

Ado
izpr997

WOLIX

CADERNOS DE LINGUA

Actas do:
**SIMPOSIO DE LEXICOGRAFÍA ACTUAL:
ELABORACIÓN DE DICCIONARIOS**

ORGANIZADO POLA
REAL ACADEMIA GALEGA
E O CENTRO DE INVESTIGACIÓNS
LINGÜÍSTICAS E LITERARIAS
"RAMÓN PIÑEIRO"

Santiago de Compostela e A Coruña,
9-11 de outubro de 1995

CONSTANTINO GARCÍA
ISABEL GONZÁLEZ FERNÁNDEZ
MANUEL GONZÁLEZ GONZÁLEZ

EDITORES

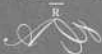
A N E X O 3

WOLIX

REAL
ACADEMIA
GALEGA



R E A L
A C A D E M I A
G A L E G A



R(05)/3

Actas do:

**SIMPOSIO DE LEXICOGRAFÍA ACTUAL:
ELABORACIÓN DE DICCIONARIOS**

ORGANIZADO POLA
REAL ACADEMIA GALEGA
E O CENTRO DE INVESTIGACIÓNS
LINGÜÍSTICAS E LITERARIAS
"RAMÓN PIÑEIRO"

Santiago de Compostela e A Coruña,
9-11 de outubro de 1995

R.123



CADERNOS DE LINGUA

Actas do:

SIMPOSIO DE LEXICOGRAFÍA ACTUAL: ELABORACIÓN DE DICCIONARIOS

ORGANIZADO POLA
REAL ACADEMIA GALEGA
E O CENTRO DE INVESTIGACIÓNS
LINGÜÍSTICAS E LITERARIAS
"RAMÓN PIÑEIRO"

Santiago de Compostela e A Coruña,
9-11 de outubro de 1995

CONSTANTINO GARCÍA
ISABEL GONZÁLEZ FERNÁNDEZ
MANUEL GONZÁLEZ GONZÁLEZ

EDITORES

A N E X O . 3

REAL
ACADEMIA
GALEGA



Director: Constantino García
Vicedirector: Manuel González González
Coordinador: Xosé Luís Regueira
Secretaria: María do Carme Pazos Balado

Coordinadores de área:

Carlos Díaz Abraira, Xermán García Cancela, Antón López Dobao,
Miguel Pérez Pereira, Modesto A. Rodríguez Neira

Colaboracións e correspondencia:

Orixinais para publicación e correspondencia en xeral enviaranse ó
coordinador:

Xosé Luís Regueira Fernández
r/. Hórreo, 31-2º
15702 Santiago de Compostela

Subscripción e intercambio: Dirixirse a
Editorial Galaxia, S. A.
r/. Reconquista, 1 - 36201 Vigo

© Real Academia Galega
A Coruña, 1990

Edita: Real Academia Galega
r/. Tabernas, 11 - 15001 A Coruña

I.S.S.N.: 1130-5924

Título clave: Cadernos de Lingua
Título abreviado: Cad. Ling.

Depósito Legal: C-1.642-1992

Deseño: M. Janeiro

Imprime: Obradoiro Gráfico, S. L.
Avda. Rebullón, 49 - Mos

Esta publicación é posible gracias á colaboración económica da Consellería de
Cultura da Xunta de Galicia.

LIMIAR

Os días 9, 10 e 11 do mes de outubro de 1995 tivo lugar en Santiago de Compostela e na Coruña o *Simposio de lexicografía actual: elaboración de dicionarios*, que reuniu un grupo dos mellores especialistas actuais en temas lexicográficos.

Aínda que había algún antecedente interesante, como son os casos do Bacharel Olea ou, posteriormente, de Sarmiento e Sobreira, a lexicografía galega en forma de dicionario comeza no século XIX, coas obras de Francisco Javier Rodríguez, Juan Cuveiro Piñol e Marcial Valladares. Os tres foron autores de dicionarios bilingües galego-casteláns, que tiñan un gran cariño pola lingua galega, pero que carecían da formación suficiente para levar a cabo a súa empresa co rigor esixible no seu tempo. Dentro do mesmo ámbito da lexicografía bilingüe galego-castelá cabe salientar, xa no século XX, o labor realizado pola Real Academia Galega (co seu dicionario inconcluso publicado entre os anos 1913-1928), por Filgueira Valverde-Tobío Fernández-Magariños Negreira e Cordal Carús, por Leandro Carré Alvarellos, por Xosé Luís Franco-Grande e, sobre todo, por Eladio Rodríguez González. Tamén houbo algún achegamento no sentido inverso castelán-galego; tal é o caso do Dicionario das Irmandades da Fala, e das dúas obras de José Santiago Crespo Pozo.

A situación cambia radicalmente a partir dos anos 80, coa aparición de dicionarios monolingües de certa madurez, sobre todo coa publicación do *Diccionario Xerais da lingua* (Vigo, 1986), o *Diccionario da lingua galega* da editorial Ir Indo (Vigo, 1986), o *Diccionario da lingua galega* do Instituto da Lingua Galega e Real Academia Galega (A Coruña, 1990), o *Diccionario da lingua galega* de Obradoiro-Santillana (Vigo, 1995), e algún outro. Tamén dentro da lexicografía bilingüe cabe salientalo *Diccionario normativo galego-castelán* da editorial Galaxia (Vigo, 1988) e o *Diccionario Xerais castelán-galego* (Vigo, 1990). Todas elas son obras dunha altura lexicográfica aceptable. Se grande foi o camiño percorrido nos últimos anos na lexicografía galega, tamén son moitos os problemas aínda por resolver.

Neste momento o Seminario de Lexicografía da RAG ten na imprenta un novo dicionario galego, e está comezando xa os traballos preparatorios doutra versión máis ampla desta obra, que conterà

entre 45000 e 50000 entradas. No mesmo seminario estase a traballar nun dicionario bilingüe castelán-galego e noutro italiano-galego.

Paralelamente no Centro de Investigacións Lingüísticas e Literarias "Ramón Piñeiro" estase deseñando un modelo teórico da estrutura dun dicionario, que sexa aplicable a un conxunto de dicionarios bilingües de diversas linguas que terán o galego como lingua de partida ou de chegada.

Pareceunos que o momento era suficientemente importante como para sentarse a reflexionar durante uns días sobre os problemas que presenta a realización de dicionarios tanto monolingües coma bilingües. Por iso convocamos un grupo de lexicógrafos de contrastada valía e experiencia, que puidesen cos seus coñecementos axudar a achanda-las dificultades e orientar con bases sólidas os proxectos en marcha.

Era necesario coñece-lo avance das realizacións lexicográficas nos últimos anos, era necesario reflexionar sobre a importancia dos corpus textuais para a elaboración de dicionarios e sobre o seu deseño óptimo, cumpría examina-las necesidades que debe cubrir cada tipo de dicionario, sentíase a urxencia dunha discusión sobre a macroestructura e a microestructura ideal do dicionario de lingua: ¿como se debe trata-lo problema da homonimia?, ¿cal é a sintaxe máis axeitada da definición lexicográfica?, ¿que función deben cumprir os exemplos no dicionario?, ¿como se debe trata-la gramática?, ¿como convén representa-lo nivel de lingua?, etc., etc. Estes problemas, entre outros, foron amplamente tratados en relatorios encargados a especialistas e debatidos en mesas redondas, que sempre resultaron animadas e estimulantes.

Foron tres densas xornadas dun labor intenso, pero gratificante, e desde estas páxinas queremos agradecer a tódolos participantes o seu entusiasmo e a súa entrega.

A primeira sesión (luns, día 9 de outubro) tivo lugar no Centro de Investigacións Lingüísticas e Literarias "Ramón Piñeiro". Despois da apertura do Simposio cunhas palabras de D. Manuel Regueiro Tenreiro, Director Xeral de Política Lingüística da Consellería de Educación da Xunta de Galicia, Isabel González deu lectura á conferencia de R. Simone, profesor na Terza Università degli Studi de Roma, quen por problemas xurdidos a última hora non puido asistir, e que versou sobre os *Verbi sintagmatici*. Seguiu logo a disertación de Manuel Seco, da Real Academia Española, sobre a *Microestructura del Dicionario del Español Actual*, e cerrou a sesión da mañá a investigadora J. Rey-Debove, que falou sobre *Les domaines respectifs de l'exemple et de la citation dans les dictionnaires de langue actuels*.

Pola tarde tivo lugar unha mesa redonda sobre os *diccionarios monolingües*, moderada por Constantino García, Director do Seminario de Lexicografía da Real Academia Galega. Nela discutíronse algúns dos problemas que máis preocupan na confección dun diccionario monolingüe, entre os que cabe salientar: os criterios que deben primar para a distinción de voces homónimas dentro do diccionario (etimolóxico, conciencia lingüística, relacións sintagmáticas...); a información que debe leva-lo artigo dun diccionario monolingüe; a necesidade de marca-lo contorno na definición; etc.

