (Avesani, 1990). En la prominencia H alineada con la sílaba inacentuada siguiente se actualiza un pico en esa sílaba siguiente y un tono L en la sílaba acentuada. El patrón es de overshooting 'posrealización del pico' y el tono resultante es L* + H (Ladd, 1996: 211; Avesani, 1990).

Según los criterios de lectura con respecto a la asociación del tono con el texto, los resultados del español son diferentes y contradictorios. Según Face (2000) y el grupo ToBI, en los enunciados declarativos se observa consistentemente el empleo de dos tipos de tono: L+H* en posición nuclear y L*+H en posición prenuclear. Las diferencias de realización se atribuyen en este caso a la existencia de diferentes categorías fonológicas. Prieto (1998) y Prieto *et al.* (1995, 1996) indican la realización de un solo tono H*, con el diacrítico estrella en la sílaba acentuada y fuerte. Hualde (2002) observa la presencia también de un tono, pero (L+H)*, con dos estados estables en la sílaba acentuada, indicado por el paréntesis. Sosa (1999) observa una multiplicidad de tonos, aunque existe un frecuencia de realización de tonos del tipo L*+H (Díaz Campos, 2002; Mc Gory y Díaz Campos, 2002). El problema de la alineación tonal es también crucial en otras lenguas: en griego (*et al.*, 2000; en catalán (Prieto, s.d.); en catalán central, italiano de Nápoles e italiano de Pisa (Prieto *et al.*, s.d.). Para resolver este problema en algunas lenguas romances ya indicadas, Prieto (s.d.) y Prieto *et al.* (s.d.) encuentran una solución basada en dos asociaciones del tono y la unidad métrica: (1) la codificación de la asociación fonológica primaria entre el tono H o L y la unidad métrica, preferentemente, la sílaba acentuada, y (2) la codificación fonológica secundaria, los tonos H o L anclados a alguna frontera prosódica (moras, sílabas o palabras fonológicas). El resultado es sumamente abstracto y de difícil comprobación experimental.

En este trabajo se propone una revisión del problema de asignación tonal basado en criterios diferentes. Se proponen varias estrategias para la solución de este problema: (1) un criterio de segmentación y medida tonal, (2) relaciones psicofonéticas en el contexto tonal y en la sílaba acentuada y fuerte, (3) decisiones con respecto al diacrítico estrella. Se observan y se presentan soluciones en corpus hispanoamericanos, peninsulares e insulares.

2. PROCEDIMIENTOS

2.1. Corpus de Buenos Aires

Es un corpus de narrativa espontánea emitido por un hablante masculino adulto, de nivel sociolectal alto. En el material global se seleccionan frases entonativas y luego los acentos tonales en el pretonema o antes de pausas intermedias o la frontera final. No se seleccionan las palabras con fenómenos fonéticos de coarticulación. Esta realización fonética es de gran frecuencia en los corpus de habla espontánea. Además, se descartan los acentos tonales con baja intensidad en la realización.

2.2. Corpus de Cuba

Se analizan los materiales emitidos por un hablante masculino adulto de nivel sociolectal alto, los discursos semiespontáneos se obtienen en una mesa redonda difundida en una emisión radiofónica. Se analizan las frases entonativas y los acentos tonales. Se estudian primordialmente los pretonemas; la causa es que en el discurso de este hablante, las pausas perceptivas posteriores a un tonema son excepcionales, se registran solamente

algunos casos. En los corpus se estudian los acentos tonales paroxítonos, la frecuencia de aparición más relevante.

2.3. Corpus de Madrid (CREA)

El corpus de español peninsular, dialecto de Madrid, es parte del Corpus Oral del Castellano, una muestra de discurso humanístico (Laboratorio de Lingüística Informática, Universidad Autónoma de Madrid, Real Academia Española. Se analizan los discursos de un hablante femenino de nivel sociolectal alto, emitidos en una emisión radiofónica. Los materiales son segmentados en frases entonativas y luego se estudian los acentos tonales que integran los pretonemas, preferentemente.

2.4. Corpus de Castilla (DIES)

Las muestras se toman del proyecto DIES-RTVP (*Difusión Internacional del Español por Radio, Televisión y Prensa*, un corpus que registra los materiales lingüísticos de los medios de comunicación social, en este caso, los medios de España Peninsular (proyecto de la Universidad de Alcalá de Henares). Los materiales elegidos son radiofónicos. Se seleccionan sobre seis tipos de programas o estratos: informativos, debates, retransmisiones deportivas en directo, comentarios deportivos, programas culturales y miscelánea. Se graban 60 medias horas (1800 minutos). De ese material global se eligen los estratos debates (religioso). El debate registra la producción de un hablante masculino, castellano, con un nivel sociolectal alto. La taxonomía registra los acentos tonales obtenidos sólo en el pretonema.

2.5. Corpus de Tenerife

Se analiza un corpus radiofónico perteneciente a un hablante femenino de Santa Cruz de Tenerife. El nivel sociolectal de la hablante es alto, es una animadora radial. Los materiales se segmentan previamente en frases entonativas, luego se seleccionan los pretonemas. Se extractan los acentos tonales paroxítonos integrados por dos a cinco sílabas.

3.1. Espacio tonal

Se segmenta el espacio tonal de cada hablante (el máximo valor tonal al mínimo valor tonal) en microespaciones tonales, en cada acento tonal. La segmentación se efectúa según umbrales psicofonéticos (las diferencias apenas perceptibles entre tonos contrastados: 1.50 semitono). Dentro de cada microespacio tonal rigen los niveles estáticos alto (H) y bajo (L).

3.2. Umbrales tonales

Pamies Bertrán *et al.* (2002) realizan tests de discriminación para determinar los umbrales tonales en donde rigen los niveles estáticos alto y bajo. Los estímulos son oposiciones de sílabas [*ba ba*] de habla natural. Se manipula la señal natural por medio de cambios tonales sintetizados. En la producción del hablante masculino se utilizan pares de estímulos con diferencias de 0.54 semitono, 0.99 semitono, 1.51 semitono y 2.07

semitonos. En la producción del hablante femenino se utilizan pares de estímulos con cambios de la fundamental de 0.47 semitono, 0.99 semitono, 1.51 semitono y 1.97 semitono. Los estímulos se presentan a un panel de oyentes, dialectos de Barcelona y de Granada, para obtener juicios de discriminación psicoacústica. Los resultados indican que los oyentes pueden discriminar diferencias tonales entre sílabas adyacentes con diferencias de 1.50 semitono. Estos hallazgos son similares a los obtenidos en otras lenguas germánicas (Nooteboom, 1997).

3.3. Segmentación acústica

El cálculo de la prominencia tonal se realiza en el pico de la fundamental en cada sílaba acentuada, fuerte, y en el contexto fonético a izquierda o a derecha de esa sílaba acentuada. En la medición de contornos silábicos de prominencia menos definida, éstos son los criterios. En las sílabas en que el contorno es plano u ondulante se alinea el contorno de la fundamental con el contorno de intensidad. El punto de medición es el valor tonal más alto que se alinea con el valor más elevado en el contorno de intensidad (ver Ladd, 1988). Los tonos en las sílabas con oclusivas se miden en el estado estable de la vocal y no en el valor más alto, en el valor de la sonoridad inmediata a la explosión (ver Eady *et al*, 1986). Se utiliza el *Speech Analyzer* 1.5 *Test Version* 15.3 (*Summer Institute of Linguistics*).

3.4. Lectura fonética y alineación tonal

En la prominencia dentro de la sílaba acentuada, fuerte, los primitivos fonológicos son **H*** o **L***, monotonales, la alineación con el diacrítico estrella. El contexto fonético a derecha y a izquierda está bajo el umbral perceptivo, 1.50 semitono. Las combinaciones bitonales con un tono opuesto en la sílaba inacentuada, a derecha o a izquierda dan estas manifestaciones de superficie estas alineaciones y estos primitivos fonológicos. En la prominencia H alineada con la sílaba inacentuada precedente se realiza un pico en esa sílaba precedente y un tono L en la sílaba acentuada. El efecto es de *preshooting* 'prerrealización del pico' y el primitivo fonológico es H + L*. En la prominencia H alineada con la sílaba inacentuada siguiente se actualiza un pico en esa sílaba siguiente y un tono L en la sílaba acentuada. El efecto es de *overshooting* 'posrealización del pico' y el tono fonológico es L* + H. Cuando todo el contexto fonético débil, a derecha y a izquierda, manifiesta valores tonales sobre el umbral perceptivo, L, con respecto a la sílaba acentuada, fuerte, H, el tono subyacente es (L + H)* + L. El paréntesis mantiene la bitonalidad subyacente, aunque el patrón tonal sea *hat pattern* 'bajo-alto-bajo'. La taxonomía se inspira, libremente, en Hualde (2002). El diacrítico estrella puede aparecer dentro o fuera del paréntesis. La alineación tonal se registra en el contexto a derecha y a izquierda o en la sílaba fuerte. El contexto es fuerte o débil (H o L) la sílaba acentuada es fuerte o débil (H o L). El resto de los valores tonales es transicional (Ladd, 1996: 42). Sólo se analizan los primitivos fonológicos, es decir, los niveles o estados estables tonales (Ladd, 1996: 43). Esta alineación tonal tienen en cuenta las diferencias fonológicas y las diferencias lingüísticas, esto es, comunicativas. El valor comunicativo está codificado en la producción y decodificado en la percepción de los tonos.

3. LA ALINEACIÓN TONAL EN LOS CORPUS

En la Fig. 1 se observa un segmento de la frase entonativa...*este radioteatro lo escuchaban por lo general las mucamas que quedaban en la casa trabajando...*, en el segmento se registra el acento tonal *mucamas*. El hablante, masculino, pertenece al dialecto de Buenos Aires. En este acento tonal, el contexto fonético débil, a derecha y a izquierda de la sílaba fuerte, está sobre el umbral de 1.50 semitono (st.) y es L. Es un caso de *hat pattern*. El paréntesis mantiene la bitonalidad del acento tonal, es (L+H*)+L. El valor tonal de la sílaba a la izquierda de la fuerte es 39.87 st. y la sílaba fuerte tiene un valor de 46.56 st., los dos valores están sobre el umbral perceptivo. Una relación similar se observa entre la sílaba fuerte y el contexto de la derecha, débil y L. El diacrítico estrella le pertenece a la sílaba fuerte y acentuada, H*.

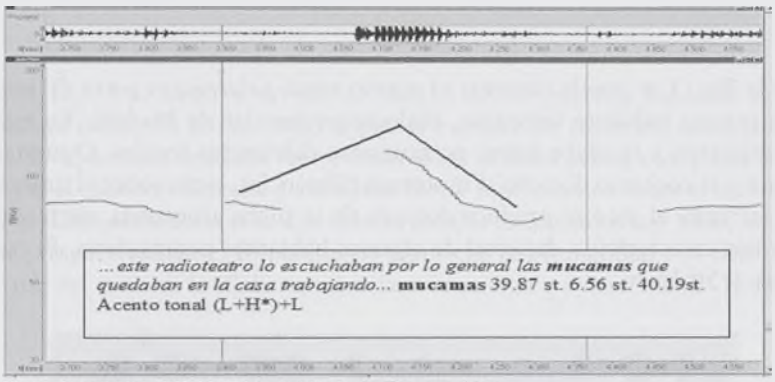


Fig. 1 Acento tonal (L+H*)+L, hablante de Buenos Aires.

La Fig.2 muestra un acento tonal obtenido en el discurso radiofónico de un hablante masculino, de Cuba. La frase entonativa es...*Israel posee suficientes misiles, vectores, aviones de combate y bombarderos...*, el acento tonal es *suficientes*. La relación entre el contexto fonético de la izquierda y la sílaba fuerte está sobre el umbral perceptivo, los valores son 54.89 st. y 60.25 st. En cambio, el contexto derecho y la sílaba fuerte no manifiesta diferencias tonales significativas, está bajo el umbral perceptivo. El tono es, entonces, L+H*.

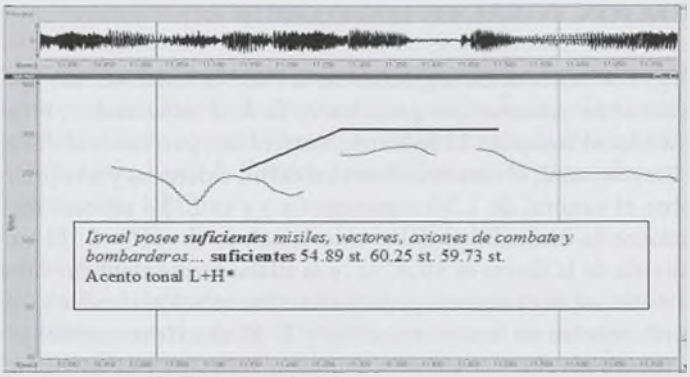


Fig. 2 Acento tonal L+H*, hablante de Cuba.

En la Fig. 3 se puede observar el acento tonal *palabras*, es parte de una muestra producida por un hablante femenino, dialecto peninsular de Madrid. En este caso, el contexto izquierdo y la sílaba fuerte no muestran diferencias tonales. Opuestamente, la sílaba fuerte y el contexto derecho sí muestran diferencias, están sobre el umbral perceptivo. En este caso, el pico se produce después de la sílaba acentuada, un fenómeno que podría ser una característica dialectal de algunos hablantes peninsulares. El contraste es de 53.99 st. y 56.16 st.

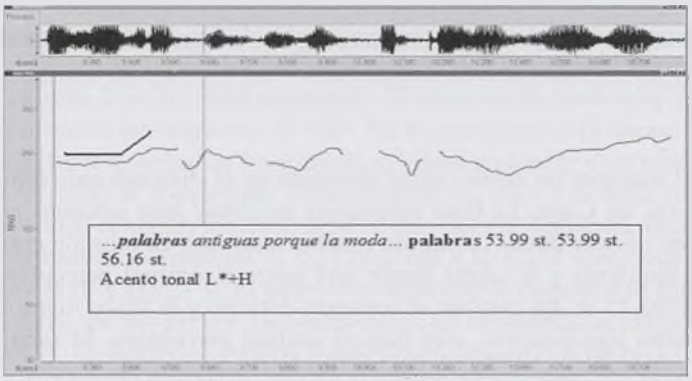


Fig. 3 Acento tonal L*+H, hablante de Madrid (CREA).

La Fig. 4 indica las características de alineación tonal del acento *lazo* emitido por un hablante masculino del área dialectal de Castilla. El contexto derecho y la sílaba fuerte están bajo el umbral perceptivo. Los valores registrados son 40.60 st. y 39.47 st., respectivamente. El acento tonal tiene el diacrítico estrella sobre la sílaba fuerte, H*.

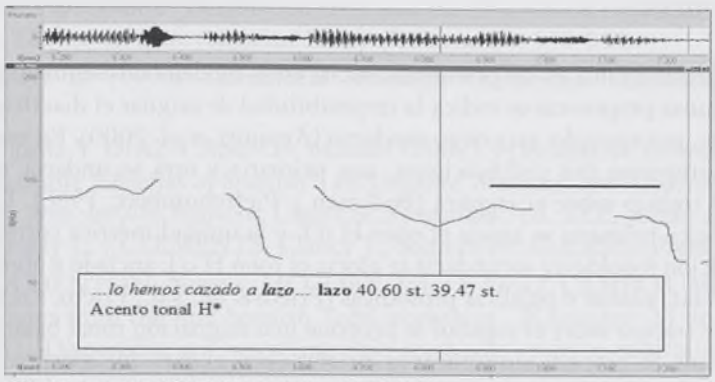


Fig. 4 Acento tonal H*, hablante de Castilla (DIES).

En la Fig. 5 se observa un acento tonal emitido por una hablante de Tenerife, una animadora radial. El acento tonal *despejado* muestra un recorrido tonal particular: el contexto a la izquierda y la sílaba fuerte están sobre umbral perceptivo y esta sílaba es H, mientras que el contexto derecho y esa sílaba fuerte está también sobre el umbral perceptivo, pero la sílaba acentuada es L. Los valores tonales escalonados son 54.07 st., 57.75 st. y 59.44 st., respectivamente. El paréntesis mantiene la bitonalidad y la barra oblicua la doble lectura del tono en la sílaba acentuada, a izquierda y a derecha: (L+H/L*)+H.

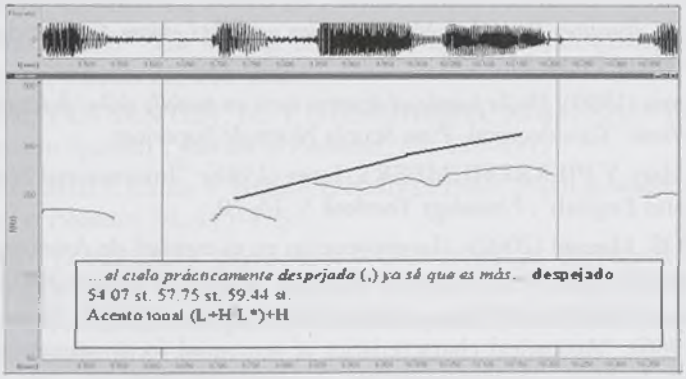


Fig. 5 Acento tonal (L+H/L*)+H, hablante de Tenerife.

4. CONCLUSIÓN

La alineación tonal es un problema crucial en el modelo autosegmental y métrico (AM). En algunas propuestas se indica la imposibilidad de asignar el diacrítico estrella a la sílaba fuerte, por ejemplo, en griego moderno (Arvaniti, *et al.* 2000). En varias lenguas romances se proponen dos codificaciones, una primaria y otra secundaria, este recurso basado en un trabajo sobre el japonés (Beckman y Pierrehumbert, 1986). En la codificación fonológica primaria se asocia el tono H o L y la unidad métrica correspondiente. En la codificación fonológica secundaria se asocia el tono H o L anclado a alguna frontera prosódica, moras, sílabas o palabras prosódicas (Prieto *et al.*, s.d.; Prieto, s.d.).

En este trabajo sobre el español se propone una asignación tonal basada en la percepción de tonos. Se propone simplificar la interfaz entre el dato de superficie, fonético, y la lectura fonológica. Se propone un recurso para la segmentación del continuo fonético en estados tonales estables. Luego se resuelve la lectura del contexto derecho e izquierdo de la sílaba fuerte según umbrales psicofonéticos (Pamies Bertrán *et al.*, 2002). Por último, se asigna el diacrítico estrella a la sílaba acentuada de acuerdo a la relación con el contexto, sobre el umbral perceptivo o bajo el umbral perceptivo. Lo hallazgos sugieren una taxonomía tonal que puede dar cuenta de las diferencias lingüísticas, es decir, tonos en contaste comunicativo (Toledo, 2000, 2001, 2002, 203, 2004a, 2004b, s.d.).

5. BIBLIOGRAFÍA

ARVANITI, Amalia, LADD, Robert D. y MENNEN, Ineke (2000): "What is a starred tone? Evidence from Greek", en BROE y PIERREHUMBERT (eds.) *Papers in Laboratory Phonology V: Acquisition and the Lexicon*. Cambridge: Cambridge University Press, 119-131.

AVESANI, Cinzia (1990): *Dalla parola al discorso: verso un modello della 'declinazione' intonativa in italiano*. Tesis doctoral. Pisa: Scuola Normale Superiore.

BECKMAN, Mary Y PIERREHUMBERT, Janet (1986): "Intonational Structure in Japanese and English", *Phonology Yearbook* 3, 15-70.

DIAZ-CAMPOS, Manuel (2002): "La entonación en el español de América: Un estudio acerca de ocho dialectos hispanoamericanos", *Boletín de Lingüística* 18, 3-26.

EADY, Stephen, COOPER, William, KLOUDA, Gayle, MELLER, Pamela y LOTTS, Dan (1986): "Acoustical characteristics of sentential focus: narrow vs. broad and single vs. dual focus environments", en: *Language and Speech* 29, 233-251.