A segunda sesión (martes, día 10) tivo lugar na sede da Real Academia Galega, na Coruña. Despois dunhas palabras de recibimento de D. Domingo García Sabell, Presidente da Real Academia Galega, comezou a xornada de traballo cunha conferencia de P. Beltrami, da Opera del vocabolario italiano de Florencia, quen falou sobre *Il Tesoro della lingua italiana delle Origini*; seguiu a conferencia de M. Alvar Ezquerra, catedrático da Universidade de Málaga, sobre *Cincuenta años de diccionarios Vox monolingües del español*; e pechou a sesión da mañá Jacqueline Picoche, Profesora Emérita da Universidade de Picardía, falando sobre a *Lexicographie théorique et appliquée: un dictionnaire des mots de haute fréquence*.

Pola tarde, antes da mesa redonda, fixo unha presentación do *Diccionario da Lingua Galega* da Real Academia Galega Manuel González González, membro da Real Academia Galega e director da obra. A continuación Isabel González, catedrática de Filoloxía italiana da Universidade de Santiago, presentou outras dúas obras que se están a realizar no Seminario de Lexicografía da RAG: o *Diccionario castelán-galego* e o *Diccionario italiano-galego* que está dirixindo.

De seguido tivo lugar unha mesa redonda, moderada por Isabel González, sobre *diccionarios bilingües*, na que se discutiron, entre outros temas, as peculiaridades que presenta un diccionario bilingüe fronte a un monolingüe; a selección de entradas; a estruturación da información; a discriminación de significados; os delimitadores semánticos; a ordenación das equivalencias de tradución; a función do exemplo; os indicadores sobre o nivel de lingua; a conveniencia de utiliza-la lingua de partida ou a de chegada nas explicacións, segundo a finalidade do diccionario.

O mércores 11, de novo no C.I.L.L. "Ramón Piñeiro", P. Beltrami fixo un resumo dos traballos que se están a facer no Ufficio filologico dell'Opera del Vocabolario Italiano; despois tomou a palabra Joaquim Rafel, secretario xeral do Institut d'Estudis Catalans, quen falou sobre o tema *Diccionarios y corpus textuales. Perspectivas para el catalán*; e pechou a sesión I. Sarasola, membro da Real Aca-

demia Vasca, cunha intervención sobre o tema *Lexicografía vasca: el Diccionario General Vasco*.

Tralas palabras do Director Xeral de Política Lingüística e de Constantino García, director do Seminario de Lexicografía da RAG e coordinador científico do C.I.L.L. "Ramón Piñeiro", deuse por concluído o simposio.

Esperamos que esta primeira reunión científica sobre os problemas lexicográficos non sexa senón un punto de partida, que teña continuidade nos anos vindeiros coa realización doutros simposios que nos permitan intercambiar informacións e experiencias para facer posible ofrecer ó público dicionarios cada vez máis perfeccionados, máis rigorosos e máis eficaces, que respondan en definitiva ás necesidades que demanda a sociedade actual.

O noso agradecemento á Real Academia Galega e ó Centro de Investigacións Lingüísticas e Literarias "Ramón Piñeiro", porque fixeron posible que estas xornadas se puidesen levar a cabo. Reiterámoslle-la nosa gratitude a tódolos participantes por ter respondido a esta chamada; a súa presenza redundou en beneficio de todos, e singularmente no noso propio, ó botar luz sobre moitos problemas que se nos presentaron na realización dos dicionarios que neste momento temos no tear.

Lamentámo-la ausencia de M. Salem, de Florencia, e de F. J. Hausmann, da Universidade de Erlangen-Nürnberg, que por problemas persoais non puideron estar connosco.

Aínda que R. Simone non puido estar presente, incluímos neste volume a súa conferencia, porque nola fixo chegar con anterioridade e puido ser lida.

CONSTANTINO GARCÍA
ISABEL GONZÁLEZ FERNÁNDEZ
MANUEL GONZÁLEZ GONZÁLEZ
(coordinadores do simposio)

VERBI SINTAGMATICI IN ITALIANO

RAFFAELE SIMONE
Terza Università degli studi di Roma

1. Una sottoclasse di verbi e un profilo lessematico

1.1. Riconoscere classi omogenee

Una delle preoccupazioni del linguista (e soprattutto del lessicologo e del lessicografo) dovrebbe essere quella di sapere con buone ragioni con quali classi di parole ha a che fare, e in particolare di ricercare classi omogenee per comportamento (sintattico, morfologico o semantico) entro le generali 'parti del discorso' che ci ha consegnato la tradizione. Se su questo principio tutti possono essere d'accordo, le cose vanno meno bene quando di verifichi da vicino lo stato delle nostre conoscenze su questi temi.

Per quanto riguarda l'italiano e tante altre lingue che gli sono geneticamente vicine, ad esempio, mi pare che quando si discute di lessicologia ci si accontenti normalmente di sapere che esse hanno nomi, verbi, aggettivi, e così via – cioè le classi di parole identificate dalla tradizione e poco più. Un'altra tipica assunzione implicita è che queste categorie siano costituite naturalmente da entità strettamente *monorematiche*, formate cioè da una sola 'parola'. Quindi, il vocabolario di questa lingua viene rappresentato come composto da nomi, verbi, aggettivi ecc., ciascuno dei quali è fatto esclusivamente di una parola. Su questa base sono fatti praticamente tutti i dizionari italiani pubblicati finora e molti dizionari di altre lingue.

Vista da un angolo più generale, questa negligenza si accompagna alla scarsa riflessione dei linguisti sulle 'parti del discorso', che dovrebbero essere riesaminate con un occhio moderno ma sulle quali ben poco è stato fatto (salvo isolate accezioni come Swiggers-van Hoecke 1986; Bosque 1991). Lo stesso ragionamento andrebbe fatto, oltre che dal lato superficiale, anche da quello

profondo. La semantica, e più ancora la sintassi, sta solo da poco cominciando a rendersi conto che il lessico di una lingua è organizzato secondo principi complessi. Per fare un esempio, il fatto che i ruoli tematici possano essere assegnati non solo dai verbi ma anche dai nomi (e non esclusivamente dai nomi de-verbali), già chiaro alla linguistica *fin de siècle*, ritorna a essere discusso di recente, ma non mi pare che se ne siano tratte le conseguenze appropriate sul piano della riflessione generale sui fondamenti semiotici del linguaggio – e quindi tantomeno su quello lessicologico e lessicografico.

È stato mostrato in modo convincente (ultimamente in Voghera 1994) che il modo di ragionare in termini di classi tradizionali di parole e di concepire queste come costituite da una sola sequenza di segmenti isolati tra due spazi vuoti, è insufficiente, perché non riflette né la reale organizzazione della lingua né la competenza che il parlante ne ha. In questo modo si trascura completamente una varietà di fenomeni non immediatamente visibili ma di grande rilievo: tra questi il fatto che in varie lingue (e anche in italiano) esistono intere classi di parole sintagmatiche, formate cioè la più parole che non possono occorrere l'una senza l'altra e che hanno nel loro insieme il medesimo comportamento sintattico. Queste parole sintagmatiche sono state chiamate con diversi nomi, ma quello più trasparente sembra essere *polirematiche*, che useremo del seguito di questo discorso.

Le parole polirematiche hanno proprietà specifiche. Intanto, si comportano in maniera diversa dalle parole monorematiche di cui si compongono (in particolare dalle loro parole-testa) e formano una classe a sé anche nella competenza del parlante. Chi potrebbe, ad esempio, sostituire *messa in piega* con il semplice *messa*, o *sedia elettrica* con *sedia*? Inoltre, le polirematiche si ordinano a loro volta in sottoclassi ben riconoscibili, che rendono differente *ferro da stiro da farsi sotto*, o *filo da torcere da tipo da spiaggia*. Ad esempio, il grado di separabilità dei loro elementi può essere molto variato e la loro stessa architettura sintattica può cambiare di molto. Nel seguito di questo discorso chiamerò *profili* tipi di parola polirematica via via considerati.

In italiano, profili polirematici esistono in particolare nell'ambito dei nomi, dei verbi e degli avverbi (Voghera 1994: 187), ed è un problema rilevante quello di cercar di dare a questi fenomeni un trattamento lessicologico e lessicografico appropriato. I verbi, in particolare, contengono una sottoclasse ancora più specifica, che presenta caratteristiche tutte sue e che merita di essere presa in

considerazione se vogliamo davvero che il vocabolario non sia una pura lista di parole isolate, ma una macchina strutturata che rifletta il più possibile la competenza lessicale.