FACE, Timothy (2000): "Prosodic manifestation of focus in Spanish", *Southwest Journal of Linguistics* 19(1), 45-62.

HUALDE, José Ignacio (2002): "Intonation in Spanish and the other Ibero-Romance languages: overview and status quaestionis", en WILTSHIRE, Caroline y CAMPS, J. (eds.) *Romance phonology and variation. Selected papers from the 30th Linguistic Symposium on Romance Languages*. Amsterdam: John Benjamins, 101-116.

LADD, Robert D. (1988): "Declination reset and the hierarchical organization of utterances", *The Journal of the Acoustical Society of America* 84, 530-544.

LADD, Robert D. (1996): *Intonational Phonology*. Cambridge: Cambridge University Press.

MCGORY, Julia Y DÍAZ-CAMPOS, Manuel (2002): "Declarative Intonation Patterns in Multiple Varieties of Spanish", en *Structure, Meaning, and Acquisition in Spanish: Papers from the 4th Hispanic Linguistic Symposium*,, ed. LEE, James F., GEESLIN, Kimberly L. y CLEMENTS, Clancy J., Somerville: Cascadilla Press, 73-92.

MENDOZA DENTON, Norma, MCGORY, Julia Y DÍAZ CAMPOS, Manuel (1999): "Minutes from the first Spanish ToBI workshop". [Consulta: 24 de setiembre de 2005]. Disponible en: www.ling.ohio-state.edu/~tobi/sp-tobi/minutes_formatted.html.

NOOTEBOOM, S. (1997): "The prosody of speech: melody and rhythm", en W. HARDCASTLE, W. y LAVER, J.(ed.) *The Handbook of Phonetic Sciences*, Oxford: Blackwell Publishers: 640-673.

PAMIES BERTRÁN, Antonio, FERNÁNDEZ, Ana, MARTÍNEZ CELDRÁN, Eugenio, ORTEGA, Alicia y AMORÓS, Mari Cruz (2002): "Umbrales tonales en español peninsular", en DÍAZ GARCÍA, Jesús (ed.) *Actas del II Congreso de Fonética Experimental*. Sevilla: Universidad de Sevilla, 272-278.

PRIETO, Pilar (s.d.): "En torno a la asociación tonal en el modelo métrico-autosegmental. Puntos controvertidos en su aplicación al catalán", *Revista Internacional de Lingüística Iberoamericana* 5, vol. 2 (Volumen especial sobre *Prosodic Aspects*, ed. KABATEK, Johannes).

PRIETO, Pilar (1998): "The scaling of L values in Spanish downstepping contours", *Journal of Phonetics* 26, 261-282.

PRIETO, Pilar, VAN SANTEN, Jan Y HIRSCHBERG, Julia (1995): "Tonal alignment patterns in Spanish". *Journal of Phonetics*, 23. 429-451.

PRIETO, Pilar, SHIH, Chilin Y NIBERT, Holly (1996): "Pitch downtrend in Spanish", *Journal of Phonetics* 24, 445-473.

PRIETO, Pilar, D'IMPERIO, Mariapaola y GILI FIVELA, Barbara (s.d.): "Pitch accent alignment in Romance: primary and secondary associations with metrical structure", *Language and Speech* (Volumen especial sobre *Variation in Intonation*, ed. WARREN, Paul).

TOLEDO, Guillermo (2000): "H en el español de Buenos Aires", *Langues et Linguistique* 26: 107-127.

TOLEDO, Guillermo (2001): "Taxonomía tonal en español", *Language Design* 3: 1-20.

TOLEDO, Guillermo (2002): "Acentos tonales en discursos", en DÍAZ GARCÍA, Jesús, (ed.) *Actas del II Congreso de Fonética Experimental*, Sevilla: Universidad de Sevilla, pp. 78-88.

TOLEDO, Guillermo (2003): "Modelo autosegmental y entonación: los corpus DIES-RTVP", *Estudios de Fonética Experimental* XII: 143-163.

TOLEDO, Guillermo (2004a): "Modelo autosegmental y entonación: una muestra del corpus CREA", *Revista de Filología* 22 (RFULL), 313-327.

TOLEDO, Guillermo (2004b): "Prominencia H*: una muestra de español de Cuba", *Estudios de Fonética Experimental* XIII, pp. 181-202.

TOLEDO, Guillermo (s.d.): "Modelo autosegmental y dialecto: el español de Tenerife", *Revista Internacional de Lingüística Iberoamericana* 5, vol. 2 (Volumen especial sobre *Prosodic Aspects*, ed. KABATEK, Johannes).

CHOQUE TONAL EN ESPAÑOL

GUILLERMO TOLEDO
Conicet, Argentina y Université Laval, Canadá

1. INTRODUCCIÓN

En la base teórica de la fonología métrica se indica que la representación mental de las prominencias del acento están estructuradas jerárquicamente y no en forma lineal. Una estructura métrica abstracta recobra las relaciones acentuales en la palabra o en la frase. El eje horizontal muestra la sucesión silábica. En el eje vertical se señalan los niveles abstractos de la asignación acentual. Un principio de eufonía se manifiesta por medio del juego alternante de golpes rítmicos fuertes y débiles en toda la estructura jerárquica del eje vertical. Este desarrollo prosódico impide la realización de largas secuencias de golpes débiles, la alternancia rítmica, los acento primario, secundario y terciario (ver resultados opuestos en español, un cierto isosilabismo y la prominencia primaria: Cedergren y Toledo, 1993; Toledo, 1999, 2000a, 2001a, 2002a, 2002b); asimismo, veda la presencia de golpes fuertes adyacentes, el choque acentual (Selkirk, 1984). En español se observan resultados dispares. Almeida (1993) estudia este efecto prosódico en español canario, según la técnica de la mímica de habla. El corpus está integrado por palabras del tipo *azul-claro* o *primer niño*. Los resultados muestran una tendencia alternante entre las sílabas tónicas. Se produce una reducción temporal de la primera sílaba acentuada con el fin de recuperar el efecto eufónico. Pamies Bertrán (1994) registra también efectos de alternancia entre las sílabas tónicas contiguas. En este caso los materiales lingüísticos son naturales y los informantes son peninsulares e hispanoamericanos. Los hallazgos obtenidos indican la debilidad de un acento del par contiguo, con el fin de restablecer el patrón alternante y, por consiguiente, la eufonía. Desde un punto de vista fonológico, propone una regla de neutralización, un archiacento (similar a un archifonema, pero suprasegmental), una situación fonológica parecida a los architonos en algunas lenguas tonales, por ejemplo: el chino. Toledo (1997) obtiene este patrón prosódico: el tono asciende en A1 y la duración decrece en A1; el tono desciende en A2 y la duración crece en el mismo acento. Las prominencias se compensan.

En suma, los resultados generales no muestran una tendencia a evitar el choque acentual en español, en cambio, indican un efecto de compensación en los dos rasgos fonéticos: el tono y la duración.

Prieto (2002a, 2002b, en prensa) estudia los efectos de coarticulación en choques acentuales en catalán. El análisis enfoca el problema en la asignación tonal y la reorganización temporal y en las diferencias de altura tonal de los acentos en conflicto. En el primer enfoque encuentra que el valle (L) y el pico (H) del primer acento (L1 y H1) ocurren antes con respecto al final de la sílaba; en cambio, el valle del segundo acento (L2) ocurre después con respecto al inicio de la segunda silaba. El pico del segundo acento (H2) permanece estable. Con respecto al segundo enfoque, la altura tonal, se indican estos hallazgos. L2 aumenta su frecuencia; opuestamente, L1, H1 y H2 no son afectados en su altura tonal. De ello se concluye que los puntos máximos de F0 son *targets* 'estados estables' que el hablante intenta realizar en toda condición prosódica, un resultado que confirma otro estudio previo (Prieto y Shih, 1995).

En este nuevo trabajo sobre el español de Buenos Aires y de Barcelona, se analiza el choque tonal en dos aspectos: uno, la asignación tonal propuesta en otros trabajos previos (Toledo, 2000b, 2001b, 2002c, 2003a, 2003b, 2004a, 2004b, en prensa; Toledo, este congreso) y dos, la prominencia tonal de los dos acentos tonales en conflicto. Además, se analiza la posibilidad de un foco ilocutivo en un corpus enfatizado. En este nivel se actualizan las intenciones del hablante con respecto a la importancia enfática de su mensaje (ver Toledo y Cedengren, 1993). Se analizan corpus de español de Buenos Aires y de español de Barcelona.

2. PROCEDIMIENTOS

2.1. Corpus

Se analizan dos corpus emitidos por tres hablantes de Buenos Aires (GF, IF, femeninos, GT, masculino) y un hablante de Barcelona (AF, femenino). Estos son los corpus estudiados:

Corpus 1

Amar Roma fue algo inevitable.
Buscar cosas es lo que hace siempre.
Cortar cables es su trabajo.
Dorar muebles es labor de ebanista.
Esperar siempre es su historia de amor.
Focalizar puntos en el espacio es la nueva técnica.
Ganar notas altas es la obsesión del estudiante.
Hacer buenas acciones es el deber del cristiano.
Igualar tantos es el juego.
Jaquear reinas es su deseo.

Corpus 2

El celofán verde se compra en París.
Envuelve los regalos con celofán verde de París.
El huracán negro azotó las costas de España.
Se aproxima un huracán negro peligroso.
El musulmán viejo camina por las calles.
Todos vieron al musulmán viejo caminando.
El volcán rojo desprende lava ardiente.
Se despierta el volcán rojo africano.
La sartén limpia evita las bacterias.
La receta indica una sartén limpia y aceitada.
El sacristán místico se volvió libertino.
Castigó al sacristán místico duramente.
El escocés rubio controla el modelo.
El modelo del escocés rubio es bueno.
El capitán sórdido llamó a la niña.
La niña llamó al capitán sordo y ciego.
Andrés Cabo se enojó con Amalia.
Amalia no contestó a Andrés Cabo rápidamente.
Inés Mora se fue a París.
Todos los amigos esperan a Inés Mora en el aeropuerto.

El primer corpus presenta los acentos tonales en conflicto en primera y segunda posición en las oraciones enunciativas (A*mar Ro*ma...). El segundo corpus presenta los acentos tonales en choque tonal en primera y segunda posición de la oración enunciativa (El celo*fán ver*de...) o en otras posiciones de esos enunciados y siempre en realizaciones antes del acento nuclear, esto es, en el pretonema, por ejemplo,...con celo*fán ver*de de París.

2.2. Análisis acústico y segmentación

Las mediciones se realizan en el *Speech Analyzer* 1.5 *Test Version* 15.3 (*Summer Institute of Linguistics*). El cálculo de la prominencia se efectúa en el pico de la fundamental de cada acento tonal y en el contexto fonético a izquierda y a derecha para marcar cambios de asignación tonal. Estos son los criterios en contornos conflictivos. En las sílabas en que el contorno es plano u ondulante se alinea el contorno de la fundamental con el contorno de intensidad. El punto de medición es el valor tonal más alto que se alinea con el valor más elevado en el contorno de intensidad (Ladd, 1988). Los tonos en las sílabas con oclusivas sordas se miden en el estado estable de la vocal y no en el valor más alto, en el valor de la sonoridad inmediata a la explosión (Eady et al, 1986).

2.3. Espacio tonal y umbrales tonales

Se segmenta el espacio tonal del hablante (el máximo valor al mínimo) en microespaciones tonales, en cada acento tonal. La segmentación se realiza de acuerdo a

umbrales psicofonéticos, esto es, psicoacústicos (las diferencias apenas perceptibles entre tonos contrastados). Dentro de cada microespacio tonal rigen los niveles estáticos alto (H) y bajo (L). Los umbrales tonales se determinan por medio de tests perceptivos obtenidos con técnicas de manipulación de la señal natural sometida a cambios tonales sintetizados. Los resultados de los sujetos de Barcelona y Granada indican la posibilidad de discriminación de diferencias tonales entre sílabas adyacentes con diferencias de 1.50 semitono (st.) (Pamies Bertrán *et al.*, 2002; Toledo, este congreso).

2.4. Taxonomía tonal entre acentos en choque

Los primitivos fonológicos en los acentos en choque son H* o L*, monotonales. El diacrítico estrella indica que el tono se produce dentro del desarrollo temporal de la sílaba acentuada. Las diferencias entre los dos acentos en choque se produce si uno de los acentos está sobre el umbral perceptivo, esto es, si la diferencia es mayor que 1.50 semitono. Si los acentos están bajo el umbral perceptivo, la prominencia es similar. Pueden producirse alteraciones en la asignación tonal, esto es, combinaciones bitonales debido a la presencia de otro acento en choque. Un efecto de *preshooting* 'prerrealización del pico' y el primitivo fonológico es H+L*. Asimismo, un efecto de *overshooting* 'posrealización del pico' y el primitivo fonológico resulta L*+H (Ver Avesani, 1990; ver también Toledo, este congreso).

3. RESULTADOS

En la Fig. 1 se observan los contornos de los tres primeros enunciados del Corpus 1. Son oraciones emitidas por un hablante masculino de Buenos Aires (GT). Los acentos en choque están en primera y en segunda posición, los dos están indicados por medio de los cursores. Las prominencias en choque tonal son similares, están bajo el umbral perceptivo de 1.50 semitono. Este hablante no rechaza el choque tonal, los acentos son H*.

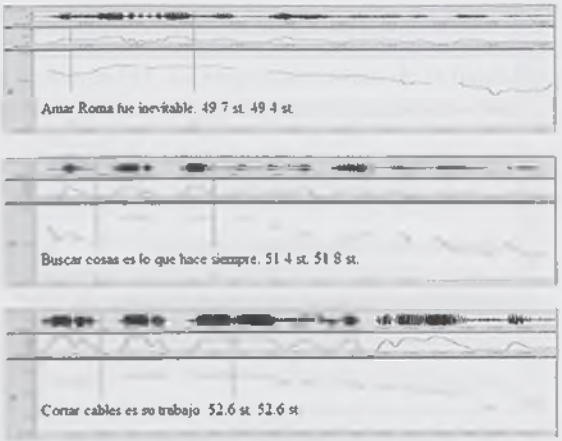


Fig. 1 Choque tonal en el español de Buenos Aires (GT)

En la Fig. 2 se indican los contornos de los enunciados 4, 5 y 6 del Corpus 1, emitidos por GT. Los resultados son similares a la Fig. 1. Este hablante produce acentos tonales en choque sin diferencias en la prominencia, todos los primitivos fonológicos son H*.

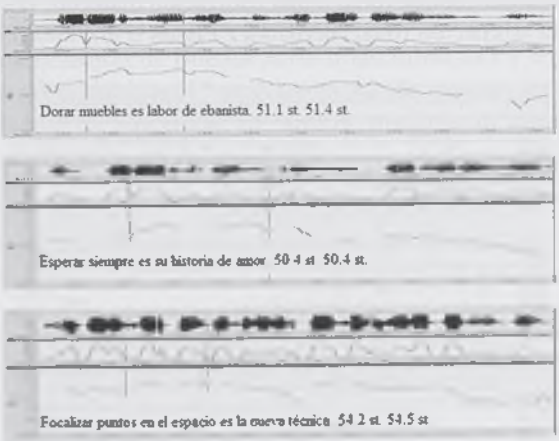


Fig. 2 Choque tonal en el español de Buenos Aires (GT).

En la Fig. 3 se consignan los contornos de los enunciados 7, 8 y 9 del Corpus 1, emitidos también por el mismo hablante, GT. La tendencia tonal es similar a los casos ya consignados: los dos acentos en choque no manifiestan cambios de prominencia, están bajo el umbral perceptivo y son H*.

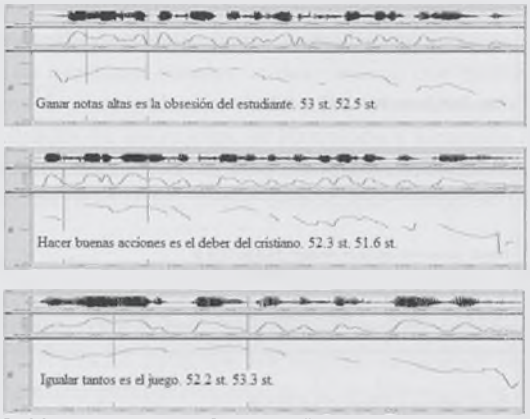


Fig. 3 Choque tonal en el español de Buenos Aires (GT)

En la Fig. 4 se muestran los contornos entonativos de los tres primeros enunciados del Corpus 1, el hablante femenino es peninsular, de Barcelona, es bilingüe. En los tres casos consignados se muestra que el acentos 2 se realiza con un fenómeno de posrealización del pico, el primitivo fonológico es L*+H. La barra vertical (negra) indica el desplazamiento de la prominencia a la sílaba inacentuada siguiente Este fenómeno de asignación tonal podría ser consecuencia del choque tonal o una realización dialectal. Los hablantes peninsulares propiciarían esta asignación tonal (Fernández Planas *et al.*, 2002).

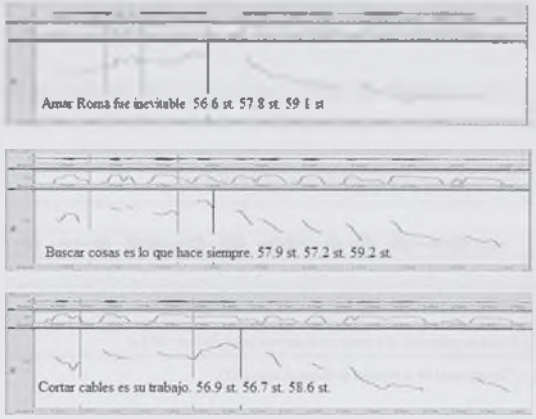


Fig. 4 Choque tonal en el español de Barcelona

En la Fig. 5 se observan los contornos de dos hablantes de español de Buenos Aires (femeninos, GF y IF) y una hablante de español peninsular (AF). Los enunciados pertenecen al Corpus 1. Las tendencia prosódicas son similares a las ya observadas. Las hablantes

de Buenos Aires presentan prominencias H* en los dos acentos en choque. Sin embargo, la asignación tonal de la hablante GF, en el A2, es H*+L; en cambio, la asignación tonal de la hablante IF, en el A2, es H*. La asignación tonal de la hablante peninsular (AF), en el A2, es L*+H, con un fenómeno claro de posrealización del pico.

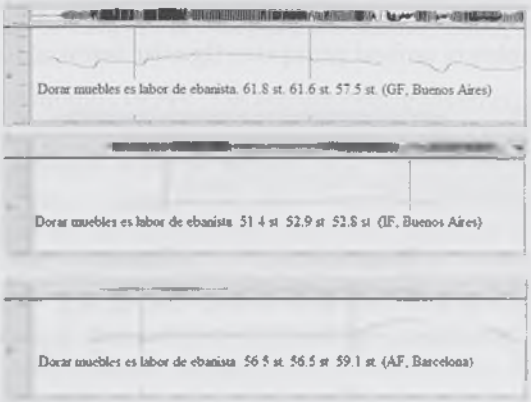


Fig. 5 Choque tonal en el español de Buenos Aires y de Barcelona

En la Fig. 6 se consignan los fragmentos de los contornos de entonación en un enunciado del Corpus 2, los dos acentos en choque tienen la posición 1 y 2. Se muestran dos contornos de dos hablantes de Buenos Aires (masculino, GT y femenino, IF). Los dos hablantes presentan prominencias H* en A1 y en A2. Se realizan los dos estados estables en plenitud. La hablante de Barcelona, en cambio, muestra un fenómeno de posrealización y el primitivo fonológico del A2 es L*+H.