1.2. Verbi sintagmatici

In questo contributo propongo un esercizio lessicologico su un frammento di lessico italiano, ma con qualche ambizione di teoria generale del lessico. Vorrei infatti cercar di mostrare che i verbi e le parole polirematiche italiane formano una intersezione che è occupata da una sottoclasse di parole omogenea (per comportamento semantico, sintattico e fonologico), che chiamerò *verbi sintagmatici* (d'ora innanzi VS), e descrivere questa classe in maniera puramente fenomenologica.

Per intenderci subito, definirò i VS come sintagmi formati da una testa verbale e da un complemento costituito da una 'particella' (originariamente un avverbio), uniti da una coesione sintattica di grado elevato, al punto che non si può commutare il VS intero con una sola delle sue parti. Si tratta quindi di costruzioni del tipo di *fare fuori, venire meno o buttare giù*. Per la verità la classe dei VS italiani contiene ancora altre sottoclassi, che ugualmente attendono di essere studiate e definite: si tratta perlomeno (a) dei verbi pronominali semplici del tipo di *vederci, capirci* (*Non ci capisco nulla*) o *starci* (*Luisa non ci sta*) o *esserci e averci* (due voci cruciali dell'italiano moderno, delle quali non c'è ancora una definizione precisa), (b) dei verbi pronominali multipli come *prendersela, farsela* (*Quei due se la fanno da un pezzo*) o *intendersela*, o (c) dei verbi sintagmatico-pronominali, come *farsi accanto, farsi avanti, farsi incontro, farsi indietro, farsi sotto*, o come *farsela addosso o andarsene via*. Tutte queste categorie sono altrettante proprietà originali dell'italiano e richiederebbero una considerazione ravvicinata anche per definire più nel dettaglio la caratterizzazione tipologica della morfologia di questa lingua. Nel seguito, però, mi occuperò soltanto dei VS nella loro configurazione minima, cioè quella che ha il profilo lessematico Verbo + Avverbio.

Il termine con cui designo questa classe di verbi, VS, è ovviamente un calco dell'inglese *phrasal verbs*. Adopero questo termine perché i VS italiani formano (come cercherò di mostrare) una classe molto prossima ai *phrasal verbs* inglesi.

C'è però una differenza di trattamento e di fortuna tra i *phrasal verbs* inglesi e i loro omologhi italiani. I primi sono molto numerosi e rispondono ad un profilo lessematico tipico delle lingue

germaniche, trovando analoghi più o meno rigorosi in tedesco (verbi separabili e inseparabili, da *nachdenken* a *entstehen*) e in nederlandese (con la stessa fenomenologia del tedesco: *nadenken*)¹. Per conseguenza sono riconosciuti da tempo come una sottoclasse rilevante di forme, essendo dotati di una semantica, di una sintassi e di una fonologia di parola e di frase particolari (Quirk et al. 1972: 815). I VS italiani invece sono notevolmente meno numerosi e sembrano essere isolati nell'ambito della famiglia romanza², e, pur essendo un profilo produttivo e di notevole frequenza d'uso, sono sensibilmente meno visibili.

E in effetti, il loro diritto di cittadinanza non è riconosciuto praticamente da nessuno. Non c'è ad esempio un solo vocabolario moderno che li presenti come classe di parole autonoma e li consideri meritevoli di formare lemma a sé: i vocabolari italiani registrano sì i VS, ma li dissolvono nel lemma dedicato al verbo-testa. Anche le grammatiche si comportano in questa maniera: non ho trovato una sola grammatica italiana di vaste dimensioni che faccia parola dei VS, né sotto i capitoli morfologici né sotto quelli sintattici.

Insomma, se non è possibile negare che la categoria dei VS offra all'italiano una risorsa importante e caratterizzante, nessuno l'ha mai descritta e riconosciuta – un singolare ma non sorprendente caso di cecità alla propria lingua.

L'idea che voglio sostenere qui è che i VS costituiscono una classe diversa da quella formata dai soli verbi-testa e che quindi è indispensabile trattarla in modo autonomo e appropriato. In altre parole, *venire meno* non è in nessun senso un caso particolare di *venire*, né *buttare giù* è un caso particolare di *buttare*. Ma la mia preoccupazione non è solo lessicografica. Definendo una sottoclasse omogenea come quella dei VS, si tocca il problema generale di revisionare la nostra concezione delle 'parti del discorso', in termini sia superficiali che profondi, e quindi di riflettere da un angolo nuovo sui fondamenti della teoria del linguaggio (Simone 1990: 284 ss.). Deve esserci infatti almeno qualche buona ragione semiotica che giustifichi l'esistenza di una sottoclasse omogenea di verbi, e indagando su questa ragione semiotica possiamo forse capire un po' meglio il modo in cui si organizzano le lingue.

¹ Sui verbi con particella tedeschi, parzialmente somiglianti a quelli sintagmatici italiani descritti qui, si veda Stiebels-Wunderlich (1994).

² Considerazioni interessanti sul 'tipo romanzo' (e in particolare italiano) e il 'tipo germanico' di VS si trovano in Schwarze (1985: 357 ss.).

In aggiunta (ma non è un argomento marginale) i VS definiscono un profilo lessematico tipologicamente nuovo, piuttosto produttivo e capace di possibili sviluppi. Abbiamo infatti buone ragioni per ritenere che nell'ambito romanzo questi verbi siano una innovazione specificamente italiana. In altre lingue romanze si registrano (da rilievi informali che ho fatto) soltanto casi isolati e contati di VS: così lo spagnolo *quedarse afuera* "restare fuori" o *quedarse atrás* "restare indietro", *seguir adelante* "continuare" o *echar abajo* "buttare giù", il francese *aller dedans* "andare dentro" = "starci (in uno spazio limitato)".

I VS costituiscono anche una questione diacronica rilevante: essi sembrano essere infatti non un portato del fondo toscano dell'italiano, ma piuttosto un affioramento in italiano di un profilo lessematico dialettale di un tipo che chiamerei approssimativamente 'settentrionale'³. Per fare solo un esempio, nella parlate venete appaiono preferibilmente VS per una varietà di significati: *smontare giù* "scendere", *andare dentro* "entrare", *andare indietro* "indietreggiare", e perfino *saltare su* "contraddire protestando", *perdersi via* "distrarsi", *andare dietro* "corteggiare". I VS pongono quindi il problema di stabilire quanto l'italiano deve, dal punto di vista della sua architettura fondamentale, ai suoi dialetti e attraverso quale via questi sono riusciti ad imporre una risorsa così particolare.

2. Un corpus approssimativo

2.1. Una lista aperta

Non sono in grado di dire quando VS esistano in italiano, anche perché (come ho detto prima) mancano basi di dati appropriate. Ho già osservato che non sono tanto numerosi come in inglese e nelle lingue germaniche in generale. Nondimeno costituiscono ugualmente una classe quantitativamente rispettabile.

³ Considerazioni in Schwarze (1985: 364-365), che accenna al fatto che, in semplici test di accettabilità di espressioni con VS (come *uscire fuori* o *venire fuori*) contro espressioni con verbi semplici di significato somigliante (*uscire*), la prima forma tende ad essere preferita da parlanti di dialetti settentrionali.

Ho provato a costruirne una lista approssimativa, ed ho ottenuto la seguente:

1. andare addosso
2. andare avanti
3. andare dietro
4. andare fuori
5. andare giù
6. andare incontro
7. andare indietro
8. andare lontano
9. andare sopra
10. andare sotto
11. andare su
12. andare via
13. andare vicino
14. buttare addosso
15. buttare fuori
16. buttare giù
17. buttare via
18. correre contro
19. correre dietro
20. correre incontro
21. correre via
22. dare addosso
23. dare via
24. dire bene
25. dire male
26. entrare dentro
27. essere accanto
28. essere contro
29. essere dentro
30. essere fuori
31. essere giù
32. essere giù
33. essere senza
34. essere su
35. essere via
36. fare aposte
37. fare fuori
38. finire lontano
39. finire sotto
40. finire vicino
41. guardare avanti
42. guardare dietro
43. guardare lontano
44. lasciare dentro
45. lasciare fuori
46. mandare fuori
47. mandare giù
48. mandare su
49. mandare via
50. mettere avanti
51. mettere contro
52. mettere da parte
53. mettere dentro
54. mettere dietro
55. mettere fuori
56. mettere giù
57. mettere indietro
58. mettere sopra
59. mettere sotto
60. mettere su
61. parlare davanti
62. parlare dietro
63. passare avanti
64. passare dietro
65. passare sopra
66. passare su
67. passare via
68. pensare su
69. porre accanto
70. porre avanti
71. porre dietro
72. porre su
73. portare avanti
74. portare giù
75. portare indietro
76. portare sopra
77. portare sotto
78. portare su
79. portare via
80. portare vicino
81. prendere sotto
82. prendere su
83. remare contro
84. restare dentro
85. restare fuori
86. restare senza

- | | |
|---------------------|----------------------|
| 87. restare senza | 111. tirare dietro |
| 88. saltare dentro | 112. tirare giù |
| 89. saltare fuori | 113. tirare su |
| 90. saltare su | 114. tirare via |
| 91. saltare via | 115. togliere via |
| 92. sbattere dentro | 116. uscire fuori |
| 93. sbattere fuori | 117. uscire incontro |
| 94. scappare via | 118. venire accanto |
| 95. stare accanto | 119. venire addosso |
| 96. stare addosso | 120. venire avanti |
| 97. stare avanti | 121. venire bene |
| 98. stare contro | 122. venire fuori |
| 99. stare dietro | 123. venire giù |
| 100. stare giù | 124. venire incontro |
| 101. stare insieme | 125. venire indietro |
| 102. stare sopra | 126. venire male |
| 103. stare sotto | 127. venire meno |
| 104. stare su | 128. venire sopra |
| 105. stare vicino | 129. venire su |
| 106. strappare via | 130. venire via |
| 107. tenere dietro | 131. venire vicino |
| 108. tenere fuori | 132. volere bene |
| 109. tenere sotto | 133. volere male |
| 110. tirare avanti | |

Questa lista⁴ è, come ho detto, approssimativa, e non sarà difficile trovare altri elementi per completarla. In ogni caso essa non dovrebbe risultare composta da più di alcune decine di voci (al momento sono poco più di centotrenta). Ciò sottolinea il paradosso che ho accennato all'inizio: i VS italiani sono pochi di numero, ma svolgono un ruolo importante, per frequenza e per registro, nell'uso reale della lingua.

2.2. Proprietà *prima facie*

Dalla lista riportata sopra saltano agli occhi alcuni fatti importanti. I VS sono nella stragrande maggioranza verbi di movimento (*venire, andare*) e stativi (*stare, restare*). La lista si ottiene infatti quasi per intero da intersezioni di verbi di movimento o stativi come *andare, venire, portare, mettere*, ecc. e, rispettivamente, *stare, restare*, con avverbi indicanti posizione, distanza o direzione

⁴ Questa lista è costruita in modo molto prudentiale, anche se non tutte le voci che ne fanno parte hanno uno statuto omogeneo. Non sono sicuro, ad esempio, che *volere bene* e *volere male* abbiano la stessa natura degli altri verbi della lista.

come *su*, *giù*, *accanto*, *vicino*, *dietro*, *via*, *incontro*, ecc. Questo fenomeno è tutt'altro che sorprendente, anche se richiederebbe una spiegazione: infatti, numerose lingue che hanno verbi con particella (dal tedesco, all'inglese ecc.) utilizzano questo specifico profilo proprio per segnalare la direzione del movimento (cfr. Stiebels-Wunderlich 1994).

Su questa base si può supporre che la funzione dello specifico profilo lessematico del VS sia principalmente quella di rendere dettagliata l'informazione di verbi di movimento dal significato generico. Si può anche pensare una delle funzioni generali di VS sia quella di attribuire o enfatizzare con un quasi-affisso (l'avverbio post-verbale) un tratto di movimento nel verbo. In particolare, l'aggiunta di *via* come complemento sembra aggiungere una marca di movimento ad una varietà di verbi che in forma assoluta non hanno questo tratto:

- (1) *lavare via* "togliere via lavando" \approx *lavare*
- (2) *raschiare via* "togliere via raschiando" \approx *raschiare*
- (3) *grattare via* "mandar via grattando" \approx *sciogliere*

Non tutti i VS sono, però, verbi di movimento, anche quando il verbo-testa lo è: VS come *venire meno*, *buttare giù* nel senso di "deprimere", *stare senza*, *volere male* hanno semantiche diverse dai verbi di movimento, che si riflettono in sintassi diverse.

Si può osservare anche che questi verbi hanno qualcosa a che fare con verbi a reggenza come *cantare (su)* o *stare (su)*. Può essere utile dire qualche parola per distinguere le due classi. Mi pare che la differenza principale sia la seguente: nei verbi a reggenza propriamente intesi la particella (che è una preposizione) non contribuisce alla formazione del significato dell'insieme, o, per dirla diversamente, non assegna ruolo tematico, ma serve solo per stabilire il collegamento sintagmatico col nominale che segue. Nei VS la particella è invece essenziale per la formazione del significato. Inoltre, i verbi a reggenza, se appaiono con particella, non possono essere usati senza un SN Oggetto, mentre i VS sì. Non possiamo allora avere *Conto molto su*, ma semmai *Conto molto su di lui*; al contrario, possiamo avere *È uno che tira via* oppure *Una sigaretta tira un po' su*, senza alcun nominale dopo la particella.

3. Argomenti analitici per i VS

Presenterò ora argomenti di diversa natura per sostenere che i VS formano una classe omogenea e devono essere trattati in modo conseguente.

3.1. Argomenti semantici

Proprio come nel caso dei *phrasal verbs* inglesi, anche i VS (sia quelli di movimento che gli altri) hanno una semantica che non è sempre compositiva: il significato dell'insieme non si deduce infatti da quello dei suoi componenti. Distinguiamo comunque tre casi:

(a) quelli in cui la particella ripete o intensifica l'informazione del verbo:

- (4) uscire ≈ uscire fuori
- (5) entrare ≈ entrare dentro
- (6) scappare ≈ scappare via

È facile vedere che in questo caso è possibile anche usare il verbo testa da solo.

(b) Quelli in cui la particella aggiunge una marca circa la natura del movimento ad un verbo che indica sì un movimento ma di tipo non-marcato. In questi casi, la particella serve soprattutto a trasformare un verbo di moto generico in uno in cui è specificata la direzione del moto e in qualche caso la sua relazione rispetto all'emittente dell'enunciato:

- (7) mettere ≈ mettere su
- (8) tirare ≈ tirare via
- (9) saltare ≈ saltare fuori

(c) I VS in cui la coppia Verbo + Avverbio ha un significato del tutto imprevedibile a partire da quello delle parti. Ecco alcuni esempi:

- | | |
|------------------|---|
| (10) buttare giù | "far cadere dall'alto in basso"
"deprimere, rattristare" |
| | "scrivere (qualcosa) in modo frettoloso e approssimativo" |
| (11) buttare via | "gettare via"
"sperperare, dissipare" |
| (12) fare fuori | "eliminare, ammazzare, licenziare" |

(13) mettere dentro	"porre in un luogo chiuso" "incarcerare"
(14) passare su	"muoversi calcando su qualcosa" "ignorare, non curarsi di"
(15) portare avanti	"spostare muovendo in avanti" "sviluppare, elaborare, diffondere"
(16) stare su	"trovarsi in alto" "essere di umore positivo, essere in forma"
(17) tirare via	"strappare" "compiere un lavoro in modo approssimativo"
(18) venire bene	"riuscire, avere successo, essere esatto (di operazioni matematiche e sim.)"
(19) venire giù	"scendere" "crollare"

È ovvio che i VS della seconda e della terza classe non possono essere sostituiti dai soli verbi-testa, perché questi non basterebbero a convogliare il significato dell'insieme. In questi casi la prevedibilità semantica del VS è nulla, e la situazione è complicata ulteriormente dal fatto che spesso lo stesso VS ha insieme un valore di tipo (a) e uno di tipo (c). Ad esempio, *mettere su* in alcune varietà locali significa, oltre che "collocare qualcosa in alto", anche "aizzare, istigare" con oggetto [+umano], così come *buttare giù* significa, oltre che "gettare in basso", anche "scrivere qualcosa in modo rapido e frettoloso".

Come si vede, dal punto di vista semantico siamo quindi pienamente nel caso dei *phrasal verbs* inglesi, dove a volte la semantica si legge nella combinazione degli elementi del sintagma (*set up*), in altri invece il lemma deve avere una entrata lessicale autonoma, perché non è analizzabile. Tale è ad esempio il caso di *think out*, che varrebbe lateralmente più o meno "pensare fuori", ma significa in effetti "pensare approfonditamente", di *drop by* che vale "comparire all'improvviso", e simili.

3.2. Argomenti sintattici

Ho accennato prima che i VS manifestano una particolare coesione interna. Vale la pena di osservare che nel corpus riportato sopra si osservano diverse categorie per quanto riguarda la separabilità e la dislocabilità dei due costituenti del VS. Osserviamo gli esempi seguenti:

- (20) a *Ci ho pensato su* molto
 b *Ci ho pensato molto su*
- (21) a *Ho portato via* il bambino
 b *Ho portato subito via* il bambino
 c **Ho portato il bambino via*
 d *Ho portato il bambino via da casa*
- (22) a *Quella telefonata ha buttato giù* mio fratello
 b **Quella telefonata ha buttato mio fratello giù*
 c *Quella telefonata ha buttato mio fratello giù dal letto*
- (23) a *Gli è andato addosso*
 b *Gli è andato con la macchina addosso*
 c *È andato con la macchina addosso al fratello*
- (24) a *È venuto meno* l'appoggio dei suoi amici
 b **È venuto l'appoggio dei suoi amici meno*

Come si vede, la separabilità di questi VS è piuttosto variegata. In generale, nei VS di movimento i due costituenti sembrano poter essere separati l'uno dall'altro solo in due casi:

(a) se si interpone un costituente leggero e non argomentale (*Ho portato subito via il bambino*),

(b) se la particella del VS è a sua volta testa di un sintagma (*Quella telefonata ha buttato mio fratello giù dal letto*, dove *giù* è testa del sintagma *giù dal letto*, e quindi non è propriamente la particella del VS).