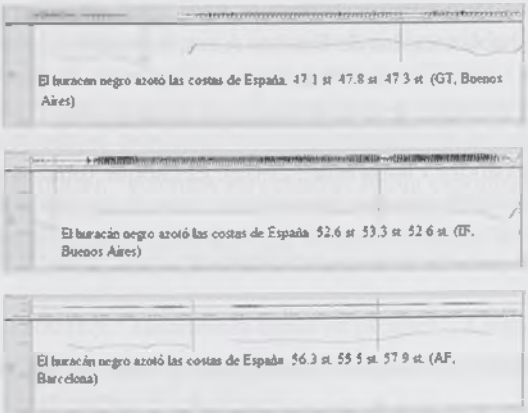


Fig. 6 Choque tonal en el español de Buenos Aires y de Barcelona

En la Fig. 7 se pueden observar algunos fragmentos de contornos de entonación en un Corpus 1 emitido con diferencias pragmáticas, esto es, ilocutivas. La hablante de Buenos Aires (GF) graba emisiones del corpus con enunciados enfáticos y otros neutros. En el fragmento superior y en el fragmento (posición media en la Figura), el A1 tiene prominencias más destacadas que el A2, los dos acentos están sobre el umbral perceptivo (A*mar* Roma... y Bus*car* cosas...). En el fragmento inferior es el A2 el que se destaca por énfasis; los dos están sobre el umbral perceptivo (Igualar *tan*tos...).

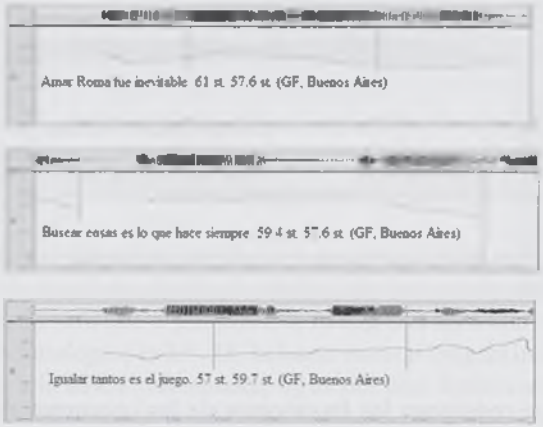


Fig. 7 Choque tonal en un corpus enfatizado, hablante de Buenos Aires

4. CONCLUSIÓN

Los hallazgos obtenidos en esta investigación confirman las diferentes estrategias en la asignación tonal entre los hablantes de Buenos Aires y los hablantes peninsulares. La tendencia general de esos hablantes, los de Buenos Aires, es mantener los valores de A1 y A2, sin rechazar el choque tonal. De manera opuesta, la hablante de Barcelona propugna una posrealización del pico en todas las emisiones, los primitivos son L*+H en el A2. Esta tendencia, sin embargo, podría ser dialectal y no un fenónemo de cambio de asignación tonal en casos de choque (Fernández Planas *et al.*, 2002). El choque tonal es sensible a diferencias pragmáticas debido al interés enfático del hablante: uno de los dos acentos se privilegia en prominencia para indicar el valor ilocutivo de ese ítem y de ese referente.

5. BIBLIOGRAFÍA

ALMEIDA, Manuel (1993): "Alternancia temporal y ritmo en español", *Verba* 20: 433-443.

AVESANI, Cinzia (1990): *Dalla parola al discorso: verso un modello della 'declinazione' intonativa in italiano*. Tesis doctoral. Pisa: Scuola Normale Superiore.

CEDERGREN, Henrietta y TOLEDO, Guillermo (1993): "Rhythm and compression in Caribbean Spanish", *The Journal of the Acoustical Society of America* 93 (2): 2297.

FERNÁNDEZ PLANAS, Ana María, MARTÍNEZ CELDRÁN, Eugenio, SALCIOLI GUIDI, Valeria, TOLEDO, Guillermo y CASTELLVÍ VIVES, Joan, "Taxonomía autosegmental en la entonación del español peninsular", en DÍAZ GARCÍA, Jesús (ed.) *Actas del II Congreso de Fonética Experimental*, 2002, Sevilla: Universidad de Sevilla, 180-186.

PAMIES BERTRÁN, Antonio (1994): "Los acentos contiguos en español", en *Estudios de Fonética Experimental* VI. Barcelona: Universidad de Barcelona, pp. 91-111.

PAMIES BERTRÁN, Antonio, FERNÁNDEZ, Ana, MARTÍNEZ CELDRÁN, Eugenio, ORTEGA, Alicia y AMORÓS, Mari Cruz (2002): "Umbrales tonales en español peninsular", en DÍAZ GARCÍA, Jesús (ed.) *Actas del II Congreso de Fonética Experimental*. Sevilla: Universidad de Sevilla, 272-278.

PRIETO, Pilar (2002a): "Efectos de coarticulación en choques acentuales", seminario invitado al Coloquio *Tonía, dimensiones fonéticas y fonológicas*, México, El Colegio de México.

PRIETO, Pilar (2000b): "Coarticulation and stability effects in tonal clash contexts in Catalan", Proceedings of the Speech Prosody 2002 Conference, ed. BEL, Bernard y MARTIEN, Isabelle, Aix-en-Provence: Laboratoire Parole et Langage, 587-590.

PRIETO, Pilar (s.d.): "Effects in tonal clash cotexts in Catalan", *Jouunal of Phonetics*.

PRIETO, Pilar y SHIH, Chilin (1995): "Effects of tonal clash on downstepped H* accents in Spanish", *Proceedings of EUROSPEECH '95, Fourth European Conference on Speech Communication and Technology*, Vol. 2, 1307-1310.

SELKIRK, Elizabeh (1984): *Phonology and Syntax. The Relation between Sound and Structure*. Cambridge, Massachusetts y Londres: The MIT Press.

TOLEDO, Guillermo (1997): "Prominencia melódica y temporal: La colisión acentual en el español", en *Estudios de Fonética Experimental*. Barcelona: Universidad de Barcelona.

TOLEDO, Guillermo (1999): "Jerarquías prosódicas en el español", *Revista Española de Lingüística* 29: 69-104.

TOLEDO, Guillermo (2000a): "Acentos en español: habla espontánea", *Estudios Filológicos* 35: 125-137.

TOLEDO, Guillermo (2000b): "H en el español de Buenos Aires", *Langues et Linguistique* 26: 107-127.

TOLEDO, Guillermo (2001b): "Taxonomía tonal en español", *Language Design* 3: 1-20.

TOLEDO, Guillermo (2001a): "Acentos en el español: un corpus de conversación", *Estudios de Fonética Experimental* XI: 121-142.

TOLEDO, Guillermo (2002a): "Reglas del acento en paroxítonos: el español peninsular", *Estudios Filológicos* 37: 133-149.

TOLEDO, Guillermo (2002b): "Acentos en español: un problema para la fonología métrica", *Verba* 29: 119-138.

TOLEDO, Guillermo (2002c): "Acentos tonales en discursos", en DÍAZ GARCÍA, Jesús (ed.) *Actas del II Congreso de Fonética Experimental*, Sevilla: Universidad de Sevilla, pp. 78-88.

TOLEDO, Guillermo (2003a): "Modelo autosegmental y entonación: los corpus DIES-RTVP", *Estudios de Fonética Experimental* XII: 143-163.

TOLEDO, Guillermo (2003b): "Modelo autosegmental y entonación: el español de Cuba", en *Lingüística 2003, Conferencia Científica Internacional, Instituto de Literatura y Lingüística «José Antonio Portuondo Valdor», La Habana, Cuba del 24 al 29 de noviembre de 2003*, CD-ROM ISBN: 959-7152-04-5.

TOLEDO, Guillermo (2004a): "Modelo autosegmental y entonación: una muestra del corpus CREA", *Revista de Filología* 22 (RFULL), 313-327.

TOLEDO, Guillermo (2004b): "Prominencia H*: una muestra de español de Cuba", *Estudios de Fonética Experimental* XIII, pp. 181-202.

TOLEDO, Guillermo (en prensa): "Modelo autosegmental y dialecto: el español de Tenerife", *Revista Internacional de Lingüística Iberoamericana* (RiLi), Frankfurt/Madrid.

TOLEDO, Guillermo (este congreso): "Alineación tonal en español".

TOLEDO, Guillermo y CEDERGREN, Henrietta (1993): "Focus in Caribbean Spanish", *Learned Societies Conference. Linguistics*, Ottawa, Canadá.

ANÁLISIS ACÚSTICO DE LAS VOCALES DEL PORTUGUÉS EUROPEO PRODUCIDAS POR HISPANOHABLANTES

Ana Margarida C. Vaz Da Silva, Rosa Lídia Coimbra,
Antonio J. S. Teixeira, Lurdes Castro Moutinho
Universidad de Aveiro (Portugal)

INTRODUCCIÓN

La proximidad geográfica de Portugal y España acerca sus dos lenguas nacionales. Sin embargo, la historia marcó caminos divergentes para ambos países, dejando marcas lingüísticas que hicieron posible que hoy nos encontremos ante dos lenguas: el portugués y el español[1]. Si la proximidad entre dos lenguas puede, en un determinado momento de la adquisición de una nueva lengua (L2), facilitar el proceso de su aprendizaje, también es sabido que, sobre todo en una fase inicial, el aprendiz usa casi de forma inconsciente como estrategia de comunicación el recurso de competencia hacia su lengua materna (L1).

En lo que se refiere a las dos lenguas que aquí nos ocupan -portugués y español- las semejanzas son inequívocas, pero también es cierto que sus diferencias se manifiestan a diferentes niveles de la estructura lingüística, no constituyendo una excepción lo relacionado con los dominios fonético y fonológico.

La falta de estudios realizados sobre la importancia y la utilidad de la fonética en la enseñanza y aprendizaje de las lenguas extranjeras, en especial en la enseñanza del portugués a hablantes de español, es corroborada por la escasa bibliografía encontrada en este dominio. Si la proximidad de las dos lenguas puede en cierta medida ser facilitadora en la adquisición de L2, también puede dificultar el progreso de su aprendizaje. La ausencia de una discriminación correcta de los sonidos de L2 puede conducir a errores de pronun-

1 Se opta por esta designación, de acuerdo con la indicación de Antonio Llorente, que considera el castellano como variedad regional de la lengua española en la comunidad de Castilla y León. También la Real Academia Española, tras su formación, optó por referirse a la Lengua Española, considerando el Castellano como un estadio de la evolución hacia el Español actual.

ciación que sólo podrán ser corregidos si se realiza esa reflexión/discriminación previa, a través de la comparación de los dos sistemas lingüísticos en cuestión[2].

A pesar de existir varios aspectos fonéticos relevantes, optamos por el estudio de las vocales orales[3] por conllevar dificultades añadidas en la adquisición del portugués, por la gran discrepancia que se verifica entre el número de vocales en cada una de las lenguas, como fácilmente se deduce por la observación de los triángulos acústicos del portugués (Martins, 1998) y del español (Celdrán, 1994: 188).

En verdad, y sabiendo que el sistema vocálico oral de la lengua portuguesa presenta una diversidad de fonemas superior -nueve vocales- a la del sistema vocálico del español, que cuenta apenas con cinco, es nuestra intención verificar qué tipo de estrategias son usadas cuando un hispanohablante es conminado a producir sonidos vocálicos que no se encuentran contemplados en su lengua.

Con ese objetivo, fueron analizados diversos *corpora* previamente grabados. Posteriormente se interpretaron sus resultados para posibilitar la descripción de las diferentes realizaciones fonéticas de las vocales del portugués producidas por hispanohablantes.

1. LOS *CORPORA*

Constitución de los corpora

Los *corpora* contemplaron producciones en portugués y español por hablantes nativos complementadas por producciones del portugués como lengua extranjera por parte de hispanohablantes. Surgen, de este modo, tres *corpora*, con las siguientes características:

- *Corpus* 1: dieciocho palabras del portugués (ver cuadro 1), donde están contempladas todas las vocales orales del portugués, en posición acentuada e inacentuada, estando siempre que fuera posible entre consonantes oclusivas sordas.
- *Corpus* 2: diez palabras en español, cubriendo todos los contextos en que pudieran aparecer las vocales de esta lengua (ver cuadro 2).
- *Corpus* 3: treinta palabras, también portuguesas (ver cuadro 3), grabadas en el ámbito de otro proyecto[4] y reutilizadas en este trabajo, sirviendo como complemento del *corpus* 1.

---

2 Este aspecto es muchas veces descuidado por el profesor, incluso porque la mayoría de las veces los métodos destinados a la enseñanza de lenguas extranjeras, en particular los destinados a la enseñanza del portugués, casi no se refieren a estos aspectos, como es fácil de constatar a través de una consulta de los materiales existentes. Se hace, si es que se hace, una breve referencia a aspectos fonéticos, para centrarse después sobre aspectos morfológicos y sintácticos de L2.

3 Aunque el estudio de la adquisición de las vocales nasales sea un estudio interesante, por el hecho de que encontrarse completamente ausentes en el español, no habrá cabida para ellas en el ámbito del presente artículo.

4 Este *corpus* fue grabado en el ámbito del Proyecto "Fonética Aplicada al Procesamiento del Habla: Las Nasales del Portugués" (proyecto financiado por el programa POSI, Fundación para la Ciencia y la Tecnología, ya concluido). Más información disponible en http://www.ii.ua.pt/cidlc/gcl/fap/.

Corpus de palabras del portugués Informante hispanohablante	
Fonema (SAMPA)[5]	Contexto -entre oclusivas sordas
[a]	sapato, macaco
[6]	atacar, incapaz
[E]	raquete, pateta
[e]	capete, chupeta
[@]	antepor, piquetar
[i]	apito, mesquita
[o]	bacoco, batoco
[O]	pacote, pipocas
[u]	astuto, batuta

Cuadro 1 - *Corpus* 1

Corpus de palabras del español Informante hispanohablante	
Fonema (SAMPA)	Contexto -entre oclusivas sordas
[a]	ataque, zapato
[e]	capete, cateto
[i]	pupitre, repique
[o]	capote, estopa
[u]	batuta, disputa

Cuadro 2 - *Corpus* 2

Corpus de palabras do portugués Informante portugués	
Fonema (SAMPA)	Contexto
[a]	pato, capa, baba
[6]	pacote, capete,capacete gaguejar
[E]	peca, pateta
[e]	capete, capacete ampulheta
[@]	degrau
[i]	tipo, tito, guida
[o]	pintor, conquistador tambor, mentiroso connosco, totoloto
[O]	hipopótamo, pote totoloto, copo, bigode
[u]	cupido, buda, narigudo

Cuadro 3 - *Corpus* 3

Los informantes

En términos generales, los informantes tenían edades comprendidas entre los 25 y los 35 años, todos de sexo masculino, alumnos de la enseñanza preuniversitaria o universitaria.

5 El alfabeto SAMPA (*Speech Assessment Methods Phonetic Alphabet*) es un alfabeto fonético susceptible de lectura por el ordenador sin necesidad de instalación de fuente propia, ya que se sirve de los caracteres disponibles en el teclado. Fue

Tanto los informantes del *corpus* 1 como los del *corpus* 2 tienen como L1 el español y pertenecían, en el momento de las grabaciones, al nivel avanzado del *Curso de Português para Extranjeros* de la Universidad de Aveiro.

Los datos que constituyen el *corpus* 3, obtenidos en el decurso del Proyecto referido, fueron producidos por un informante portugués de la región de Aveiro.

Pretendemos, en el caso del *corpus* 1, obtener producciones orales del portugués por hispanohablantes. Los datos del *corpus* 3, producido por el informante portugués, permitirán un análisis comparativo con los datos obtenidos para el *corpus* 1. Los datos del *corpus* 2, producido por los mismos informantes del *corpus 1*, pero esta vez conteniendo palabras del español, serán comparados con los datos de los *corpora* 1 y 3.

El análisis de cada uno de los *corpora* y el cruce de datos entre sí deberá permitirnos determinar la distribución de las vocales del portugués producidas por hispanohablantes, para definir las características de esos hablantes cuando realizan las vocales portuguesas.

Recogida de los corpora

Las grabaciones fueron todas realizadas en el laboratorio de Fonética de la Universidad de Aveiro, en una sala no insonorizada, habiendo sido utilizada en el caso del *corpus*3 la recogida directa en un ordenador portátil usando el sistema CSL 4400 de Kay Elemetrics Corp y un micrófono Shure SM48. Durante la grabación del *corpus*, fue recogida, además de la señal de voz, información acerca del comportamiento de las cuerdas vocales a través de la utilización del electroglotógrafo (EGG). Los otros dos *corpora* fueron grabados directamente en un ordenador personal, grabándose apenas la señal acústica. Para los tres *corpora*, el programa utilizado para grabación y análisis acústico de la señal fue el programa SFS (Speech Filing System)[6].

Durante la grabación del *corpus* 3 fueron presentadas al informante imágenes para inducirlo a la producción de la palabra pretendida, grabando dos repeticiones de la misma. En el *corpus* 1 y 2 la recogida fue efectuada en base a la lectura de listas de palabras que se repitieron cinco veces.

Procesamiento posterior y anotación de los corpora

El programa utilizado para la grabación y segmentación del *corpus*, el SFSWin arriba mencionado permite visualizar todas las formas de onda en simultáneo o individualmente (ver figura 1): oscilograma -variación de la amplitud de la onda sonora a lo largo del tiempo- y el espectograma -representación visual de la resolución de la onda sonora en términos de componentes de frecuencia (medida en hertz[7]) a lo largo del tiempo (unida-

---

desarrollado a finales de los años 90 en el ámbito de un proyecto de un grupo de foneticistas y comenzó a ser aplicado a las lenguas de la Comunidad Europea, teniendo cada una de ellas una tabla propia. Tablas e información disponibles en http://www.phon.ucl.ac.uk/home/sampa/home.htm.

6 Este programa puede ser obtenido gratuitamente en http://www.phon.ucl.ac.uk/resource/sfs/

7 El Hertz (Hz) es una unidad de medida de frecuencia, numéricamente equivalente a ciclos por segundo (100 Hz = 100 ciclos por segundo). Un sonido es tanto más agudo cuanto mayor sea su frecuencia.

des de milisegundos). El SFSWin permite también operar sobre determinados segmentos seleccionados, oír, borrar, anotar, etc. El programa posee una herramienta que permite la visualización de las formas de onda de la señal acústica, en tiempo real.

Tras la grabación se procedió a la organización de los ficheros. Fueron constituidos directorios por informante, conteniendo las varias repeticiones efectuadas.

La anotación de los ficheros contempló dos niveles: la delimitación de la palabra completa y la segmentación de la vocal para analizar, así como del contexto consonántico que la antecede y la procede. En este proceso fue utilizado el alfabeto SAMPA para el portugués y para el español.

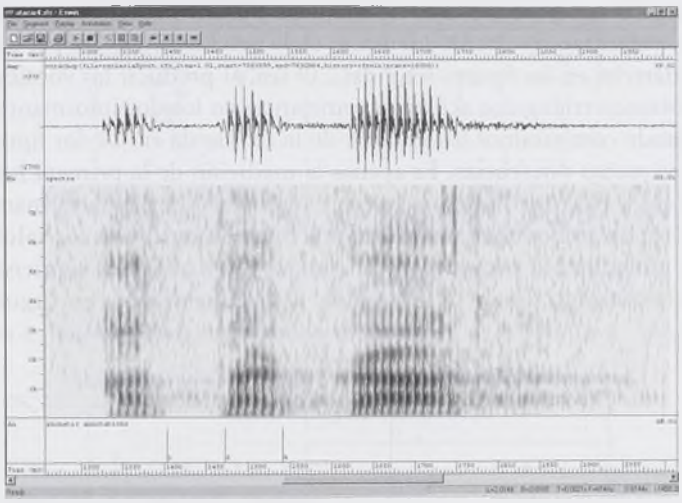


Figura 1 - Ejemplo de un fichero anotado (palabra "atacar" del *corpus* 1)

Para facilitar el análisis, los *corpora* fueron convertidos para su utilización por el sistema EMU[8]. En esta conversión fueron también calculadas las formantes.