Si può quindi dire che, quando la particella non è attratta da un sintagma ulteriore, la sua separabilità del verbo-testa è bassa o nulla. Nella seconda sottoclasse (verbi, cioè, non di movimento, come *venire meno*) la separabilità sembra ancora più ridotta. Questa specificità dei VS di movimento dipende dal fatto che la particella che li integra può, con un oggetto appropriato, assegnare un ruolo tematico. Quando invece il VS è preso nel significato di tipo (c), cioè totalmente imprevedibile, non può essere integrato da nessun oggetto, e la separabilità dei due elementi si azzerava.

Un uguale grado di coesione si osserva considerando la possibilità di dislocare i costituenti dei VS. Si osservino gli esempi seguenti:

- (28) **È meno che è venuto il suo appoggio*
 (29) **È giù che è si è buttato mio fratello*
 (30) *?È senza che sono rimasto*
 (31) *?È insieme che stanno Carlo e Luisa*

- (32) ?È via che sono andati
 (33) È su che sta
 "Sta in alto", ma * "È di buon umore"

Si possono trovare esempi ulteriori, ma credo di poter dire che quando il verbo è usato sintagmaticamente la dislocabilità dei suoi elementi è bassa o bassissima.

5. Ulteriori test

Una verifica del fatto che i VS formano una sottoclasse propria si può avere sottoponendoli a qualche ulteriore test.

5.1. Nominalizzazione

Propongo anzitutto un test di nominalizzazione. I verbi 'normali' sono usualmente nominalizzabili (salvo che ragioni lessicali e morfologiche non lo impediscano):

- (34) a Che Luigi *sia venuto* a casa mia mi ha fatto piacere
 (35) b La *venuta* di Luigi a casa mia mi ha fatto piacere
 (36) a Che le si *tirino* i capelli non piace a Luisa
 (37) b Le *tirate* di capelli non piacciono a Luisa

I VS corrispondenti non possono essere nominalizzati. Si vedano gli esempi seguenti:

- (38) a Che tu *sia venuto su* è stato utile
 (39) b *La tua *venuta su* è stata utile
 (40) a Ci è dispiaciuto che ci *sia venuto meno* il suo appoggio
 (41) b *La *venuta meno* del suo appoggio ci è dispiaciuta

L'unica nominalizzazione possibile di questi verbi sembra essere ovviamente quella costituita dall'infinito nominale, una risorsa accessibile a qualsivoglia voce verbale italiana, ma sintatticamente molto particolare perché conserva la caratterizzazione argomentale del verbo:

- (42) a Che tu *tiri via* il lavoro non sta bene
 (43) b Il tuo *tirar via* il lavoro non sta bene

Il fatto singolare dell'impossibilità di nominalizzare può essere considerato indizio di una ancora insufficiente grammaticalizzazione dei VS, che operano come verbi nella configurazione completa delle loro possibilità ma non sono ancora in grado di dar luogo a nominalizzazioni accettabili.

5.2. Inserzione di un SN complemento

Una riprova ancora più forte della coesione dei VS si ha esaminando il comportamento dell'avverbio postverbale. In italiano tanto le preposizioni quanto gli avverbi degradati a preposizioni possono essere seguiti da un nominale con cui fanno sintagma:

- (44) Metti il libro [sulla tavola]
(45) Metti il libro [sotto il tavolo]

Quando sono transitivi e quindi possono avere un Oggetto, i VS si comportano in altro modo. Il nominale che segue l'avverbio del VS non è un componente del sintagma formato da Avverbio + Nome, ma è retto dall'intero VS. La struttura di una clausola come *Porta su la scala* non è quindi (a), bensì (b):

- (46) a * [Porta] [su la scala]
 b [Porta su] [la scala]

Ciò vuol dire che il SN che segue il verbo non ha sintatticamente nulla a che fare con l'avverbio che lo precede: non è il complemento di un Sprep, ma è l'Oggetto dell'intero VS. Quando il verbo è sintagmatico, insomma, dopo la sua particella c'è un confine di sintagma che non può essere scavalcato.

5.3. Comportamento fonologico

Il confine di sintagma di cui parlavo prima si osserva con una chiarezza ancora maggiore verificando il comportamento fonologico della particella. Quando questa fa sintagma col nominale che la segue, normalmente ha luogo un sandhi tra essa e l'articolo:

- (47) Siamo andati [su#la montagna] → Siamo andati [sulla montagna]

Quando il verbo è un VS è seguito da un nominale oggetto, invece, il processo di sandhi si blocca:

- (48) Abbiamo messo su il caffè → *Abbiamo messo *sul* caffè

- (49) L'inflazione ha portato su i prezzi → *L'inflazione ha portato
sui prezzi

Gli esempi seguenti, che possono essere interpretati in entrambi i modi, rendono ancora più evidente il fenomeno:

- (50) a I commercianti *tirano sul* prezzo
"Non cedono sul prezzo"
b I commercianti *tirano su il* prezzo
"Aumentano il prezzo"
- (51) a Mettilo dietro, il vetro
b Mettilo die[tril]vetro

Dal punto di vista fonologico, quindi, l'avverbio del VS rappresenta il confine destro del sintagma. Da ciò consegue un particolare comportamento dell'avverbio rispetto all'accento, che costituisce un'ulteriore riprova fonologica della autonomia dei VS rispetto ai verbi polirematici in generale.

Normalmente, gli avverbi che formano sintagma col nominale seguente sono atoni e l'accento di sintagma cade sul nominale stesso (52a):

- (52) a [Sàltano][suibambìni]
b [Saltanosù], i bambini

Gli avverbi che sono parte, invece, di un VS, sono tonici, tendono a formare parola fonologica con la parte verbale del VS e, in posizione finale, portano il rilievo maggiore, come del resto accade anche coi *phrasal verbs* inglesi (52b, 53, 54). In taluni casi, è possibile anche percepire un leggero colpo di glottide tra la vocale finale della particella e la vocale iniziale del nominale seguente (lo indichiamo in 52):

- (53) Abbiamo [portatosù i] bambini
(54) a Ci sono i bambini da [portaresù]
(55) a *Ci sono i bambini da [portàre] su

Non a caso i VS possono avere una cancellazione della vocale finale negli infiniti, e formare con la particella una parola fonologica unica:

- (56) Non ha voluto [venirsù]
(57) Preferisco [stargiù]

A me pare che queste evidenze fonologiche provino ad abbondanza che i VS costituiscono ormai, perfino nella competenza fonologica, delle unità compatte. Ho già accennato prima che

l'avverbio post-verbale può essere interpretato come un quasi-affisso con la funzione di elaborare il significato del verbo: ciò significa che i VS sono soggetti a regole di formazione di parola, le quali, come ogni regola di questo tipo, può comportare dei riaggiustamenti fonologici. Per questo, un modo non banale di interpretare i fenomeni fonologici che ho sommariamente descritto (spostamento sistematico dell'accento e cancellazione di vocale nel verbo) consiste proprio nel considerarli come riaggiustamenti conseguenti all'applicazione di una regola morfologica.

6. Indizi sociolinguistici

Un ultimo rilievo prima di concludere. I VS hanno in molti casi dei verbi monorematici corrispondenti per significato:

- (58) andare avanti \approx avanzare
- (59) andare indietro \approx arretrare
- (60) mettere dentro \approx arrestare
- (61) pasare su \approx trascurare
- (62) sbattere fuori \approx espellere
- (63) tirare avanti \approx campare
- (64) venire fuori \approx uscire

I verbi monorematici sono semanticamente meno trasparenti dei VS, almeno in buona parte dei casi, cioè quando i VS vengono presi nella loro interpretazione letterale. Per questo è legittimo predire che sarà nella varietà informali che i VS tenderanno ad apparire più di frequente, proprio in quanto sono ottenuti dalla combinazione di due parole di alta frequenza e di notevole generalità, una parte delle quali porta un significato che si può leggere con qualche approssimazione nell'insieme risultante. Un'altra interpretazione possibile consiste nel supporre (come Schwarze 1985) che i VS tenderanno ad affiorare (o perlomeno ad essere meglio accettati) nella competenza linguistica di dialettografi settentrionali, anche se personalmente preferisco la prima interpretazione.

Concludo dicendo che mi pare che abbiamo sufficienti ragioni (semantiche, sintattiche, fonologiche, e infine sociolinguistiche) per considerare i VS una classe di parole a sé stanti. È ora quindi di ammetterli come lemmi a sé nei trattamenti lessicografici, ma soprattutto di studiarne le proprietà più da vicino, magari approfittando di questo per riavviare una riflessione sulle classi di

parole e la loro diversa maniera di organizzarsi e di codificare nel lessico.

Riferimenti bibliografici

- Bosque, J. (1991): *Las categorías gramaticales*, Síntesis: Madrid.
- Carrera Díaz, M. (1984): *Curso de lengua italiana*, vol. 1, Ariel: Barcelona.
- De Mauro, T. et al. (1993): *Lessico di frequenza dell'italiano parlato*. Franco Angeli: Milano.
- Franchi De Bellis, A.- Savoia, L. (a c. di) (1985): *Sintassi e morfologia della lingua italiana d'uso. Teorie e applicazioni descrittive*, 2 voll. (= SLI 24), Bulzoni: Roma.
- Quirk, R. et al. (1972): *A Grammar of Contemporary English*, Longman: London.
- Schwarze, C. (1985): 'Uscire' e 'andare fuori': struttura sintattica e semantica lessicale, in Franchi De Bellis-Savoia (1985: 355-371).
- Swiggers, P.-van Hoecke, W. (a c. di) (1986): *Mots et parties du discours*, Peeters: Louvain-Paris.
- Simone, R. (1995): *Fondamenti di linguistica*, Laterza: Bari (1990¹).
- Stiebels, B.-Wunderlich, D. (1994): *Morphology feeds syntax: the case of particle verbs*, "Linguistics", 32: 915-968.
- Voghera, M. (1994): *Lessemi complessi: percorsi di lessicalizzazione a confronto*, "LeSt", 29: 185-214.

LA MICROESTRUCTURA DEL DICCIONARIO DEL ESPAÑOL ACTUAL

MANUEL SECO
Real Academia Española

Es premisa indispensable para hablar del *Diccionario del español actual* la presentación del contexto en que este proyecto nació. La significación y las características de la obra solo se comprenderán adecuadamente si se tiene en cuenta el escenario lexicográfico que sirvió de fondo a su nacimiento.

La peculiaridad de la tradición lexicográfica del español es el *academicocentrismo*. En esa tradición se encuentran, de un lado, el *Diccionario* de la Academia Española; de otro, todos los demás. Pero la relación entre el primero y los segundos es la misma que rige entre el Sol y los planetas.

El *Diccionario* de la Academia, el centro de ese sistema solar, nació durante la primera mitad del siglo XVIII. Era una extensa obra, en 6 volúmenes, que se confeccionó sobre una base documental, rica para su tiempo, de la cual se imprimieron numerosas citas dentro de los artículos como testimonio de la realidad de los sentidos definidos y también como ilustración de los mismos. Por ello este diccionario recibió más tarde el nombre de *Diccionario de autoridades*. Pese a su calidad, que la situaba entre las mejores de su siglo en Europa, no logró ver más que un tomo de una segunda edición.

Lo que hizo la Academia, en 1780, fue producir una versión abreviada del primer diccionario, en un solo volumen. La reducción se consiguió a costa de sacrificar todas las citas en que estribaba el valor científico del *Diccionario de autoridades*. A pesar de esta mutilación, la versión abreviada fue bien acogida por el público, que la reverenció como heredera del prestigio ganado por la primera obra.

Esta versión reducida es la que, con el nombre de *Diccionario académico* por excelencia, o *Diccionario común*, se ha editado veinte veces más hasta 1992. Es cierto que a lo largo de esos 212 años y de esas 20 ediciones no se ha llevado a cabo una sola revisión profunda y

sistemática de la obra, ni se ha ido más allá de introducir enmiendas y adiciones esporádicas. Esto no es obstáculo para que, en el mundo hispanohablante, el respeto popular al *Diccionario* de la Academia sea ilimitado y se haya mantenido incólume a través de los siglos.

Consecuencia natural de esta fe colectiva en una institución y en una obra es el papel central que siempre ha desempeñado el *Diccionario* académico en toda la lexicografía española. Es un fenómeno que no tiene paralelo en ninguna de las lenguas de nuestro entorno cultural. La influencia del repertorio académico, tanto en su macro como en su microestructura, es visible, en mayor o menor grado, en los diccionarios no académicos. No es de extrañar que María Moliner, autora de uno de los más apreciados entre ellos, escribiese en 1966 que "absolutamente todos los diccionarios españoles" habían tomado del académico sus definiciones. Y no excluía el suyo propio.

Este aire de familia ha sido en definitiva un factor empobrecedor para todo el género. Porque las producciones no académicas de más calidad, a pesar del interés de sus innovaciones particulares -y que yo he resaltado con justicia en otras ocasiones-, no han llegado a romper el cordón umbilical con la madre Academia. No deja de existir un aspecto positivo en este hecho, ya que sin duda hay en el *Diccionario* académico mucho material valioso. Pero no es menos verdad que hay un lado negativo, y es que también abundan en él los datos erróneos y los análisis desacertados, como consecuencia, principalmente, del desfase cronológico entre la composición de la obra original y la fecha en que la recibe el lector.

El *Diccionario* de la Academia, tras dos siglos de ediciones, no refleja adecuadamente la evolución de la lengua a lo largo de ese tiempo. Hay en él definiciones que se imprimieron por primera vez en la primera mitad del siglo XVIII y que se reproducen palabra por palabra en la edición de 1992. Añádase a esto la desorientación en cuanto a la vigencia cronológica de muchas voces y de muchas acepciones. Palabras y sentidos cuyo último testimonio conocido corresponde a doscientos años atrás aparecen en el *Diccionario*, en no escasa proporción, sin ninguna advertencia sobre su decrepitud, como si su circulación fuese normal en los finales del siglo XX. Y añádase también el empobrecedor criterio de selección del léxico, basado con frecuencia en principios puristas, y encima con desigual rigor.

La necesidad de una reforma radical del *Diccionario* de la Academia fue denunciada por primera vez en 1945, por Ramón Menéndez Pidal, en el prólogo a la primera edición del *Diccionario Vox*, dirigido por Samuel Gili Gaya. Allí señalaba Menéndez Pidal la urgencia de crear un inventario total del léxico, cuyo objetivo había de ser

el registro tanto de la lengua usada en el momento presente como de la lengua de las épocas pasadas del idioma, "formando un diccionario histórico total, sin criterio selectivo". Y mencionaba como posible modelo de ese inventario total el Diccionario inglés de Oxford. Solamente a partir de una obra así se podrían superar las deficiencias de que adolecía el *Diccionario* académico común.

Muy poco después, en 1947, trazaba Julio Casares el proyecto académico del *Diccionario histórico de la lengua española*, en el que la propuesta de Menéndez Pidal y el modelo de Oxford ocupaban lugar preeminente. Esta obra, prevista en 25 volúmenes, empezó a publicarse en 1960. Cada uno de sus artículos contiene el estudio más completo hasta ahora desarrollado de la palabra, con información no solo sobre su contenido semántico, su sintaxis y su fraseología, sino también sobre su vigencia pasada y actual y su extensión geográfica en España y fuera de España; todo ello realizado de primera mano sobre una documentación suficiente y con ejemplos de uso real, acepción por acepción. Algo que en nuestro país no se hacía desde 1739. La obra en marcha, acorde con la concepción de Menéndez Pidal, constituye en este momento el único punto de partida sólido para reedificar el *Diccionario* común de la Academia.

Por desgracia, las circunstancias difíciles que desde el mismo inicio de su publicación han acompañado a este ambicioso proyecto del *Diccionario histórico* —sin duda el más importante de la Academia desde su fundación— hacen que no podamos vislumbrar en un horizonte próximo los efectos beneficiosos, tan necesarios, de esta obra sobre la lexicografía tradicional española.

En 1969, viendo yo la escasa energía con que era impulsada la construcción de una obra que tenía que ser la piedra angular de una nueva lexicografía española, y la profunda incompreensión que rodeaba la empresa, estudié la posibilidad de intentar por mi cuenta esa renovación de nuestra lexicografía en una escala reducida y dentro de los límites mucho menos ambiciosos impuestos por una extremada escasez de medios. Así nació mi proyecto del *Diccionario del español actual*.

La confección del *Diccionario del español actual* comenzó en 1970. Hoy, felizmente, 25 años después, ya está en fase de maquetación por ordenador como último paso previo a la impresión. Tendrá casi 4.500 páginas, cifra superior a la de los diccionarios de la Academia y Moliner juntos. Todos sabemos muy bien que la importancia de un diccionario no se calibra por el número de páginas; si ofrezco este dato es solo con el fin de dar una idea parcial de la cantidad de esfuerzo que ha llevado la construcción de esta obra.