Tras la conversión, se aprovechó para efectuar una verificación de la anotación inicialmente efectuada en SFS. Las diferentes facilidades de los dos sistemas, cuando son combinados, permitieron una mayor precisión de la anotación

8 EMU es una colección de herramientas informáticas para la creación, manipulación y análisis de bases de datos de habla, permiciendo búsquedas basadas en la estructura secuencial y jerárquica de las anotaciones. Puede ser obtenido gratuitamente en http://sourceforge.net/projects/emu.

2. RESULTADOS DEL ANÁLISIS

Después de segmentada y anotada la señal, se siguió a la fase de medición. Para tal, fueron tomadas las medidas de la primera y segunda formantes (F1 y F2) en cada una de las vocales. Con el fin de obtener resultados fiables en cada una de las vocales, la medida fue retirada en la mitad de su duración, por considerarse la parte más estable en el momento de la realización.

Los gráficos presentados en las figuras 2, 3 y 4 muestran los valores obtenidos en cada una de las varias realizaciones de los sonidos de los diferentes *corpora* (parte inferior de las figuras) y de las medias para cada uno de los fonemas (parte superior de las figuras).

Comparando la figura 2 con la figura 3, verificamos que los dos informantes hispanohablantes presentan resultados idénticos en la producción del *corpus* 2 (comparar los gráficos de la derecha en las figuras referidas). O sea, al producir las vocales de su lengua materna, se obtienen triángulos acústicos semejantes en los dos informantes.

Así, cuando comparamos los gráficos de la izquierda en las dos figuras, encontramos tan sólo pequeñas diferencias. Es apenas la medición de la primera informante (F1) de las vocales [a] y [6] que difiere un poco, y también la segunda formante (F2) de los sonidos [u] y [o]. En ambos casos se verifica una ligera inversión en los valores obtenidos. Sin embargo, globalmente podemos decir que los dos hablantes siguieron las mismas estrategias en la construcción de su versión del triángulo acústico portugués.

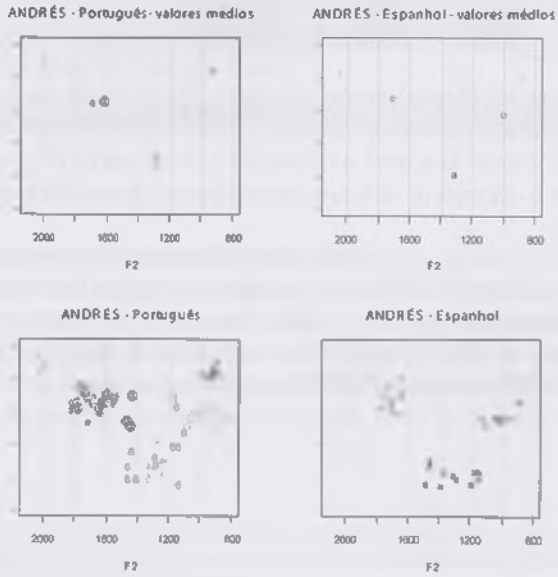


Figura 2 - Valores medios y realizaciones de los sonidos en el *corpus* 1 y 2 – informante 1

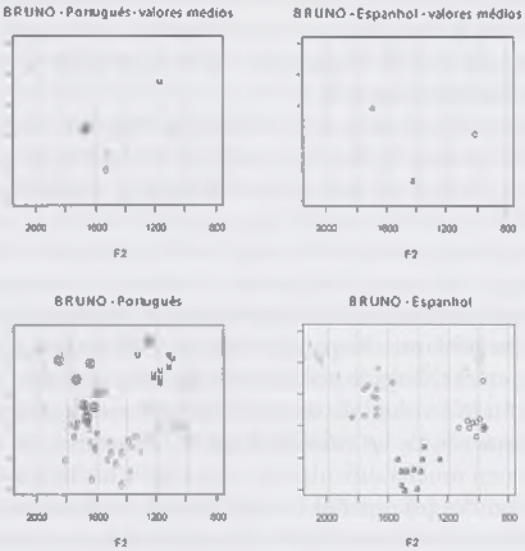


Figura 3 - Valores medios y realizaciones de los sonidos en el *corpus* 1 y 2 – informante 2

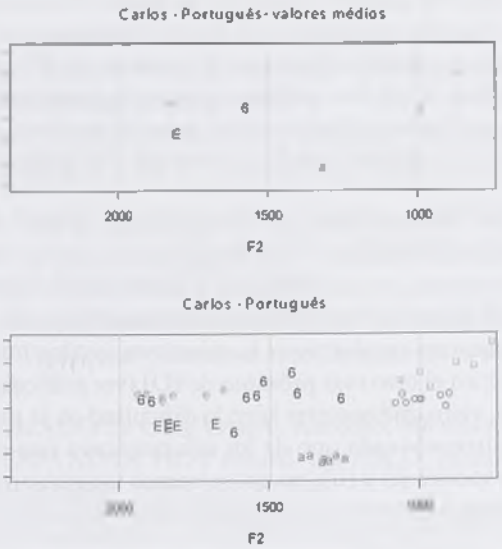


Figura 4 - Valores medios y realizaciones de los sonidos en el *corpus* 3

Partiendo de estos resultados, comparamos los triángulos acústicos obtenidos por la producción de palabras portuguesas por el informante portugués (figura 4) y por los informantes españoles (gráficos de la izquierda en las figuras 3 y 4) y verificamos algunas discrepancias, que pasaremos a exponer.

En primer lugar, verificamos que el hablante portugués, a pesar de haber grabado la palabra "degrau" (ver cuadro 4), donde se podría eventualmente producir la vocal [@] en la primera sílaba, no produjo ese sonido, que fue total y sistemáticamente elegido por él. El análisis de la señal acústica demostró que la producción de la palabra fue [dgraw] y no [d@graw]. Esta es, además, una elisión muy común en nuestra lengua[9].

En segundo lugar, nos parece importante destacar las realizaciones de [E], [e] y [@]. El hablante portugués elide la [@], como vimos, y presenta para [E] y [e] valores de F1 bien distintos. Ya los hablantes hispanohablantes colocan estas tres vocales en valores medios muy próximos en el triángulo acústico (ver gráfico superior izquierdo de las figuras 2 y 3) y la realización individual de estos sonidos es bastante inconsistente y dispersa (ver gráfico inferior izquierdo de las mismas figuras), mostrando de un modo muy claro que estos hablantes tienen mucha dificultad en distinguir los tres sonidos entre sí. Comparado con la realización del [e] español (ver gráficos de la derecha en las figuras 2 e 3), verificamos que es en esa misma zona donde ellos colocan las tres vocales portuguesas.

Una tercera observación se refiere a la producción de las vocales [a] y [6]. Mientras que el hablante portugués presenta una colocación media bien distinta de estas vocales en el triángulo acústico (ver gráfico superior de la figura 4), principalmente en términos de F1 y en las realizaciones individuales de estos sonidos, nunca presenta sobre posiciones entre las dos vocales (ver gráfico inferior de la figura 4); esto mismo no sucede con los hispanohablantes. Tan sólo se registra una ligera diferencia en el F1 entre [a] y [6] y, aún así sin consistencia, ya que el informante 1 presenta un F1 medio superior para la [6] y el informante 2 para la [a] (ver gráfico superior izquierdo de las figuras 2 y 3). Al mismo tiempo, en las realizaciones individuales de estos sonidos, tanto uno como el otro acaban por concentrar en la misma zona las dos vocales (ver gráfico inferior izquierdo en las figuras referidas).

En cuarto lugar, debemos destacar la realización de las vocales [O] y [o], que también presentan algunos problemas. Los hispanohablantes no fueron consistentes en la producción de estas vocales, en su distinción y colocación en oposición a la vocal [u]. Podemos, por ejemplo, observar (ver gráficos de la derecha en las figuras 2 y 3) que la posición relativa de estas tres vocales no es la misma en los dos informantes: en un caso, [o] está próximo de [u]; en el otro más próximo de [O] (ver gráficos inferiores). Esta constatación es interesante, visto que muestra bien la dificultad en la producción de esta oposición [O]/[o]/[u], recurriendo cada uno de los informantes a una estrategia diferente.

En cuanto a las vocales [i] y [u], no encontramos ningún problema, ya que los tres informantes las colocan igualmente en los valores máximos de F1 y las distinguen en F2

9 Tuvimos ocasión de comprobarlo en el decurso de los análisis efectuados con 24 informantes en el ámbito del proyecto referido en la nota 5.

respectivamente por los valores máximo y mínimo. Como excepción documentamos la alteración en F2 del segundo informante que, caso de no ser atribuible a error de medida (hipótesis que no podemos poner totalmente de lado debido a la corta duración de estos sonidos para estos informantes y condiciones de grabación) podrá ser consecuencia de la estrategia utilizada para la distinción [o]/[O], que pasa por la producción de [o]. A pesar de eso, no será en principio previsible que surjan dificultades en la producción de [i] y [u], en el PE, por parte de los hispanohablantes.

3. CONCLUSIÓN

A partir del breve análisis realizado, pudimos concluir que los sonidos vocálicos orales del portugués europeo (PE) no presentan todos las mismas dificultades, para una correcta realización, por parte de los hablantes hispanohablantes.

En esta búsqueda, dimos cuenta de las vocales que realmente presentan problemas en la producción de PE por parte de estos hablantes e identificamos los parámetros acústicos que denotan estas dificultades.

Tal materia merecería ser objeto de un análisis más profundo, para determinar con mayor rigor las diferencias en los sistemas vocálicos de las dos lenguas. Sólo esta toma de conciencia permitirá crear los automatismos necesarios en situación de confrontación lingüística, principalmente en la enseñanza/aprendizaje de lenguas y traducción simultánea.

BIBLIOGRAFÍA

ALARCOS LLORACH, Emilio: *Fonología Española*. Col. Biblioteca Románica Hispánica, Editorial Gredos S.A., Madrid, 1992.

ALCINA, J; BLECUA, J. Manuel: *Fonética y Fonología*, GRAMÁTICA ESPAÑOLA, Letras e Ideas, Editorial Ariel S. A., Barcelona, 1988, pp.195-483.

CAMARGO, Z. "Da Fonação à Articulação - Princípios Fisiológicos e Acústicos". *Fonoaudiologia*. Ano 2, N.º 2, 1999. (também disponível on-line em: http://www.fonoaudiologia.org.br/revista/f0799-2.htm)

CELDRÁN, E, Martínez. *Fonética*. Barcelona: Editorial Teide, 1994.

LADEFOGED, Peter. *Vowels and Consonants: An Introduction to the Sounds of Languages*. Blackwell, 2000.

LLORENTE MALDONADO de GUEVARA, Antonio, *Variedades del español en España*, LA LENGUA ESPAÑOLA HOY (coord. por SECO, Manuel y Salvador), Fundación Inca-March, Madrid, 1995, pp. 87-94.

MARTÍNEZ CELDRÁN, Eugenio: "En torno de las vocales del Español: análisis y reconocimiento", *Estudios de Fonética Experimental*, Vol. VII, Promociones y Publicaciones Universitarias S.A., Barcelona, 1995, pp.197-218.

MARTINS, Mª Raquel Delgado: *Ouvir falar: introdução à fonética do Português*, Editorial Caminho, Lisboa, 1998.

MATEUS, Mª Helena Mira: *Aspectos da Fonologia do Português*, Centro de Estudos Filológicos, Instituto de Alta Cultura, Lisboa, 1977.

MOUTINHO, Lurdes de Castro. *Uma Introdução ao Estudo da Fonética e Fonologia do Português*, Colecção Plátano Universitária, Plátano – Edições Técnicas, Lisboa, 2000.

QUILIS, Antonio; FERNANADEZ, Joseph A.: *Curso de fonética y fonología* españolas, Consejo Superior de Investigaciones Científicas, Madrid, 1992.

TEIXEIRA, António et al..:"An Acoustic Corpus Contemplating Regional Variation for Studies of European Portuguese Nasals", *4th International Conference On Language Resources And Evaluation (LREC)*, Lisboa, 2004.

VAZ DA SILVA, Ana M. C.; et al.: "Quanto mais pior: considerações acerca da aquisição das vogais orais do Português Europeu por hispanófonos", *Cadernos de PLE* 3, Universidade de Aveiro, Aveiro, 2003.

AGRADECIMIENTOS

Los autores agradecen a Guillermo Rodríguez Vilar la traducción de este artículo del portugués al español.

EL "TROUGH E EFECT": EL USO DEL ULTRASONIDO EN EL ESTUDIO DEL MOVIMIENTO LINGUAL DURANTE LA OCLUSIÓN EN LAS OCLUSIVAS BILABIALES

YOLANDA VÁZQUEZ ÁLVAREZ, NIGEL HEWLETT
Uva_es@yahoo.com, NHewlett@QMUC.ac.uk
Speech & Hearing Sciences / Speech Science Research Centre (SSRC), QMUC. Edimburgo (U.K.)

INTRODUCCION

La producción de las oclusivas bilabiales no implica el movimiento de la lengua que sería necesario para crear una constricción en la mayoría de las consonantes. El gesto articulatorio mas importante para la producción de esta clase de oclusivas es la oclusión de los labios por lo que la lengua no tendría necesidad de moverse durante la posición intervocálica. Sin embargo, Houde (1968) sería el primero en descubrir por medio de grabaciones cineradiográficas lo que mas tarde pasaría a ser llamado el 'trough effect'. Este estudioso encontró "perturbaciones linguales relacionadas con la fase de oclusión" en forma de movimento descendiente de la lengua en las consonantes oclusivas en un contexto vocal simétrico. Otros estudios (Bell-Berti & Harris 1974; Gay 1975) que utilizaron la técnica electromiográfica (EMG) también encontraron una notable desactivación en la actividad del genioglossus. Utilizando EPG, McAllister y Engstrand (1992) informaron de reducciones importantes en el número de contactos en el EPG lo que sugiere un movimiento descendiente de la lengua alejándose del paladar duro. Recientemente, Lindblom et al. (2002) utilizaron el formante F2 y datos de rayos X de un informante para afirmar que este fenómeno demuestra una activación que es segmento-por-segmento, en oposición al modelo de Öhman por el cual la trayectoria de la lengua es vocal-por-vocal con un gesto independiente de la lengua para la consonante. Una explicación vinculada al control motor favorecería una explicación donde los articuladores son activados dependiendo de los objetivos a alcanzar con una base segmental de señales neuronales (Lindblom et al. 2002). Por otro lado, existe una posible explicación aerodinámica rechazada por Lindblom et al. pero respaldada por Svirsky et al. (1997) y Fuchs et al. (2004) por la cual este movimiento lingual se podría explicar por la necesidad de la lengua de acomodar la

presión creada en la cavidad oral y que es necesaria para producir una explosión adecuada para la consonante oclusiva bilabial. Hoole et al. (1998) sugiere que este movimiento lingual anticiparía las fuerzas aerodinámicas y planearía su moviemiento para aprovecharse de la dirección y magnitud de estas fuerzas. A pesar de las diferentes teorías sobre la naturaleza del trough effect, hay pruebas fiables que demuestran la existencia de este fenomeno. La cuantificación de este 'trough' o 'movimiento descendiente de la lengua' es todavía una incógnita. Los resultados de anteriores estudios podrían haber sido engañosos debido a la gran variación entre los informantes. Esta variación pondría en duda estos resultados ya que cuando era posible visualizar el contorno de la lengua sólo se estudiaban uno o dos informantes. Pero cuando más informantes eran incluidos en el estudio, los investigadores estimaban la posición de la lengua y su movimiento a través de medicciones indirectas utilizando técnicas como el análisis acústico (frecuencias de los formantes), EPG (contacto de la lengua y el paladar) y EMG (medida de la actividad muscular).

En este estudio hemos querido investigar la variación de los informantes con respecto a este fenómeno utilizando mediciones del movimiento lingual en tres vocales diferentes y dos consonantes oclusivas labiales producidas por diez informantes diferentes.

DISEÑO EXPERIMENTAL

Informantes y materials

Diez informantes participaron en este experimento. Todos ellos eran hablantes nativos de uno de los acentos del Inglés Británico y estudiantes de logopedia. Los materiales utilizados fueron secuencias simetricas del tipo VCV donde la consonante (C) era /b/ o /p/ y la vocal (V) el monoptongo /i/, /u/ o /a/. Cada secuencia se situó en el interior de una frase fija "I said ____ too". Un total de 600 (30 respuestas: 6 VCVs – 5 repeticiones x 2 sesiones x 10 informantes) fueron recogidas al final del experimento. El sistema de ultrasonido propiedad de Queen Margaret Univerity College (QMUC), Edimburgo fue utilizado para acquirir la imagen dinámica del contorno de la lengua junto con la señal acústica.

Instrumentación y montaje experimental

La posición del transductor de ultrasonido y el montaje experimental están ilustrados en la Figura 1. El movimiento de la lengua fue adquirido con la ayuda de un equipo de ultrasonido (Merlin Ultrasound Scanner; Type 1101, B-K Medical A/S, Herlev, Denmark), junto con la señal acústica. El transductor (Endovaginal End-fire Transducer; Type 8561, B-K Medical A/S, Herlev, Denmark) se fijó justo debajo del mentón del informante gracias a un casco de motocicleta modificado (ver Fig. 1) y diseñado por el departamento de ingeniería mecánica de la universidad Heriot Watt – Edimburgo. Un standoff acústico de 7.5 mm (Sonokit soft, Sonogel –Germany) fue utililzado entre el mentón y el transductor tal y como lo recomiendan Stone et al. (1995). La frecuencia elegida para el funcionamiento del transductor fue 6.5 MHz con un sector de imagen de 160^0. Las imágenes fueron obtenidas a una frecuencia de acquisición de 25 imágenes por segundo (ips).

Figura 1: *Esta ilustración muestra el casco de motocicleta modificado para sujetar el transductor justo debajo del mentón del informante.*

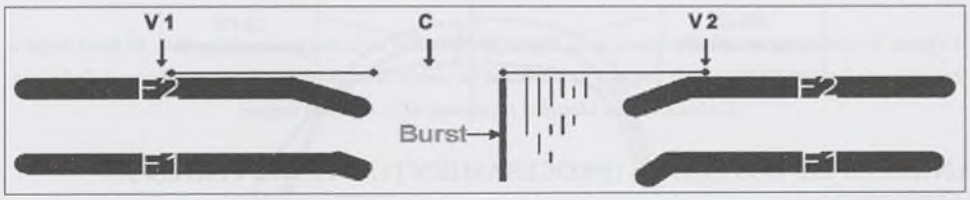


Figura 2: *Ilustración esquemática de las tres anotaciones utilizadas en la señal acústica. La leyenda 'burst' en el gráfico identifica la barra de explosión de la consonante oclusiva.*

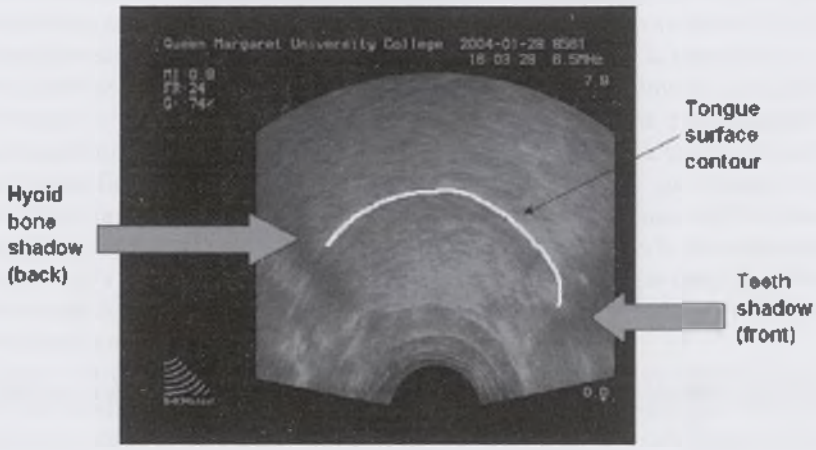


Figura 3: Ilustración de cómo el contorno de la lengua fue delineado en la imagen de ultrasonido. El hueso hoyid se encuentra en la parte posterior de la lengua y la sombra de los dientes se encuentra en la parte posterior de la imagen. La línea blanca identifica el contorno de la superficie lingual.

ANALISIS DE LOS DATOS (PROCESAMIENTO DEL MOVIMIENTO LINGUAL)

Las imágenes recogidas con el equipo de ultrasonido presentaban un formato de sección sagital sincronizada con la señal acústica durante la adquisición de las imágenes. Los datos fueron analizados utillizando el programa Articulate Assistant (Copyright©2005, Articulate Instruments Ltd.).

Medición temporal, cualitativa y cuantitativa del movimiento lingual

Primero, se anotaron tres puntos diferentes en la señal acustica (ver Fig. 2): 1) mitad de V_1, 2) punto central de la fase de oclusión en la oclusiva bilabial, 3) momento durante la V_2. Este último punto fue calculado tomando en cuenta la distancia desde la mitad de V_1 hasta el final de esta vocal. A continuación, este cálculo fue utilizado desde la barra de explosión de la consonante para indicar la tercera anotación en V_2. De esta manera establecimos una distancia simétrica desde ambos lados de la oclusión de la consonante hacia las vocales. Por último, el contorno de la lengua fue indentificado y delineado en cada una de las imágenes de ultrasonido correspondientes a cada anotación (ver Fig.3).

En la sección de resultados también se incluye una media de varios contornos de la lengua que proporcionan ilustración del contorno lingual y su movimiento. Las coor-

denadas x e y del contorno de la lengua fueron exportadas a Matlab para poder comparar los diferentes contornos linguales de las vocales con aquellos de la consonante.

Para el análisis cuantitativo se utilizó una barra para medir el movimiento lingual basado en el punto más alto del contorno C (ver Fig. 4) pero no necesariamente el punto más elevado de la lengua durante las vocales o después de la oclusión.

La Figura 5 muestra un ejemplo de cómo los datos que reflejan el moviento lingual se expondrán en la sección de resultados. La barra de error de la izquierda muestra una diferencia negativa (movimiento descendiente de la lengua) de la distancia entre el contorno V_1 y C. La barra de error de la derecha muestra una diferencia positiva (movimiento ascendente de la lengua) de la distancia entre el contorno C y V_2. Estas dos medidas seran utilizadas como medidas de desplazamiento lingual (MTD). El 'trough effect' sera identificado como tal solamente cuando el primer MTD es negativo y el segundo positivo.

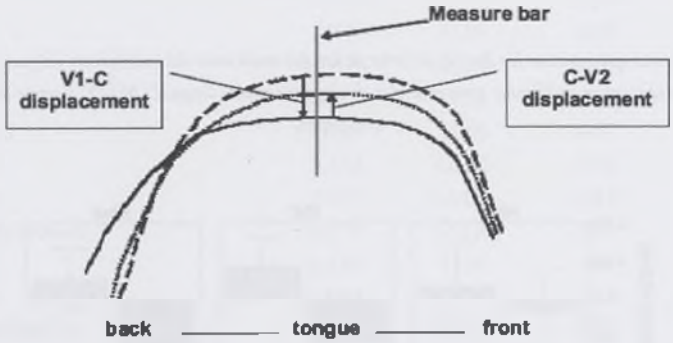


Figura 4: Ilustración esquemática del movimiento lingual desde V_1 a C y desde C a V_2. La leyenda 'measure bar' identifica la barra utilizada para medir el movimiento lingual.

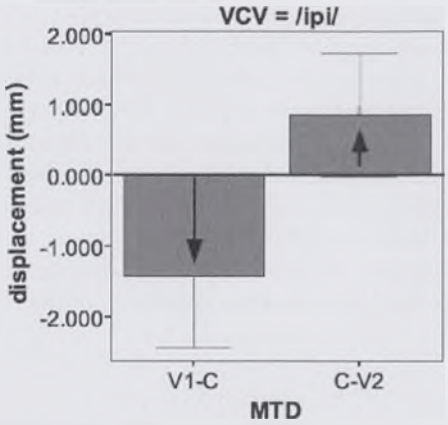


Figura 5: Ilustración que muestra las barras de error de las dos mediciones del movimiento lingual para la secuencia /ipi/. Estas dos medidas seran utilizadas como medidas de desplazamiento lingual (MTD) y seran descritas en milímetros (mm).

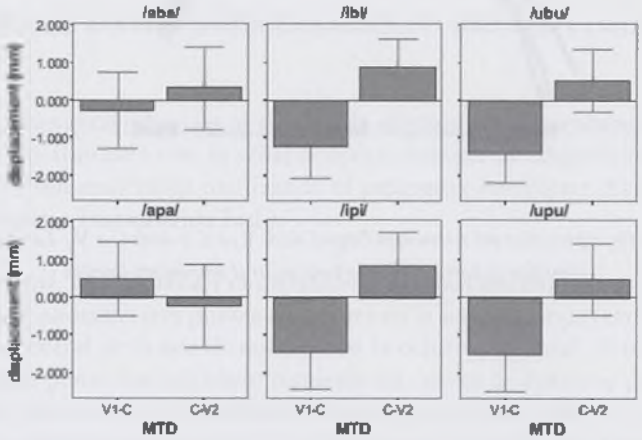


Figura 6: Medias del movimiento lingual por cada una de las diferentes secuencias VCV.

RESULTADOS

Resultados cuantitativos del movimiento lingual

Medimos el movimiento de la lengua desde V_1 a C y desde C a V_2 en los datos recogidos en dos sesiones separadas por un descanso de 20 minutos. Se realizó un análisis de ANOVA de medida repetida con 4 factores: medida de desplazamiento lingual, diferentes tipos de vocal, identidad labial y sesión. La Figura 6 muestra las medias del movimiento lingual para cada vocal, consonante oclusiva bilabial y desplazamiento lingual tomando los datos de ambas sesiones.

Las medias presentadas en la Figura 6 muestran un comportamiento diferente de la vocal /a/ si se compara con las secuencias que contienen la vocal /i/ o /u/.

La tabla de datos incluida a continuación es un resumen de los resultados del análisis ANOVA:

Factor	F	df	Sig.	Significance
Sessions	1.030	(1,9)	.337	
Bilabial	.251	(1,9)	.628	
Vowels	10.676	(2,18)	006 +	**
MTD	28.202	(1,9)	.000	***
Sessions*bilabial	.186	(1,9)	.667	
Sessions*vowels	3.340	(2,18)	.058	
Bilabial*vowels	2.873	(2,18)	.083	
Sessions*bilabial*vowels	1.279	(2,18)	.303	
Sessions*MTD	9.136	(1,9)	.014	*
Bilabial*MTD	6.007	(1,9)	.037	*
Sessions*bilabial*MTD	.220	(1,9)	.650	
Vowels*MTD	25.087	(2,18)	.000	***
Sessions*vowels*MTD	1.102	(2,18)	.353	
Bilabial*vowels*MTD	16.974	(2,18)	.000	***
Sessions*bilabial*vowels*MTD	.874	(2,18)	.434	

+ Greenhouse-Geisser correction. Sig.: (***) = p<0.001, (**) = p< 0.01, (*) = p<0.05.

La interacción sesion *MTD (F (1,9)= 9.136, p<0.05) fue significativa. Esta interacción fue causada por una reducción en la magnitud del desplazamiento lingual de la primera a la segunda sesión. Esto pudo haber sido provocado por un efecto de aprendizaje por parte de los informantes.

Los factores MTD (F (1,9)= 28.202, p<0.001) y vocal (F (1.22,11.00)= 10.676, p<0.01) con correccion Greenhouse-Geisser para el último, fueron significativos. Un análisis de *Post hoc* (Paired-samples t-test) del factor MTD no mostró una diferencia significativa entre las dos medidas de desplazamiento lingual en las secuencias /aba/ y /apa/. Por lo tanto, no encontramos un 'trough effect' en las secuencias VCV con vocal /a/.

Por otro lado, paired-samples t-test revelaron una diferencia significativa entre las dos MTDs – la primera medida fue negativa y la segunda positiva – en las secuencias /ipi/ (t (9) = -7.558, $p < 0.001$), /ibi/ (t (9) = -5.485, $p < 0.001$ y /ubu/ (t (9) = -5.618, $p < 0.001$). Por lo tanto, encontramos el 'trough effect' en ambas secuencias VCV /i/ y /u/. Por el contrario, no encontramos un efecto significativo ni del factor vocal ni del factor identidad labial.

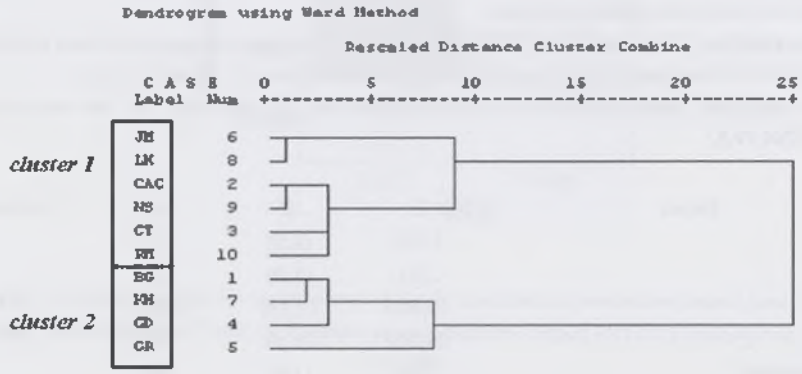


Figura 7: Dendrograma para la secuencia /aCa/.

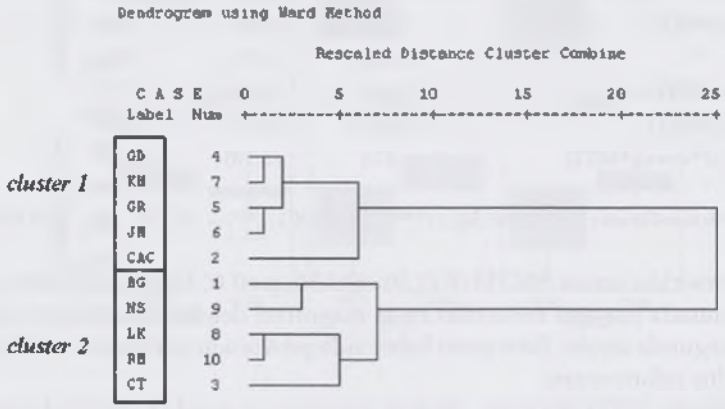


Figura 8: Dendrograma para la secuencia /iCi/.

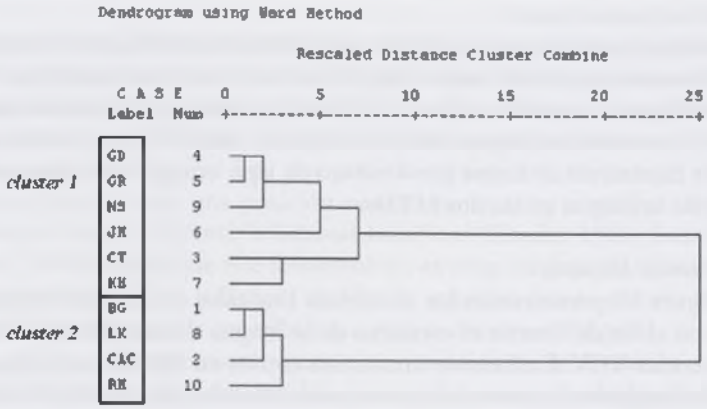


Figura 9: Dedrograma para la secuencia /uCu/.

Resultados cualitativos: Variación del 'trough effect' entre los diferentes informantes

En esta parte del articulo ofreceremos un análisis cualitativo de las diferentes realizaciones de este efecto entre los diferentes informantes.

Analisis cluster

Un analisis cluster jerárquico fue aplicado a los datos de este estudio con el fin de averiguar como los informantes eran agrupados dependiendo de la similaridad de sus MTDs. Este análisis nos permitió observar similaridades y diferencias entre los informantes. Una serie de dendrogramas fueron creados con SPSS utilizando el metodo cluster Ward (distancia Euclidiana) que es un tipo de análisis multivariante que evalua las diferencias entre los clusters. Los dendrogramas muestran los informantes en diferentes grupos para las secuencias VCV /a/, /i/ y /u/.

Los resultados para las secuencias /aCa/ se muestran en la Figura 7. Se identificaron dos clusters diferentes: el primer cluster identifica a todos los informantes con un movimiento lingual de tipo 'trough' con un movimiento lingual descendente desde V_1 hasta la oclusión de la consonante y un consiguiente ascenso siguiendo la oclusion hasta V_2. El segundo cluster identifica a todos los informantes con un movimiento lingual contrario al 'trough effect' donde la lengua se encuentra en una posición más elevada durante la oclusión de la consonante que durante V_1 y V_2.

Los resultados para las secuencias /iCi/ se muestran en la Figura 8. Se identificaron dos clusters diferentes: todos los informantes mostraron un movimiento lingual tipo 'trough' para las secuencias /iCi/ pero también mostraron diferencias en su producción. El primer cluster identifica a los informantes con un movimiento lingual de tipo 'trough'

pero la magnitud del movimiento lingual en las dos MTDs fue más pequeña que en los informantes del segundo cluster.

Los resultados para la secuencia /uCu/ se muestran en la Figura 9. Se identificaron dos clusters diferentes: el primer cluster identifica a todos los informantes que mostraron un movimiento lingual descendente desde V_1 hasta la oclusion pero mucha variación en la dirección del movimiento lingual desde la oclusión hasta V_2. Los informantes en el segundo cluster mostraron un fuerte movimiento de tipo 'trough' sin ninguna variación en la dirección de la lengua en las dos MTDs.

Media de los contornos linguales

En la Figura 10 presentamos los contornos linguales de algunos informantes representativos con el fin de ilustrar el contorno de la lengua durante la producción de las diferentes secuencias VCV. Utilizamos una rutina escrita en Matlab para crear la media de los contornos linguales a partir de cinco repeticiones de cada vocal y la oclusiva bilabial sorda /p/. La punta de la lengua esta situada siempre a la derecha. Los contornos linguales en la Fig. 10 muestran no sólo una gran diferencia en la dirección y magnitud del movimiento lingual, pero también una gran diferencia en la configuración del contorno de la lengua entre los informantes. Los informantes CT y BG en la secuencia /ipi/ muestran un movimiento lingual durante la oclusión de la consonante de dirección y magnitud muy similares pero la configuración de la lengua es muy diferente. Por otra parte, si comparamos la producción de esta misma secuencia /ipi/ del informante CT con la del informante GR, la configuración de la lengua es muy similar en los dos informantes pero la magnitud del movimiento lingual varia.

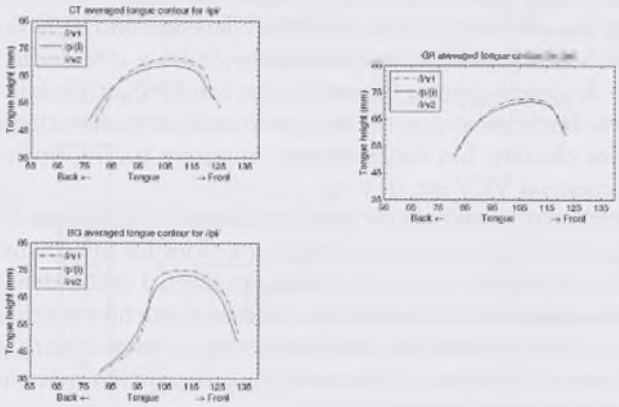


Figura 10: *Media (5 repeticiones) de los contornos linguales de los informantes CT, BG y GR durante la producción de V_1 (línea guión+punto), oclusiva labial /p/ (línea contínua) y V_2 (linea de guiones) en la secuencia /ipi/. Estos tres informantes eran mujeres.*

CONCLUSIONES

Los resultados de este experimento respaldan la reivindicación por parte de Linblom et al. (2002) de la existencia del 'trough effect' en las secuencias VCV donde la consonante intervocálica es una oclusiva bilabial. Sin embargo, dado que las secuencias VCV con vocal /a/ no mostraron un movimiento ascendente durante la fase de oclusión en la consonante, esto sugiere que una posible explicacion del 'trough effect' como un movimiento lingual hacia una posición neutral no es muy plausible. Además, la falta de movimiento lingual durante la bilabial nasal /m/ (Houde, 1968; Svirsky et al. 1997; Fuchs et al., 2004) indica que este fenómeno no es simplemente un efecto coarticulatorio de la oclusion bilabial. Los hallazgos de Houde sugieren que la presión intraoral podría ser un factor importante que contribuiría a este movimiento lingual.

En este estudio encontramos una gran variación en la extensión del movimiento lingual individualmente y entre los diferentes informantes dependiendo de la secuencia VCV. Lindblom et al. (2002) informaron de un movimiento lingual de 3 mm basándose en solamente un informante, mientras que Houde (1968) encontró un movimiento lingual descendente de 2.5 mm, también esta vez basandose en un solo informante. En este estudio encontramos una media de 1.5 mm de movimiento lingual durante la oclusión en las secuencias /iCi/ y /uCu/ basándonos en los datos obtenidos de diez informantes.

La media de los contornos linguales mostraron una diferencia no sólo en la magnitud del movimiento lingual pero también en la configuración del contorno de la lengua entre diferentes informantes. Esta variación puede haber sido causada por diferentes estilos de habla, diferencias fisiológicas o incluso puede haber sido debida a una causa aerodinámica.

En una futura investigación estudiaremos de que manera las limitaciones aerodinámicas estan relacionadas con este fenómeno. Nos centraremos en establecer si la presión intraoral podría ser la causante de la variación observada entre los diferentes informantes y si el control motor podría estar anticipando estas fuerzas aerodinámicas.

AGRADECIMIENTOS

Estamos agradecidos a Alan Wrench, de Articulate Instruments, que diseñó el programa utilizado para analizar los datos y a Bruce Davis, de Heriot-Watt Univeristy, que diseñó y adaptó el casco utilizado en esta investigación. El equipo de ultrasonido fue adquirido gracias a una beca del Fondo de Inversion en Investigacion Cientifica.

REFERENCIAS

BELL-BERTI, F.; HARRIS, K. (1974): "More on the motor organization of speech gestures". *Haskins Laboratories:* Status Report on Speech Research SR-37/38, 73-7.

FUCHS, S.; HOOLE, P.; BRUNNER, J.; INOUE, M. (2004): "The trough effect – an aerodynamic phenomenon?". In *From Sound to Sense* - MIT meeting.

GAY, T. (1975): "Some electromyographic measures of coarticulation in VCV-utterances. *Haskins Laboratories: Status Report on Speech Research* SR-44: 137-145.

HOOLE, P.; MUNHALL, K.; MOOSHAMMER, C. (1998): "Do Airstream Mechanisms Influence Tongue Movement Paths?". *Phonetica 55 (3)*: 131-46.

HOUDE, R. A. (1968) "A study of tongue body motion during selected speech sounds. *Speech Communications Research Laboratory Monograph* No.2, Santa Barbara, California.

LINDBLOM, B.; SUSSMAN, H.M.; MODARRESI, G.; BURLINGAME, E. (2002) "The trough effect: implications for speech motor programming". *Phonetica* 59 (4): 245-262.