Como he dicho, los trabajos de este *Diccionario* se han desarrollado siempre con extremada pobreza de recursos. Los tres responsables de la obra hemos tenido que trabajar siempre a tiempo parcial. Toda la tarea, hasta diciembre de 1993, ha sido exclusivamente artesanal, ya que, cuando la empezamos en 1970, los ordenadores (aquellos ordenadores de tarjetas perforadas) tenían precios prohibitivos para las personas normales. Cuando el ordenador, por fin, entró en nuestras casas y empezaba a ofrecer prestaciones verdaderamente interesantes, ya era tarde para incorporarlo a la compilación del *Diccionario*. Solo desde 1994, gracias a la colaboración de nuestros editores, contamos, en la fase de preparación del texto para la imprenta, con los equipos informáticos necesarios y con tres colaboradores expertos destinados especialmente a esta operación.

El corpus documental que ha servido de base para la redacción de la obra es, por tanto, íntegramente manual. Como también ha sido íntegramente manual la redacción misma.

El objetivo del proyecto ha sido romper el círculo cerrado de la tradición lexicográfica del español, superando la habitual dependencia respecto al *Diccionario* académico y poniendo en práctica una metodología más rigurosa que la empleada hasta ahora en la elaboración de los diccionarios de nuestra lengua. Es una avanzada en el intento de descubrir cuál es el léxico real del español actual: por un lado, dejando el catálogo limpio de todas las voces y sentidos que ya no están en uso y que, con etiqueta de "anticuados" o frecuentemente sin ella, abruma las páginas de los diccionarios usuales, especialmente el de la Academia; y, por otro lado, paralelamente, registrando numerosísimas voces y sentidos que pertenecen a la lengua de nuestro tiempo y que no se encuentran en esos diccionarios.

Esta fundamental renovación metodológica tenía como puntos claves, en primer término, el establecimiento de una sincronía precisa para la descripción del léxico contemporáneo, y segundo, la creación de un corpus documental.

La contemporaneidad está definida por una sincronía cuyos límites se fijaron entre el año 1955 y la fecha prevista para la conclusión de la primera redacción. Como esta conclusión se fue retrasando, y por otra parte no era deseable que el componente de actualidad quedase distorsionado por una distancia grande entre el cierre cronológico del corpus y la publicación efectiva del *Diccionario*, al final el límite *ad quem* del léxico descrito quedó definitivamente fijado en los finales del año 1993.

El dominio geográfico ideal para un *Diccionario del español actual* es, lógicamente, el conjunto de los países que hablan esta lengua.

Pero las enormes dificultades técnicas para abarcarlo nos impusieron desde el principio, por puro realismo, limitarnos en esta primera edición al español de España.

El léxico descrito en el *Diccionario* es el de la lengua común, en sus diversos niveles socioculturales y en sus diversos registros. No se excluyen las formas regionales aceptadas en el nivel culto del español hablado en la respectiva zona y siempre capaces de entrar en el vocabulario pasivo de los hablantes del resto del país. Y se incluye también, porque es una exigencia de todos los diccionarios modernos, una proporción notable de términos técnicos correspondientes a ámbitos especiales de la actividad y del saber, aunque limitando su inclusión a aquellos que nuestra documentación nos advierte estar en trance de penetrar en el vocabulario pasivo e incluso activo de los hablantes medios.

La selección del léxico no es en modo alguno intuitiva o introspectiva, sino dependiente de los datos suministrados por nuestra documentación. Hemos llevado a cabo un despojamiento selectivo de textos españoles, bastante numerosos, impresos dentro de los límites de la sincronía establecida: de 1955 a 1993. Esos textos son de tres órdenes principales: a) publicaciones periódicas; b) obras didácticas de enseñanza media o de divulgación; c) obras literarias de diversos géneros. En este último apartado, se excluye la poesía, por ser la forma de expresión más creativa y por tanto la más rebelde al sistema de la lengua. Para la incorporación al *Diccionario* de una voz o de un sentido no nos ha bastado la mera presencia de un testimonio en nuestro fichero, que pudiera representar una creación ocasional; le hemos exigido su confirmación, bien por ulteriores testimonios, bien por informaciones de carácter objetivo.

Una de las características más visibles del *Diccionario* obedece precisamente a nuestro propósito de demostrar la existencia de los usos que en él describimos. Igual que los grandes diccionarios históricos, igual que los grandes diccionarios descriptivos, el nuestro presenta, siguiendo a cada definición, una o más citas, con indicación exacta de su procedencia, con el fin de cumplir la doble función de certificar la realidad de lo expuesto en nuestra descripción y de ilustrar esta descripción exhibiendo la palabra en vivo, en plena actuación dentro de un contexto.

Consecuencia inevitable de esta característica es la extensión de los artículos, visiblemente mayor que la de los diccionarios españoles normales. No ha de sorprender que nuestro diccionario requiera, para 75.000 entradas, las cerca de 4.500 páginas que dije antes.

No es solo la presencia sistemática de citas la razón de la extensión de los artículos; también lo es la presencia de informaciones de carácter sintáctico y pragmático difíciles de encontrar todavía en otros diccionarios españoles.

¿Cuál es el destinatario de nuestra obra? Sin ninguna duda, el hablante culto. Más precisamente: el hablante que tiene unos conocimientos gramaticales que son elementales para el lingüista, pero que no todo el mundo posee, ni mucho menos. En realidad, la mayoría de los diccionarios están dirigidos a ese ideal personaje; pero no son pocos los que lo hacen sin plena conciencia de ello. El simple hecho de asignar a un verbo la etiqueta "transitivo" o "intransitivo" implica la suposición de que el lector tiene unas nociones de sintaxis que, según mi experiencia, no suelen ocupar un lugar en la cabeza del ciudadano medio. Lo que ocurre es que el lector va derecho a las definiciones y se contenta con que estas le ayuden a interpretar la palabra que no entendía. La distancia entre el metalenguaje del lexicógrafo y la comprensión del lector se salva con frecuencia gracias a la imaginación o a la intuición que este aporta. Y gracias a esto, el diccionario *funciona* para muchos a pesar de los escollos que interponemos los diccionaristas.

Como es natural, es el lector con un nivel de cultura que incluya ciertas nociones básicas de gramática quien tiene más capacidad para obtener de un diccionario, y especialmente del nuestro, el máximo rendimiento. En nuestro *Diccionario* este perfil de lector es sin duda el más capacitado para comprender de manera adecuada las diversas facetas de la estructura de los artículos, dentro de los cuales el punto de vista gramatical desempeña un papel eminente. Nuestro *Diccionario* tiene como una de sus principales características el ofrecer para cada unidad léxica y para cada acepción una información sintagmática, y —en menor medida— pragmática, habitualmente ausentes de los diccionarios de nuestra lengua. Parece lógico, pues, que la organización de la entrada tenga como punto de partida una perspectiva sintáctica.

No olvido la sensata observación de Bo Svensén: a medida que los diccionarios modernos tienden a dar información gramatical cada vez más detallada, el ciudadano corriente, gracias a los nuevos rumbos de la enseñanza, se encuentra cada vez más pobre del bagaje mental necesario para ello. "El diseño de los diccionarios y el conocimiento de sus usuarios están, en cierta medida, evolucionando en direcciones opuestas."

Pero, en definitiva: aunque nuestro usuario ideal es el lector culto, con una cierta familiaridad con la gramática, contamos tam-

bién con un nivel de lectores menos ideal y más numeroso que también se va a servir del diccionario, ciertamente con menos eficacia, pero, en último término, de manera *suficiente*, para resolver sus necesidades.

Por otra parte, hemos pensado también en un usuario especial: el que utiliza el español como segunda lengua. Aquí está incluido en primer lugar el estudiante avanzado que ya se ha soltado de la mano del diccionario bilingüe. Al lado de él, el que ya ha salido del terreno del estudiante y entra en el terreno del estudioso. Hemos tenido muy presente también al profesor, nativo o no nativo, que ha de enseñar a extranjeros. Y naturalmente, al traductor, tantas veces ansioso y desesperado consultante de los diccionarios monolingües. Todos ellos son siempre los que más preguntas hacen al diccionario. Por supuesto, más que los hablantes corrientes. Por eso, en conjunto, la información que nuestro diccionario ofrece está pensada no menos para el usuario extranjero y para el usuario "profesional" de diccionarios que para el hablante nativo común.

El enfoque rigurosamente sincrónico del *Diccionario del español actual* lleva como consecuencia la exclusión de toda información etimológica, la cual solo tiene verdadero sentido en aquellas obras en que la entrada es objeto de atención histórica.