McALLISTER, R.; ENGSTRAND, O. (1992) "Interpretations of tongue movement patterns in VCV sequences". In D. Huber (Ed.): *Papers from the Sixth Swedish Phonetics Conference.* Technical report No. 10, pp.115-120 (Department of Information Theory, School of Electrical and Computer Engineering, Chalmers University of Technology, Gothenburg).

ÖHMAN, S.E.G. (1967): "Numerical model of coarticulation". *Journal of the Acoustical Society of America* 41: 310-319.

STONE, M.; DAVIS, E.P. (1995): "A head and transducer support system for making ultrasound images of tongue/jaw movement". *Journal of the Acoustical Society of America* 98 (6): 3107-3112.

SVIRSKY, M.; STEVENS, K.; MATTHIES, M.; MANZELLA, J.; PERKELL, J.; WILHELMS-TRICARICO, R. (1997): "Tongue surface displacement during bilabial stops". *Journal of the Acoustical Society of America* 102: 562-571.

ESTRUCTURA JERÁRQUICA DE LA PROSODIA DEL ESPAÑOL ORAL POR MEDIO DE XML

EDUARDO VELÁZQUEZ[1]
Doctorante, Freie Universitat Berlin

1. INTRODUCCIÓN

Toda interacción comunicativa consiste en una serie de turnos conversacionales entrelazados por los hablantes y la mayoría de estos turnos están compuestos de enunciados. Sin embargo, cada turno también puede contener solapamientos de enunciados, apoyos paralingüísticos, pausas, interrupciones o producciones ininteligibles.

Dependiendo de sus funciones en el discurso, los enunciados pueden ser catalogados como declaraciones, interrogaciones, interjecciones o suspensiones. Todos ellos se componen de unidades entonativas cargadas de información tonal que suele ser más relevante al inicio, alrededor del núcleo y al final de cada unidad. Estas unidades, a su vez, se componen de sílabas, disyunciones y fragmentos. Las sílabas pueden agruparse en pies para obtener la estructura rítmica de los enunciados, y cada sílaba, la unidad prosódica mínima, concentra los rasgos auditivos de tono, prominencia, volumen y duración.

A partir de esta organización jerárquica tan flexible propongo un modelo de representación computacional más cercano a la lengua oral y fácil de procesar. El sistema se fundamenta en la colocación de información prosódica en una estructura jerárquica creada en el *Extensible Markup Language* (XML), que está especialmente diseñado para representar y sacar provecho de estructuras jerárquicas. La estructura resultante permite la introducción de 1) volumen, duración, cantidad y calidad de cada sílaba, 2) distribución de pies, 3) tonos de inicio, finales y nucleares de cada unidad entonativa, 4) tipo de enunciado y 5) otros fenómenos tales como pausas, solapamientos, interrupciones, producciones ininteligibles, etc.

1 El autor agradece el financiamiento del Consejo Nacional de Ciencia y Tecnología (CONACYT).

Este lenguaje de etiquetado se utiliza actualmente en un corpus de conversaciones en las variedades mexicana e ibérica del español oral.

2. XML (EXTENSIBLE MARKUP LANGUAGE)

En la presente sección se hará una introducción a grandes rasgos del XML y del proceso de creación de una estructura jerárquica que permita el estudio de los rasgos prosódicos.

Hay tres características que distinguen al XML (Bray *et al.* 2004) de otros lenguajes de etiquetado: a) su énfasis en el *etiquetado descriptivo* en lugar del etiquetado de procesos, lo que abre la posibilidad de que el mismo documento pueda ser procesado de muy distintas maneras por medio de hojas de estilo, utilizando únicamente las partes de él que se consideren relevantes o trabajando la misma parte del archivo en distintos procesos; b) su concepto de *tipo de documento*, pues se considera que los documentos pueden ser de distintos tipos y cada tipo de documento debe definirse formalmente según sus partes constituyentes y su estructura; y c) su *independencia* de cualquier sistema de *hardware* o *software*, ya que todo documento de XML, sin importar el idioma o sistema de escritura que se emplee, utiliza el mismo método de codificación subyacente de caracteres en datos binarios, implementado por el estándar internacional Unicode, que provee un conjunto de caracteres que cubre la mayoría de sistemas de escritura actuales e históricos de todo el mundo. (Sperberg-McQueen y Burnard 2004: 2.1)

La mayoría de los documentos (por ejemplo, las transcripciones de la lengua oral) pueden dividirse en componentes (turnos y enunciados). Estas a su vez pueden dividirse en subcomponentes (frases entonativas, pies, sílabas y segmentos).

En XML, estos componentes se llaman elementos. Cada elemento representa un componente lógico de un documento. En lugar de que cada elemento describa información sobre el formato, describe el papel lógico de sus elementos, la abstracción que representa. Los elementos pueden contener otros elementos y también pueden contener las palabras, frases y otros signos que conforman lo que comúnmente se llama texto del documento. A este texto se le llama en XML datos de caracteres (*character data*). Estos datos de caracteres son codificados, como se dijo anteriormente, en el conjunto de caracteres UTF-8 (*Universal Character Set Transformation Format 8*) del estándar Unicode.

Hay además una característica de XML, la entidad externa, gracias a la cual, un documento de XML puede dividirse en distintos archivos y, según la terminología de XML, cada uno de ellos será una entidad. Esas entidades podrían incluso estar dispersas en Internet. Mientras que los elementos de XML describen la estructura lógica del documento, las entidades siguen la huella de la ubicación de las piezas de bytes que conforman un documento de XML. Esta es la estructura física del documento. (Goldfarb y Prescod 2000: 27-28)

2.1. Etiquetado

Hasta ahora se ha discutido sobre el modelo conceptual del XML (la estructura arbórea de elementos), su estrategia para codificar caracteres (el Unicode) y su mecanismo

para administrar el tamaño y complejidad de los documentos (las entidades). Sobre lo que no se ha hablado aún es sobre la manera en la que se representa la estructura lógica del documento y en la que se vinculan todas las entidades físicas: el etiquetado (*markup*).

Las características generales de la sintaxis del etiquetado en XML son las siguientes:

- El etiquetado se diferencia de los datos de caracteres por el uso de caracteres especiales llamados *delimitadores*. Se puede decir que una etiqueta (*tag*) es todo lo que se encuentre entre un signo de menor que „<" y uno de mayor que „>", o entre un signo de *et* „&" y un punto y coma „;". (Goldfarb y Prescod 2000: 30)
- El nombre de las etiquetas es sensible a la diferencia entre mayúsculas y minúsculas. Por lo tanto, <TAG>, <Tag> y <TaG> se interpretarían como tres etiquetas distintas.
- Los espacios en blanco fuera del etiquetado siempre se conservan, mientras que los espacios entre los delimitadores se pueden preservar o ignorar. (Goldfarb y Prescod 2000: 724-726, Sperberg-McQueen y Burnard 2004: 2.3.1)

2.2. Propuesta de estructura jerárquica prosódica para documentos de XML

En esta sección se presentará la estructura jerárquica prosódica que nos permitirá tanto el señalamiento de los fenómenos prosódicos dentro del texto, como su especificación por medio de valores reales.[2] En esta estructura (fig. 1), el elemento raíz se llamará <Transcript>, que consta de encabezado, <Header>, y texto, <Text>.

Todos los subelementos del elemento opcional <Header> son lo que se denomina como hojas, excepto <Participants>, dado que no tienen más elementos subordinados y su contenido es a base de datos de caracteres. Los elementos <Class> y <Acoustic_quality> son diferentes a los demás porque tienen atributos. En la fig. 1, se indican en letra de estilo regular los atributos necesarios y en letra cursiva los atributos opcionales. El atributo *Type* de <Acoustic_quality> tiene tres valores posibles predefinidos (A, B o C), mientras que el atributo *Type1* de <Class> tendrá un valor a base de datos de caracteres. El elemento <Participants> puede tener uno o más elementos <Speaker>, que tienen a su vez una serie de subelementos, y algunos de ellos tienen sus propios atributos.

Del lado derecho en la fig. 1 se ve que el único descendiente directo de <Text> es <Turn>, el cual tiene un indicador de frecuencia que especifica que debe aparecer una o más veces y que tiene dos atributos: *Name* y *Trans*. *Name* es un atributo obligatorio que admite cualquier tipo de caracteres, lo cual nos permitirá introducir los códigos asignados a

2 La idea de la representación de la lengua humana en un lenguaje de etiquetas no es nueva. Distintos equipos de trabajo con una mayoría de especialistas en informática se dedican a hacer posible la navegación por la web semántica (por ejemplo, Berners-Lee *et al.* 2001), basada en una clasificación conceptual de la información; mientras otros se ocupan de la síntesis del habla (*Speech Synthesis Markup Language*, SSML, Burnett *et al.* 2004) y el reconocimiento de voz (*Voice Extensible Markup Language*, VoiceXML, McGlashan *et al.* 2004). De mayor interés para los científicos del lenguaje son las aplicaciones EXMARaLDA (Schmidt 2004, 2005) y TEI (Sperberg-McQueen y Burnard 2004), cuyos principios se relacionan en cierta medida con mi propuesta.

los nombres de cada participante de la conversación. El atributo *Trans* es opcional, puesto que los tipos de transición en los que un turno sea la continuación del turno anterior (cont) o en los que un turno sea producido simultáneamente (overlap) serán considerados como excepciones y deberán ser señalados por medio de dicho atributo.

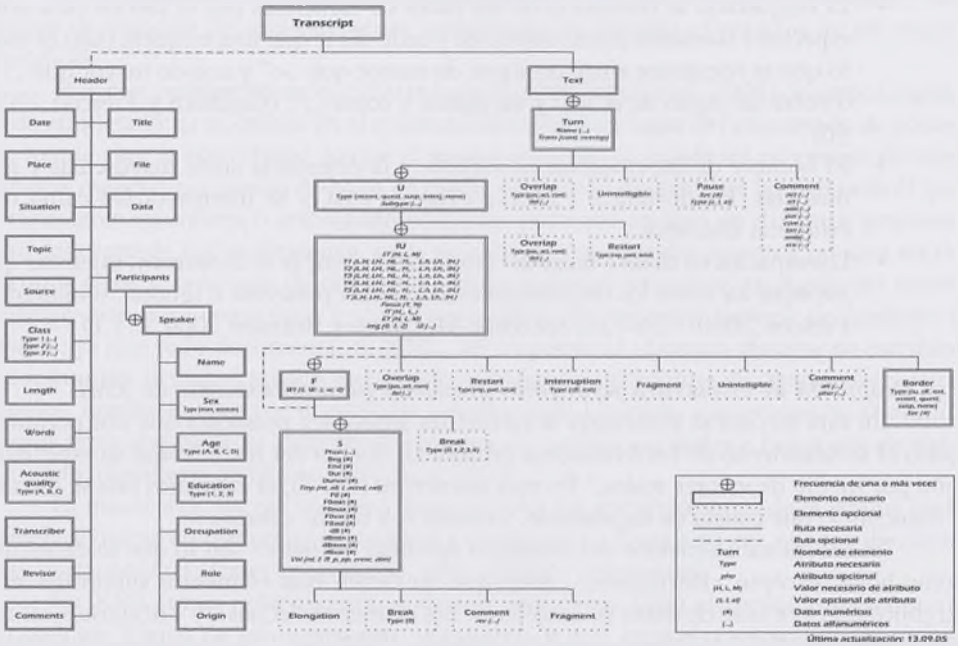


Fig. 1: Representación gráfica de la estructura jerárquica prosódica de PROSODY.DTD.
Fuente: Diseño propio

A su vez, cada <Turn> tendrá necesariamente un subelemento <U> (*utterance*). Los subelementos restantes de <Turn> son opcionales. Con <Overlap> se deberá indicar el inicio y el final de los fragmentos producidos simultáneamente, de manera pasiva (pas), activa (act) o si alguno de los textos simultáneos se extiende a lo largo de más de un turno (cont). También se podrá asignar un número o nombre de referencia por medio del atributo *Ref*. Este elemento puede aparecer a distintos niveles dentro de la estructura jerárquica.

El elemento <Unintelligible> señalará las producciones ininteligibles en el texto de la transcripción. Por otra parte, el atributo *Type* de <Pause> tiene los valores s, l y xl, mientras que *Sec* permite la especificación de la duración en segundos. Finalmente, los atributos opcionales de <Comment> ofrecen un espacio para describir los distintos tipos de fenómenos o eventos relevantes durante la conversación.

Cada enunciado, <U>, deberá contener uno o más elementos <IU>, correspon-

dientes a las unidades o frases entonativas (*intonation unit*). Opcionalmente pueden aparecer elementos <Overlap> y <Restart>. Los valores del atributo de <Restart> significan repetición (rep), reformulación parcial (part) o total.

Mucho más compleja es la declaración de <IU>, cuyos atributos sirven principalmente para la colocación de la información tonal en las distintas posiciones: al inicio (*start tone*, *ST*, y su valor en hercios, *SHz*), al final (*end tone*, *ET*), en el núcleo (*T*) y hasta cinco tonos prenucleares (*T1* a *T5*). Los valores de estos atributos, a pesar de estar ligeramente modificados para que signos como * o + no interfieran con la sintaxis de XML, corresponden al inventario de tonos del Sp-ToBI (Beckman *et al.* 2002: 32-34, Sosa 2003: 188-190), por ejemplo: L.H corresponde al tono compuesto L*+H, Lh. a L+!H*, y lH. a ¡L+H*.

Además hay los atributos opcionales *Focus* (con valor afirmativo o negativo) y el tono intermedio, *IT* (*intermediary tone*), que podría tener el valor alto o bajo. Y por último, los atributos *img* (*image*) e *id* (*identity number*), cuyo uso se explicará más adelante.

En realidad todos los subelementos de <IU> son opcionales, excepto <Border>, sobre todo con la finalidad de que la estructura también permita documentos de XML menos detallados, sin estructura rítmica o tonal. <Border> es un elemento con el cuál se puede aprovechar los signos gráficos al final de cada unidad entonativa y especificarlos por medio del atributo *Type*, incluso se podrá suministrar la duración abstracta en segundos utilizando el atributo *Sec*. Además, en este nivel de la estructura se introducen los elementos <Interruption> y <Fragment>, mientras que a <Comment> se le agregan dos nuevos atributos que aparecen únicamente en este nivel: *alt* y *pho*.

De mayor importancia son las líneas punteadas que nos llevan a <S> (*syllable*), directamente por una de las rutas o pasando primero por <F> (*foot*). De este modo no será absolutamente necesario introducir la información sobre la estructura rítmica cuando sea más práctico analizar las sílabas como descendientes directas de la unidad entonativa.

Al nivel de la sílaba se requiere también del elemento <Break>, que también corresponde a las características descritas en el inventario del Sp-ToBI.

En caso de que se incluya la estructura rítmica de los pies métricos, se podrá echar mano de los valores del atributo *Wt* (*weight*) de <F>, con los cuales podrá especificarse el peso de los pies como s (*strong*), w (*weak*) ó 0 para los casos de pies libres.

Cada <S> tiene una serie muy importante de atributos: el valor fonético (*Phon*); inicio, final, duración, variación relativa y tempo de la sílaba (*Beg*, *End*, *Dur*, *Durvar*, *Tmp*); su frecuencia fundamental, con valores mínimos, máximos, variación relativa y estandarizada (*F0*, *F0min*, *F0max*, *F0var*, *F0std*), así como su intensidad, con valores mínimos, máximos, variación relativa y estandarizada, y volumen relativo (*dB*, *dBmin*, *dBmax*, *dBvar*, *dBstd*, *Vol*).

Con el subelemento <Elongation> se señalarán las posiciones en las que ocurra un alargamiento segmental. Los elementos <Break> y <Comment> ya habían sido introducidos previamente, pero con el valor 0 del atributo *Type* de <Break> se podrán señalar los límites morfológicos de dos sílabas dentro de un hiato, y con el nuevo atributo *rec* de <Comment> se podrá incluir la reconstrucción normativa dentro de los límites de una

sílaba. El elemento <Fragment>, que también es un subelemento de <IU>, señala una producción fragmentada.

Al final, este modelo debe ser traducido en una definición de tipo de documento de XML (*Document Type Definition* o DTD) que actúe como la gramática según la cual se validan todos los documentos que se declaren conformarse según la definición PROSODY.DTD.

2.3. Creación de documentos XML a partir de datos de PRAAT

Las grabaciones digitales adaptadas a este sistema de transcripción pertenecen a un corpus de conversaciones espontáneas entre hablantes de Madrid y la Ciudad de México, que a su vez representan las variedades estándar castellana y mexicana respectivamente.

La transcripción básica de estas conversaciones bajo estrictos criterios de codificación permiten una administración eficaz. Pero estas transcripciones también representan la fuente de información para las rejillas textuales (*text grids*) de PRAAT, en las cuales se hará posteriormente la transcripción fonética. En otras filas (*tiers*) se podrá representar, por ejemplo, la estructura rítmica de las sílabas y sus valores estandarizados de variación de la frecuencia fundamental o F_0 (Cantero 2002: 144-54).

La razón por la cual se prefiere utilizar el programa PRAAT en este proceso no solamente se debe a sus herramientas de análisis acústico, sino también al hecho de que permite crear subprogramas que automaticen y coordinen las funciones básicas originales.

Por medio del uso de tales subprogramas se puede extraer los fragmentos de grabación correspondientes a cada unidad entonativa o enunciado y crear imágenes de análisis acústico. Estas imágenes, en las que se puede plasmar la frecuencia fundamental, el espectrograma y el contenido de todas las filas de texto, serán desplegadas más adelante en la página *web* resultante. Los subprogramas también permiten la asignación de variables a los parámetros acústicos de cada sílaba y la repetición de este proceso por cada sílaba conformando una unidad entonativa. Estos valores podrán también irse anexando continuamente a un archivo de texto, utilizando una sintaxis similar a la del XML.

Puesto que los documentos de XML constan únicamente de texto, no hay necesidad de conversiones o adaptaciones de formatos. Todas las posibles repeticiones, los errores o los caracteres fonéticos utilizados por el programa PRAAT (incompatibles con el Unicode) pueden ser reemplazados por medio de una macro de Visual Basic.

2.4. Aplicación de estilo con XSLT y HTML dinámico

Del proceso total de creación de un documento de XML, la aplicación de estilo es seguramente la que requiere del conocimiento de una mayor cantidad de lenguajes computacionales, herramientas y aplicaciones: XSLT, XPath, HTML, CSS, JavaScript, etc.

En esta sección se presentará a grandes rasgos las partes que constituyen la hoja de estilo que denominaremos PROSODY.XSL. En esta hoja de estilo no sólo se puede dar formato al texto, sino que, por ejemplo, se puede también incluir:

1) una tabla con los contenidos del encabezado, en caso de haberlos;

2) una tabla independiente con información sobre los participantes de la conversación;
3) un sumario en el que se haga un recuento de todos los fenómenos prosódicos codificados en el documento, y
4) una lista de convenciones utilizadas a todo lo largo del texto.

Además, se puede agregar a la hoja de estilo también ciertos elementos que harán que al inicio de cada frase entonativa haya un enlace directo a una ventana de análisis donde incluso se pueda escuchar el fragmento de grabación correspondiente.

Es aquí donde entran en escena los atributos *id* e *img* del elemento <IU>, pues son estos los que proveerán la información sobre la clave con la que se vinculará la producción lingüística con su archivo de audio y sobre el tipo de imagen que se desplegará frente a cada frase entonativa. En caso de que haya un análisis y esté vinculado por medio del valor de *id*, se correrá, por ejemplo, un programa en JavaScript. Esta secuencia de instrucciones indicará al navegador que abra una ventana con medidas específicas en la que desplegará una página *web* previamente elaborada. En este documento de HTML se hará referencia a una imagen del análisis acústico en PRAAT y al archivo de audio correspondiente al fragmento.

Finalmente, es necesario indicar que el objetivo del XML no es únicamente el de producir páginas *web* en HTML, sino el de enriquecer estos documentos con una estructura jerárquica subyacente e información autodescriptiva. Esto es particularmente útil en el caso de <S>, cuyos atributos contienen importantes datos fonéticos listos para ser utilizados. Esta es la diferencia más importante entre XML y HTML: en XML permanece toda esta información almacenada y puede ser extraída en cualquier momento.

La consulta de esta información puede hacerse de manera dinámica, por ejemplo, haciendo que cada sílaba en el texto reaccione cambiando a color rojo cuando el apuntador pase sobre ella y que despliegue una cinta amarilla con la información acústica más importante. Al hacer clic sobre la sílaba, se abrirá una ventana de diálogo en la que se desplegarán todos los valores de los atributos de <S>. El resultado final en un navegador de Internet, se puede ver en la fig. 2. Para probar esta demostración en tiempo real, refiérase a la dirección electrónica: [http://www.geocities.com/utka/df1-04-1.html].

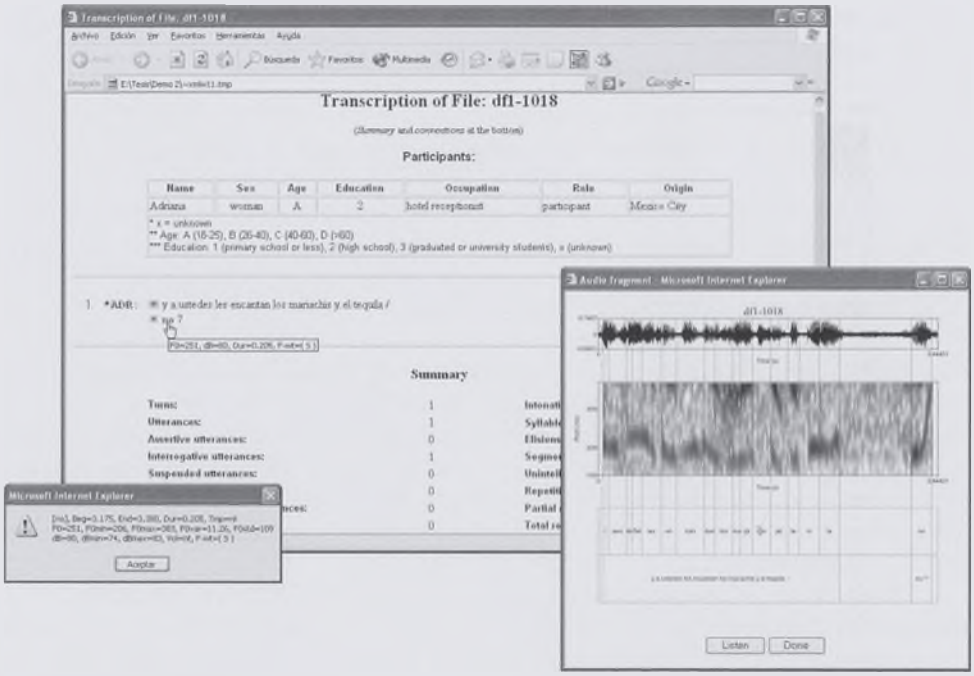


Fig. 2: Página *web* dinámica resultante tras la aplicación de estilo. Fuente: Diseño propio

3. RESUMEN

En este artículo se ha presentado un sistema de transcripción de la lengua oral basado en XML, diseñado con el propósito de ser aplicado como una herramienta para el estudio de fenómenos prosódicos en u corpus de conversaciones producidas por hablantes de las variedades españolas de Madrid y la Ciudad de México.

El proceso de transformación inicia con la transcripción básica y la marcación mínima de fenómenos prosódicos relevantes en las grabaciones de las interacciones comunicativas. Tal información constituye la fuente primaria a la hora de segmentar el archivo de audio en PRAAT. Otros análisis se agregan en nuevas filas a la rejilla textual. La información fonética y textual correspondiente a cada segmento analizado (turno, enunciado, unidad entonativa o sílaba) es extraída y escrita en un archivo de texto siguiendo una sintaxis similar a la de XML. Una vez que el resultado esté libre de errores, bien formado y que sea válido según su definición de tipo de documento, se procesa y formatea por medio de una hoja de estilo, que da lugar a un documento HTML enriquecido y dinámico (véase la fig. 3). El documento resultante se vuelve algo muy distinto de lo que comúnmente se considera un texto o una página *web*. No sólo su apariencia es importante, sino también el tratamiento de los datos por medio de la coordinación y combinación de texto, bases

de datos, imágenes y archivos de audio, para ofrecer una plataforma fácil de usar, pero muy potente.

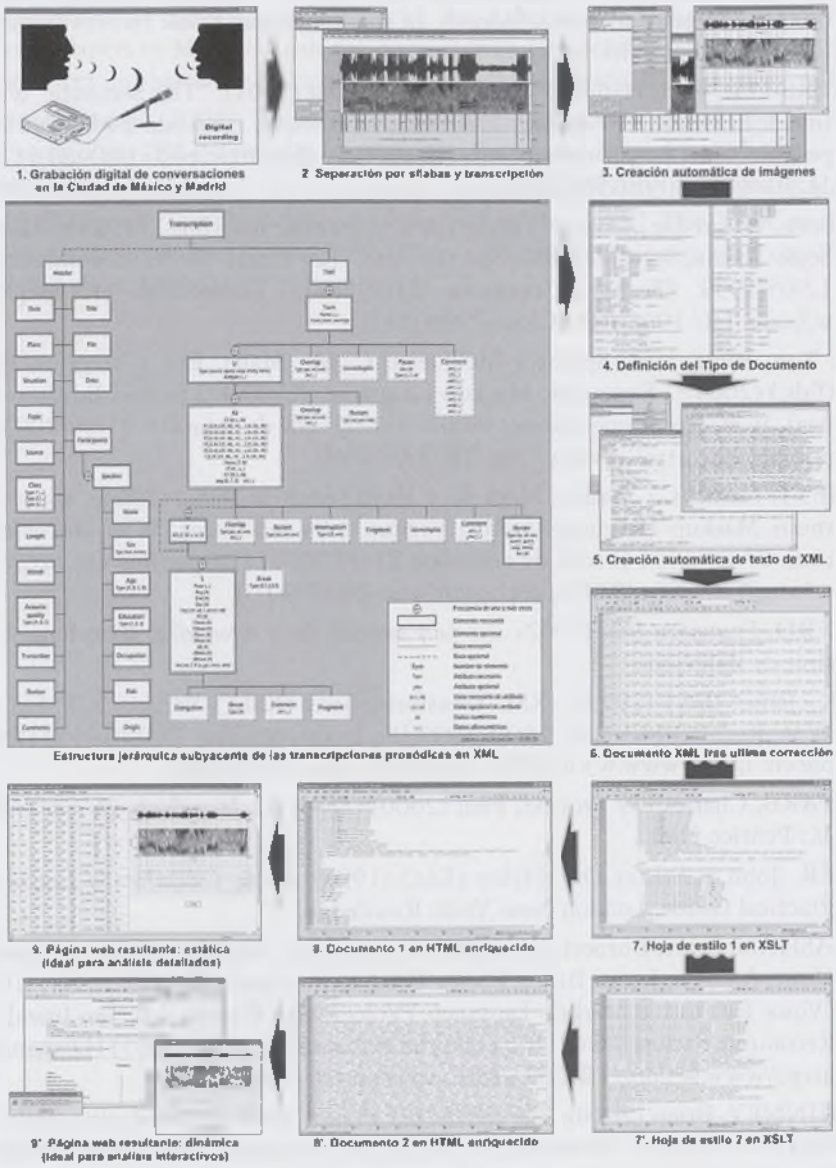


Fig. 3: Proceso de creación de documentos de XML. Fuente: Diseño propio

BIBLIOGRAFÍA

BECKMAN, Mary E.; Díaz-Campos Manuel; Tevis McGory, Julia y Morgan, Terrell A. (2002): "Intonation across Spanish, in the Tones and Break Indices framework". En: *Probus*, 14: 9-36.

BERNERS-LEE, Tim; Hendler, James y Lassila, Ora (2001): "The Semantic Web" [en línea]. En *Scientific American*: 17/05/2001 [Fecha de consulta: 21/09/2005] [Disponible en: http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21]

BOS, Bert; Wium Lie, Håkon; Lilley, Chris y Jacobs, Ian (Eds.) (1998): "Cascading Style Sheets, level 2. CSS2 Specification" [en línea]. W3C Recommendation: 12/05/1998. [Fecha de consulta: 21/09/2005] [Disponible en: http://www.w3.org/ TR/ 1998/REC-CSS2-19980512]

BRAY, Tim; Paoli, Jean; Sperberg-McQueen, C. M., Maler, Eve y Yergeau, François (Eds.) (2004): «Extensible Markup Language (XML) 1.0 Second Edition» [en línea]. W3C Recommendation: 04/02/2000 [Fecha de consulta: 21/09/2005] [Disponible en: http://www.w3.org/TR/REC-xml]

BURNETT, Daniel C.; Walker, Mark R. y Hunt, Andrew (Eds.) (2004): «Speech Synthesis Markup Language (SSML) Version 1.0» [en línea]. W3C Recommendation: 07/09/2004 [Fecha de consulta: 21/09/2005] [Disponible en: http://www.w3.org/TR/ 2004/REC-speech-synthesis-20040907/]

CANTERO, Francisco José (2002): *Teoría y análisis de la entonación*. Barcelona: Universitat de Barcelona.

CLARK, James (Ed.) (1999): „XSL Transformations (XSLT) Version 1.0" [en línea]. W3C Recommendation: 16/11/1999 [Fecha de consulta: 21/09/2005] [Disponible en: http://www.w3.org/TR/1999/REC-xslt-19991116]

GOLDFARB, Charles F. y Prescod, Paul (2000): *The XML Handbook*, 2a. ed. London *et al.*: Pentice Hall.

LAWLER, John y Aristar Dry, Helen (Eds.) (1998): *Using Computers in Linguistics*. A Practical Guide. London/New York: Routledge.

MCGLASHAN, Scott; Burnett, Daniel C.; Carter, Jerry; Danielsen, Peter; Ferrans, Jim; Hunt, Andrew; Lucas, Bruce; Porter, Brad; Rehor, Ken y Tryphonas, Steph (2004): "Voice Extensible Markup Language (VoiceXML) Version 2.0" [en línea]. W3C Recommendation: 16/03/2004. [Fecha de consulta: 21/09/2005] [Disponible en: http://www.w3.org/TR/2004/REC-voicexml20-20040316/]

MCWHINNEY, Brian (2000): *The CHILDES Project*. Tools for Analyzing Talk, Third Edition. Volume I: Transcription Format and Programs. New Jersey/London: Erlbaum.

SCHMIDT, Thomas (2002): „EXMARaLDA - ein System zur Diskurstranskription auf

dem Computer". *Arbeiten zur Mehrsprachigkeit*, Serie B (34). Hamburg. [Disponible en: http://www.rrz.uni-hamburg.de/exmaralda/Daten/4D-Literatur/AZM.pdf]

SCHMIDT, Thomas (2005): "Time-based data models and the Text Encoding Initiative's guidelines for transcription of speech". En: Arbeiten zur Mehrsprachigkeit (Working Papers in Multilingualism), Serie B (62). [Disponible en: Hamburg. http://www.rrz.uni-hamburg.de/exmaralda/Daten/4D-Literatur/SFB_AzM62.pdf]

SOSA, Juan Manuel (2003): "La notación tonal del español en el modelo Sp-ToBI". En: Prieto, Pilar (Ed.). *Teorías de la entonación.* Barcelona: Ariel. Pp. 185-208.

SPERBERG-MCQUEEN, C. M. y Burnard, Lou (Eds.) (2004): "TEI P5. Guidelines for Electronic Text Encoding and Interchange" [en línea]. The TEI Consortium. [Fecha de consulta: 21/09/2005] [Disponible en: http://www.tei-c.org/P5/Guidelines/]

LA ENTONACIÓN URBANA DE VALLADOLID: APROXIMACIÓN A SU ESTUDIO

Francisco José Zamora Salamanca, Micaela Carrera De La Red Y Mª Mar Meléndez Matías
Universidad de Valladolid

1. INTRODUCCIÓN

En esta comunicación abordamos el estudio de la entonación de la ciudad de Valladolid en el marco del proyecto Amper. Presentamos los primeros resultados del análisis acústico correspondientes a una parte del *corpus* experimental obtenido de las grabaciones de lectura de frases de una informante vallisoletana que, de acuerdo con las directrices del proyecto, no tenía estudios universitarios, estaba entre los 25 y los 45 años de edad, y había nacido en la ciudad de Valladolid. En concreto, nuestra informante estaba cercana a los cuarenta años, y su profesión era, cuando se hizo la encuesta en el mes de julio de 2003, la de ama de casa.

2. METODOLOGÍA

Las grabaciones se efectuaron, en el domicilio de la informante, en una grabadora analógica Marantz PMD-222 a la que se acopló un micrófono unidireccional MB 4000C. La informante leyó todas las frases del *corpus* experimental, y posteriormente respondió a diez cuestiones destinadas a la obtención de habla inducida. Por último, se grabaron 30 frases obtenidas como muestras de habla espontánea mediante el procedimiento de *Maptask*. Una vez grabadas todas las frases, procedimos a la digitalización de cada una de ellas por medio del programa GoldWave (versión 4.25).

De las 63 oraciones declarativas del *corpus* experimental previamente digitalizadas sólo tuvimos en cuenta, para el presente estudio, las nueve oraciones básicas sin expansión. Obtuvimos así, contando las tres repeticiones de cada frase, 27 archivos de sonido para las oraciones declarativas y otros 27 para las interrogativas (54 en total para cada informante). Estos materiales fueron posteriormente analizados utilizando el entorno MatLab con las aplicaciones informáticas específicamente desarrolladas para el proyecto

AMPER, por lo que pudimos trabajar con los archivos que contienen los valores medios para cada una de las tres repeticiones de cada frase (9 archivos para las oraciones declarativas y otros 9 para las oraciones interrogativas). En total, pues, se analizaron 18 archivos de sonido[1].

3. RESULTADOS DEL ANÁLISIS ACÚSTICO

3.1. Análisis de F0

A continuación presentamos un gráfico generado por el programa MatLab con la curva tonal que resultó de la media de las realizaciones declarativas e interrogativas de la frase con los tres acentos en sílaba llana (ver figura 1).

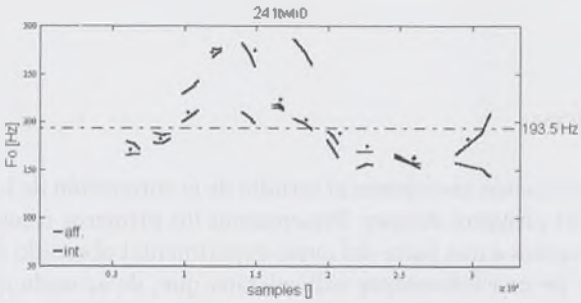


Figura 1 La guitarra se toca con paciencia

Según observamos en la figura 1, la representación de la curva de la informante de Valladolid se caracteriza por dos elevados picos tonales que vienen a coincidir con las dos vocales postónicas del pretonema: *La guitarrA se tocA con paciencia.* En la interrogativa observamos una línea tonal más elevada: la máxima elevación se hace entre la primera postónica, la vocal de la sílaba siguiente y la segunda tónica: *¿La guitarrA sE tOca con paciencia?* A partir de ahí se produce el descenso, que llega hasta la tónica final.

Una vez obtenidas todas las medias que proporciona el programa MatLab, procedimos a la tarea de estandarización (o normalización) de las curvas de F0 teniendo en cuenta tres tipos de combinaciones de acentos tonales correspondientes a los acentos inicial y final de cada frase: palabras agudas (*El saxofón se toca con obsesión*), palabras llanas (*La guitarra se toca con paciencia*) y palabras esdrújulas (*La cítara se toca con pánico*). Para ello tuvimos en cuenta en cada frase los valores centrales en herzios de cada vocal, así como el valor final de la última vocal de la frase. Resultaron así 12 valores de F0: los co-

1 Como referencia para nuestra investigación nos hemos servido principalmente de los siguientes escudios: Carrera Sabaté, *et alii* (2004), Dorta y Hernández (2004), Fernández Planas, *et alii* (2004) y Martínez Celdrán, *et alii* (2004).

rrespondientes a las 11 vocales de cada frase más el valor final de la undécima vocal (que corresponde a la denominada *juntura* terminal).

Para la estandarización de las curvas nos servimos del cálculo de porcentajes desde el valor inicial de F0[2]. En la tabla 1 presentamos los valores en herzios y en porcentajes de la curva tonal de las oraciones declarativas de la informante de Valladolid:

Sílabas	Agudas	Llanas	Esdrújulas
01. El/La/La	176 (0)	178 (0)	174 (0)
02. sa/gui/cí	179 (1,7)	186 (4,4)	209 (20,1)
03. xo/ta/ta	187 (6,2)	220 (23,5)	232 (33,3)
04. fón/rra/ra	218 (23,8)	244 (37)	224 (28,7)
05. se	215 (22,5)	195 (10,1)	178 (2,2)
06. to	205 (16,4)	206 (15,7)	200 (14,9)
07. ca	225 (27,8)	239 (34,2)	217 (24,7)
08. con	180 (2,2)	172 (-3,3)	168 (-3,4)
09. ob/pa/pá	170 (-3,4)	158 (-11.2)	155 (-10,9)
10. se/cien/ni	150 (-14,7)	161 (-9.5)	146 (-16,1)
11. sión/cia/co	157 (-10,7)	151 (-15,1)	131 (-24,7)
12. juntura final	154 (-12,5)	144 (-19,1)	125 (-28,1)

Tabla 1. Valores en herzios y en porcentajes de la curva tonal de las oraciones declarativas

Las curvas tonales que resultan son las siguientes (figura 2):

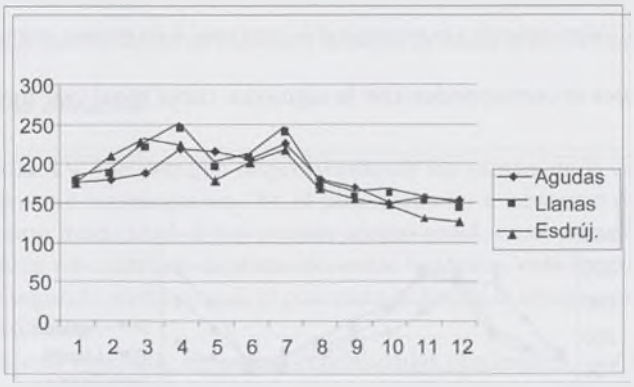


Figura 2: Curva tonal de las oraciones declarativas

2 Para el cálculo de porcentajes en la normalización (o estilización) de curvas de F0 nos servimos de la fórmula propuesta por Martínez Celdrán (2003: 126): Vest = [(Xn-Xi) x 100] / Xi, donde Vest es el valor estilizado, Xi es el valor inicial y Xn es cada uno de los valores de la frase. Véase también Cantero (2002: 147-153).

En este gráfico se observa que, mientras que en las frases con primer acento tonal en palabra llana o esdrújula el tono más alto corresponde a la postónica, en la aguda viene a coincidir el tono más alto con la sílaba acentuada. En el tonema, las agudas alcanzan el punto tonal más bajo en la pretónica, con una ligera subida en el valor final de la tónica y un nuevo descenso en la juntura terminal. Algo similar sucede con las llanas, cuya vocal tónica coincide con un leve ascenso tonal, que se continúa en la postónica con una nueva caída. Sin embargo, en las esdrújulas la caída iniciada en la tónica se va acentuando hasta llegar al final.

A continuación presentamos los valores en herzios y en porcentajes de las vocales de las oraciones interrogativas (ver tabla 2):

Sílabas	Agudas	Llanas	Esdrújulas
01. El/La/La	167 (0)	177 (0)	170 (0)
02. sa/gui/cí	183 (9,5)	181 (2,2)	226 (28,8)
03. xo/ta/ta	202 (20,9)	208 (17,5)	285 (67,6)
04. fón/rra/ra	286 (71,2)	289 (63,2)	284 (67)
05. se	301 (80,2)	290 (63,8)	240 (41,1)
06. to	241 (44,3)	241 (36,1)	210 (23,5)
07. ca	209 (25,1)	220 (24,2)	198 (16,4)
08. con	188 (12,5)	193 (9)	176 (3,5)
09. ob/pa/pá	178 (6,5)	166 (-6,2)	161 (-5,2)
10. se/cien/ni	167 (0)	157 (-11,2)	161(-5,2)
11. sión/cia/co	169 (1,1)	210 (18,6)	252 (48,2)
12. juntura final	191 (14,3)	249 (40,6)	299 (75,8)

Tabla 2. Valores en herzios y en porcentajes de la curva tonal de las oraciones interrogativas

Estos valores se corresponden con la siguiente curva tonal (ver figura 3):

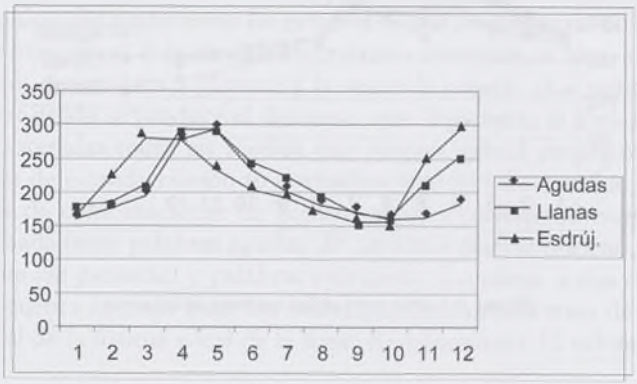


Figura 3: Curva tonal de las oraciones interrogativas

En este gráfico observamos que, en el pretonema, las esdrújulas alcanzan el pico tonal en la postónica, lo mantienen en la sílaba siguiente, y, a partir de ahí, inician un descenso que continúa en el tonema, donde, en la vocal postónica, se produce una marcada subida que se prolonga hasta un valor máximo final. Las agudas y las llanas muestran una pauta similar entre si, con la diferencia de que, en las agudas, la cima tonal prácticamente se alcanza en la tónica, y, en las llanas, esto se da en las postónicas. En el tonema, la subida comienza muy lentamente en la pretónica y se acentúa en la tónica en el caso de las agudas, mientras que, en las llanas, el ascenso se produce en la tónica, y se acentúa notablemente en la postónica (5 semitonos). En las esdrújulas, la subida se da abruptamente entre la primera postónica y la segunda (7,7 semitonos), y continúa hasta la juntura final (2,9 semitonos).

En la tabla 3 se recogen los valores en semitonos de las dos elevaciones tonales de las oraciones interrogativas. Como se puede ver en él, la máxima elevación en el pretonema se da en las frases con el primer acento tonal en palabra aguda (10,1 semitonos), mientras que la altura que se registra en las frases con acento tonal en palabra llana o en palabra esdrújula es muy similar (8,5 y 8,8 semitonos, respectivamente). En el tonema, en cambio, los valores van es ascenso: la subida es muy leve en las frases terminadas en palabra aguda (2,3 semitonos), a diferencia de lo que sucede en las llanas (7,9 semitonos) y en las esdrújulas (10,7 semitonos).

	Agudas	Llanas	Esdrújulas
Pretonema	10,1	8,5	8,8
Tonema	2,3	7,9	10,7

Tabla 3. Valores en semitonos del pico del pretonema y del ascenso del tonema en las oraciones interrogativas

3.2. Análisis de la duración

En las tablas 4 y 5 se recogen respectivamente los valores de la duración para las oraciones declarativas e interrogativas. En el pretonema no se observan diferencias sensibles entre una y otra modalidad. En el primer acento tonal, en el caso de las llanas y las esdrújulas, se marca un contraste de duración entre la tónica, más larga, y la postónica, más breve. En el segundo acento tonal, el contraste se hace a la inversa: la vocal de mayor duración es la postónica.

El tonema está marcado, en ambas modalidades oracionales, por un aumento de la duración, que, en el caso de las palabras llanas y esdrújulas, se hace más notable en la vocal final. Las interrogativas se caracterizan por un significativo aumento de la duración en la última vocal de la frase, que llega a alcanzar más del doble de la longitud de la vocal precedente (casi el triple, en el caso de las agudas).

	1	2	3	4	5	6	7	8	9	10	11
Agud.	49	58	48	46	43	43	78	40	45	51	57
Llan.	83	69	70	36	69	50	88	41	66	119	125
Esdr.	76	80	44	54	65	49	73	42	59	83	104

Tabla 4. Valores de la duración en las oraciones declarativas (en ms)

	1	2	3	4	5	6	7	8	9	10	11
Agud.	40	61	50	53	55	49	85	52	41	58	162
Llan.	62	77	65	35	61	46	92	47	62	83	167
Esdr.	53	77	45	36	55	55	79	42	52	75	177

Tabla 5. Valores de la duración en las oraciones interrogativas (en ms)

3.3. Análisis de la intensidad

De manera análoga al análisis que hemos hecho de la duración, presentamos en las tablas 6 y 7 los valores de la intensidad en las oraciones declarativas y en las interrogativas. Como se puede apreciar en dichas tablas, no hay variaciones importantes en el paso de una modalidad a la otra. Se observa, eso sí, que no todas las vocales tónicas presentan mayor intensidad que las vocales adyacentes, lo cual es un dato importante a la hora de considerar el papel de la intensidad como componente del acento en español.

Al contrario de lo que sucede con respecto a la duración, en el tonema de las declarativas se registra un descenso de la intensidad en las vocales postónicas. Aparentemente, este descenso no se da en las interrogativas; incluso, en el caso de las esdrújulas la intensidad aumenta en las postónicas.

	1	2	3	4	5	6	7	8	9	10	11
Agud.	103	104	102	99	103	101	100	101	103	100	105
Llan.	103	102	103	102	102	103	98	102	98	108	92
Esdr.	104	102	100	98	101	102	101	103	103	98	94

Tabla 6. Valores de la intensidad en las oraciones declarativas (en dB)

	1	2	3	4	5	6	7	8	9	10	11
Agud.	100	100	103	102	103	100	100	101	99	104	103
Llan.	101	102	101	101	103	98	100	100	100	103	102
Esdr.	102	102	103	102	99	101	101	101	99	105	104

Tabla 7. Valores de la intensidad en las oraciones interrogativas (en dB)

4. EL EXPERIMENTO DE PERCEPCIÓN

Se sometió a un experimento de percepción a 18 estudiantes vallisoletanos (9 informantes masculinos y 9 femeninos) y a 15 (6 masculinos y 9 femeninos) de otras procedencias. Todos ellos eran alumnos de Primer Curso de la carrera de Periodismo en la Facultad de Filosofía y Letras de la Universidad de Valladolid, con una edad comprendida entre los 18 y los 23 años. A estos estudiantes se les hizo escuchar 6 archivos de sonido sin contenido léxico-semántico con la melodía de 3 frases declarativas y otras 3 interrogativas, distribuidas aleatoriamente, de la informante vallisoletana.

Se trataba de frases con los dos primeros acentos tonales en palabra llana (*La guitarra se toca...*) y variación en el tonema (palabra aguda, *con obsesión*, palabra llana, *con paciencia*, y palabra esdrújula, *con pánico*). A continuación, los estudiantes escucharon otros 6 archivos de sonido que contenían la melodía de 6 frases leídas por una informante salmantina, también con los dos primeros acentos tonales en palabra llana y variación en el tonema[3]. Después de que escucharan dos veces las grabaciones, se pidió a los estudiantes que señalaran en cada caso si se trataba de una afirmación (frase declarativa) o una pregunta (frase interrogativa).

Los resultados fueron los que se recogen en las tablas 8 y 9. Del total de estudiantes vallisoletanos, 12 acertaron todas las respuestas; 3 fallaron una respuesta y otros 3 se equivocaron en 2 respuestas. De los archivos de sonido procedentes de la informante de Valladolid hubo 3 fallos a la hora de reconocer la frase número 2 (*¿La guitarra se toca con obsesión?*) y un fallo en la frase número 6 (*La guitarra se toca con obsesión*), mientras que, de los archivos de la informante de Salamanca, hubo 4 fallos en el reconocimiento de la frase 8 (*¿La guitarra se toca con obsesión?*) y uno en la frase 12 (*¿La guitarra se toca con obsesión?*). Se observa, por lo tanto, una acusada tendencia a confundir la modalidad de las oraciones interrogativas con tonema agudo, que son, como hemos visto, las que menor ascenso tonal presentan en las preguntas[4].

	Vallisoletanos	No vallisoletanos
Nº de aciertos	N = 18	N = 15
12 (100%)	12 (66,6%)	8 (53,3%)
11 (91,6%)	3 (16,6%)	5 (33,3%)
10 (83,3%)	3 (16,6%)	2 (13,3%)

Tabla 8. Número de aciertos en el experimento de percepción (frases declarativas e interrogativas)

3 Véase a este respecto nuestro trabajo Zamora, *et al.* (2005).

4 De los 8 estudiantes no vallisoletanos que no tuvieron ningún fallo, 4 provenían de la misma Comunidad de Castilla y León –Palencia (2), Ciudad Rodrigo (Salamanca) y Miranda de Ebro (Burgos)–, y los otros cuatro, de las siguientes Comunidades Autónomas: Andalucía (Cádiz), Cantabria (San Martín de Toranzo), Cataluña (Cornella de Llobregat, Barcelona) y Galicia (A Coruña). Los que tuvieron un fallo procedían de las dos siguientes Comunidades: Castilla y León (Aranda de Duero, Burgos, Palencia y Olmedo, provincia de Valladolid) y de Asturias (Avilés), mientras que los que tuvieron dos fallos eran castellano-leoneses (de Zamora y de Guardo, provincia de Palencia, respectivamente).

Frases	Vallisoletanos	No vallisoletanos
01. Interrogativa VA		
02. Interrogativa VA	3	3
03. Declarativa VA		
04. Declarativa VA		3
05. Interrogativa VA		
06. Declarativa VA	1	
07. Interrogativa SA		
08. Interrogativa SA	4	3
09. Declarativa SA		
10. Declarativa SA		
11. Interrogativa SA		
12. Declarativa SA	1	

Tabla 9. Número de fallos en el experimento de percepción (frases declarativas e interrogativas)

Entre los no vallisoletanos, la mayor parte de los fallos se dio de nuevo en las oraciones con tonema en palabra aguda (3 confusiones de declarativas y otras 3 de interrogativas). Sin embargo, en este grupo de estudiantes hubo también 3 casos de confusión de la modalidad de la oración declarativa *La guitarra se toca con paciencia* en la grabación procedente de la informante de Valladolid.

También realizamos, con los mismos estudiantes, un test de percepción para evaluar en qué medida dichos estudiantes podían reconocer la melodía correspondiente a la entonación de Valladolid. Para ello les hicimos oír, en orden aleatorio, una docena de archivos de sonido, con 6 frases de la informante de Valladolid y otras 6 de la informante de Salamanca. Había tres posibles respuestas: entonación vallisoletana, entonación no vallisoletana y una tercera respuesta en que el sujeto podía manifestar que no estaba seguro.

En la tabla 10 se recogen los aciertos tanto de los estudiantes vallisoletanos como de los de otras procedencias. En ella se puede apreciar que los primeros tuvieron como mínimo un 50% de aciertos y, al menos la mitad (es decir, nueve) reconocieron nueve de las doce frases (es decir, un 75%). Los no vallisoletanos, en cambio, la mayoría de ellos con unos pocos meses de residencia en la ciudad, respondieron de acuerdo con sus propias capacidades auditivas, y tuvieron por ello un porcentaje de aciertos muy variable, aunque con predominio de los que se situaban por debajo del 50%.

	Vallisoletanos	No vallisoletanos
Nº de aciertos	N = 18	N = 15
12	2	
11	1	
10	3	1
09	3	1
08	4	2
07	3	3
06	2	
05		3
04		1
03		1
02		2
01		1

Tabla 10. Número de aciertos en el test de reconocimiento de la entonación de Valladolid

5. DISCUSIÓN DE LOS RESULTADOS ACÚSTICOS Y PERCEPTIVOS

En el análisis acústico de las curvas tonales de la informante de Valladolid observamos que las oraciones declarativas y las interrogativas arrancan prácticamente a la misma altura. Sin embargo, las primeras se elevan más lentamente hasta alcanzar un primer pico tonal, que en las agudas coincide con la vocal tónica y en las llanas y esdrújulas con la postónica. Este primer pico se presenta a 3,7 semitonos en las agudas, 5,4 semitonos en las llanas y 4,3 semitonos en las esdrújulas. El segundo pico viene a coincidir con la postónica del segundo acento tonal, y se eleva, poco más o menos, a una distancia similar al primero, sin que aparentemente haya declinación. En el tonema de las declarativas, la caída del tono en las agudas es sensiblemente menor que en las esdrújulas (los porcentajes con respecto al valor de F0 inicial son, respectivamente,-12,5 y -28,1).

En las interrogativas registramos un solo pico tonal, que en las agudas se da en la tónica y se mantiene en la vocal siguiente. En las llanas y esdrújulas este pico se da en la postónica, y se mantiene en la sílaba siguiente. La altura de de esta prominencia es en las agudas de 10,1 semitonos (más de 6 semitonos que en la declarativa correspondiente); de 8,5 en las llanas (unos 3 semitonos sobre la declarativa), y de 8,8 en las esdrújulas (unos 4,5 semitonos más que en la declarativa). En el tonema, son las esdrújulas las que logran el ascenso más pronunciado (10,7 semitonos), frente a las llanas (7,9) y a las agudas, que presentan una elevación muy pequeña (2,3 semitonos).

El parámetro de la duración se aprecia, sobre todo, en el tonema tanto de declarativas como de interrogativas: en esa posición, la duración de las vocales aumenta. A este respecto, es llamativo el alargamiento de la tónica final en las interrogativas agudas, así como de la postónica final de las llanas y la segunda postónica de las esdrújulas. Por el contrario, la intensidad decrece en la postónica de las declarativas terminadas en palabra

llana o en la segunda postónica de las terminadas en esdrújula. En las interrogativas, esta disminución es mucho menor, y no llega a darse en el caso de las palabras esdrújulas.

Por último, el test de percepción probó que los informantes de Valladolid podían distinguir perfectamente la modalidad interrogativa de la declarativa, con la excepción de las interrogativas terminadas en tonema agudo. Ello puede explicarse por lo reducido del ascenso tonal de dichos tonemas, lo cual motivaría que los informantes jueces no llegaran a percibir, en las frases en cuestión, la marca de la modalidad interrogativa. En un segundo test de percepción, los informantes vallisoletanos supieron reconocer, en una proporción razonable, la entonación de Valladolid como diferente de la de Salamanca. Demostraron así su pertenencia a una misma comunidad de habla, a diferencia de lo que sucedió con los informantes de otras procedencias, quienes, en su mayor parte, no fueron capaces de reconocer tan nítidamente la entonación de Valladolid.

6. CONCLUSIÓN

Hasta aquí hemos abordado el estudio de la entonación urbana de Valladolid. Dadas las características socioculturales de nuestra informante, es de esperar que los rasgos prosódicos que hemos analizado sirvan para trazar el perfil de un "acento" vallisoletano que diferenciaría el habla de esta ciudad de la entonación del español estándar.

A la espera de poder comparar los datos de la prosodia de Valladolid con los de otras ciudades castellano-leonesas, podemos adelantar aquí, no sin ciertas reservas todavía, algunos rasgos caracterizadores, como el inicio de la curva melódica, tanto de declarativas como de interrogativas, en un tono muy bajo (el F0 no alcanzaría los 180 hz en el habla femenina), al igual que sucedería en la terminación de las declarativas, o la aparición de un pico tonal bastante elevado en el pretonema de las interrogativas, frente a un tonema, en este tipo de oraciones, que no llega a elevarse a una gran altura (sobre todo, en las terminaciones en palabra aguda).

En todo caso, como se ha demostrado en el test de pecepción, los hablantes vallisoletanos son capaces de distinguir estas particularidades prosódicas, y diferenciar el "acento" de Valladolid del de una ciudad próxima, como es Salamanca.

REFERENCIAS BIBLIOGRÁFICAS

CANTERO, F. J.: *Teoría y análisis de la entonación*, Barcelona, Edicions Universitat de Barcelona, 2002.

CARRERA SABATÉ, J.; C. VAN OOSTERZEE; A. M. FERNÁNDEZ PLANAS; L. ROMERA BARRIOS; J. ESPUNY MONSERRAT y E. MARTÍNEZ CELDRÁN (2004): «Les interrogatives al tortosí i al lleidatà. Un element diferenciador de subdialectes», *Estudios de Fonética Experimental*, XIII, 2004, pp. 157-179.

DORTA, J. y B. HERNÁNDEZ (2004): «Prosodia de las oraciones SVO declarativas e interrogativas en el español de Tenerife», *Estudios de Fonética Experimental*, XIII, 2004, pp. 225-273.

FERNÁNDEZ PLANAS, A. M.; E. MARTÍNEZ CELDRÁN; J. CARRERA SABATÉ; C. VAN OOSTERZEE; V. SALCIOLI GUIDI; J. CASTELLVÍ VIVES y D. SZMIDT SIERYKOW: «Interrogatives absolutes al barceloní i al tarragoní (estudi contrastiu)», *Estudios de Fonética Experimental*, XIII, 2004, pp. 129-155.

MARTÍNEZ CELDRÁN, E.: *El sonido en la comunicación humana. Introducción a la fonética*, Barcelona, Octaedro, 2003, 2ª edición aumentada.

MARTÍNEZ CELDRÁN, E.; A. M. FERNÁNDEZ PLANAS; V. SALCIOLI GUIDI; J. CARRERA SABATÉ Y J. ESPUNY MONSERRAT: «Approche à la prosodie du dialecte de Barcelona», *Géolinguistique*, hors série n. 3, 2005, pp. 153-175.

ZAMORA SALAMANCA, F. J.; M. CARRERA DE LA RED y M. M. MELÉNDEZ MATÍAS: «Variación geoprosódica en dos entonaciones de Castilla y León: análisis de frases declarativas e interrogativas sin expansión», *Estudios de Fonética Experimental* XIV, 2005, pp. 125-139.

Este libro, *III Congreso Internacional de Fonética Experimental*, rematouse de imprentar no obradoiro gráfico de Litonor en Santiago de Compostela, o día 21 de xullo de 2007, Día da Música

COLECCIÓN DE DIFUSIÓN CULTURAL

ÚLTIMOS TÍTULOS PUBLICADOS

CONGRESO CELSO EMILIO FERREIRO

UN TESOURO INVISIBLE

CONGRESO JOSÉ CANALEJAS E O SEU TEMPO

CONGRESO LORENZO VARELA

SIMPOSIO SOBRE ANTONIO COUCEIRO FREIJOMIL

CONGRESO PLÁCIDO CASTRO E O SEU TEMPO

ENCONTRO RAMÓN PIÑEIRO

HOMENAXE A ANTÓN ALONSO RÍOS E RAMÓN VALENZUELA

III CONGRESO DA SOCIEDADE ESPAÑOLA DE ACÚSTICA FORENSE

ISBN 978-84-453-4451-4

XUNTA DE GALICIA
CONSELLERÍA DE CULTURA
E DEPORTE
Dirección Xeral de Creación
e Difusión Cultural

52

III CONGRESO INTERNACIONAL DE FONÉTICA EXPERIEMENTAL