El problema de la polisemia y la homonimia, que la tradición lexicográfica española resuelve sistemáticamente tomando como criterio el de la identidad o diversidad etimológica, tiene en nuestro diccionario un enfoque basado en la afinidad o diversidad semántica. Para un diccionario tradicional, *pacífico* tiene como étimo el latín *pacificus*, y todo lo que en español tenga esta forma será registrado bajo un mismo lema. En cambio, nuestro diccionario separa como homónimos dos adjetivos *pacífico*: el primero significa 'de la paz' y 'amante de la paz'; el segundo, 'del océano Pacífico' (por ej., *la costa pacífica*). Otro ejemplo: por razones etimológicas, la Academia reúne en una misma entrada *parte*, nombre femenino, 'porción', y *parte*, masculino, 'comunicación'. En sincronía, el hablante se siente en presencia de dos palabras distintas, y como tales las tratamos nosotros.

La Academia y los diccionarios que la siguen separan (no de manera sistemática) en artículos diferentes la forma masculina y la femenina de nombres de seres animados, e incluso hacen un tercer artículo si existe un uso adjetivo. En el *Diccionario académico* hay una entrada *ministro* y otra entrada *ministra*. Hay una entrada para *gato* y otra para *gata*. Hay una entrada para *perro*, otra para *perra* y otra para el adjetivo *perro/perra*. Hay una entrada *maestro/maestra*

que contiene el adjetivo y el nombre masculino, y otra entrada *maestra* para el nombre femenino. En cambio, *profesor* y *profesora*, como *alumno* y *alumna*, coexisten en una entrada única. Nosotros hemos reunido bajo una sola entrada todos estos significantes con variación genérica que tienen contenidos semánticos comunes o afines. Incluso hemos agrupado en una sola entrada formas masculinas y femeninas cuyo formante genérico es especial (no importa la antigüedad o el origen de esas formas especiales) pero cuya raíz es la misma. Para nosotros, *gallo* y *gallina*, *rey* y *reina*, *actor* y *actriz*, *poeta* y *poetisa*, son la forma masculina y la femenina de una misma palabra. Otro caso, naturalmente, es de lexemas distintos, como el de *caballo* y *yegua* o el de *padre* y *madre*.

La estructura general del artículo más simple, el monosémico, es: 1º, la entrada; 2º, la información sobre la entrada; y 3º, la cita o citas.

La *naturaleza* de nombre, o de adjetivo, o de verbo, o de adverbio, de una palabra está modificada decisivamente por la *función* de esa palabra en la frase, de tal manera que esa función viene a ser su segunda naturaleza. El hecho de que esa palabra sea capaz de formar con otras unas combinaciones dotadas de un significado y una función estables, distintos de los que usa fuera de esas asociaciones, es lo que determina su ubicación dentro de una u otra sección del artículo.

Detengámonos un poco en un ejemplo. *Padre* es "por naturaleza" un nombre masculino, con bastantes sentidos que nosotros hemos reunido en nueve acepciones (algunas de ellas con varias subacepciones), las cuales abarcan desde el primario "hombre que ha engendrado a otra persona" hasta el uso en argot militar como "soldado del reemplazo anterior al último". Pero este nombre abandona su naturaleza de nombre cuando decimos *la paliza padre* o *la vida padre*, donde es un adjetivo funcional. ¿Y qué ocurre cuando *padre* adquiere un papel particular dentro de una frase como *un desajuno de padre y muy señor mío*? ¿Y qué es *padre* cuando decimos *una revista que no lee ni su padre*? ¿Y qué quiere decir una mujer cuando dice de un hombre que *está como para hacerle padre y darle las gracias*?

Vemos que la combinación *de padre y muy señor mío* funciona como adjetivo, con el sentido de 'muy grande o extraordinario'. Es tan adjetivo como el *padre* de *la paliza padre* o el de *la vida padre*. Nosotros, después de haber dedicado un extenso apartado al *padre* nombre masculino, dedicamos un segundo apartado a los usos de *padre* como adjetivo o como componente de combinaciones que tienen sentido y función adjetivos. Un tercer apartado hay que hacer con el *ni su padre* del ejemplo *una revista que no lee ni su padre*, porque ese

ni su padre es un verdadero pronombre que significa 'nadie'. En un cuarto apartado debemos situar la locución *estar* [un hombre] *para hacerle padre y darle las gracias*, que significa 'ser muy atractivo', y que evidentemente hay que considerar como verbo. Todavía hay que reservar un quinto apartado al uso que encontramos cuando alguien informa de que se ha incendiado el Instituto, y alguien responde: ¡*su padre!* Se trata, sin duda, de una interjección, expresiva de susto o sorpresa desagradable.

Ahora bien: si hacemos cinco apartados dentro del artículo *padre* atendiendo a la función sintáctica, la función sintáctica que consideramos no es necesariamente la de la voz *padre*, sino que puede ser la de cualquier combinación estable con sentido estable de la que *padre* sea componente. Tan adjetivo es la combinación *de padre y muy señor mío* en la frase *un desayuno de padre y muy señor mío*, como la palabra *padre* en la frase *la vida padre*; del mismo modo que tan sustantivo es la combinación *padre de la patria* en la frase *los padres de la patria se desvelan por nosotros*, como la palabra *padre* en la frase *Antonio es padre de dos niños*. Agrupamos siempre en un mismo apartado todas las unidades léxicas, sean univerbales o pluriverbales, comprendidas bajo la entrada, que tengan una misma función sintáctica. Cuando en un apartado se dan solo acepciones de la palabra simple, o aparecen, junto con ellas, expresiones pluriverbales, el apartado lleva la etiqueta correspondiente a la categoría: "adjetivo", "pronombre", "verbo", etc. Pero si el apartado está constituido exclusivamente por unidades léxicas pluriverbales, usamos la etiqueta "locución adjetiva", "locución verbal", "locución adverbial", etc.

No ignoro que la denominación "locución", tal como la empleamos nosotros, puede parecerles a algunos abusiva, ya que casos como las combinaciones que he citado no suelen llamarse sistemáticamente así en los diccionarios, sino alternando con otros rótulos como "frases", "expresiones" y "modismos", o bien, como hacen algunos, no dándoles ningún nombre. Como en nuestro diccionario el empleo de "locución" es unívoco y expresa con apreciable brevedad lo mismo que "unidad léxica pluriverbal", hemos decidido servirnos sistemáticamente de ese término sin mayores remordimientos. Lo más particular es que, como norma general, no usamos el término "locución" sin que vaya seguido de la especificación de su función: "adjetiva", "verbal", "adverbial", "prepositiva", etc.

En algunas entradas, el espacio dedicado a la función verbo no solo incluye las locuciones verbales tal como aquí las hemos definido, sino otro tipo de combinaciones que participan de la naturaleza verbal en cuanto que expresan una acción, pero a manera de comentario,

de deseo, de exhortación con respecto al enunciado en que se insertan. Estas combinaciones tienen la característica común de su autonomía sintáctica. Por eso las denominamos *fórmulas oracionales*. Son, en bastantes casos, frases nominales lexicalizadas. Ahora bien: no siempre son frases nominales. Puede estar presente en ellas una forma verbal, pero sin capacidad flexiva, o con muy poca.

Por ejemplo, *a buena parte vas* es una fórmula con que el hablante pondera lo inadecuado que resulta respecto a una persona citada algo que se acaba de sugerir. "¿Vas a pedir ayuda a mi tío? A buena parte vas." Pero en esta fórmula el verbo no es enteramente inmóvil; admitiría variación de persona (*a buena parte vais, a buena parte van*), y, en estilo indirecto libre, variación de tiempo ("¿Iba a pedir ayuda a su tío? A buena parte iba"). En otro ejemplo, *que te frían un ladrillo*, fórmula que manifiesta desprecio o rechazo, el esquema desiderativo es invariable; pues, aunque teóricamente cabría construir **deseo que te frían un ladrillo*, o **me gustaría que te frieran un ladrillo*, o **espero con ilusión que te frían un ladrillo*, estas variantes son inusitadas; eso sí, es posible cambiar la persona destinataria del ladrillo; por ejemplo, *que les frían un ladrillo*. Algo parecido podríamos decir de *con su pan se lo coma*, fórmula con la que se expresa la indiferencia o desinterés por lo que le ocurra a otro o por lo que haga. La total o casi total invariabilidad de todas estas fórmulas impide que las asimilemos a las locuciones verbales, como *estar más bueno* (o *más buena*) *que el pan*, o *estar para mojar pan*, o *estar de toma pan y moja*, en todas las cuales hay un núcleo verbal perfectamente flexible.

En esta modalidad fraseológica se cuentan también fórmulas sin verbo; por ejemplo, *contigo pan y cebolla*, con la que se expresa el total desinterés económico de un enamorado o unos enamorados; u *oído al parche*, que incita a prestar atención; o *y santas pascuas*, que viene a equivaler a "y ya está, y eso es todo". No se confunden nunca estas fórmulas con las interjecciones, ya que no es propia de ellas la carga exclamativa de estas últimas. Ni tampoco con los refranes o proverbios, ya que la aparición de estos en un contexto tiene el carácter de cita, tomada en este caso del archivo literario folclórico.

Pasemos ya a decir algo de las definiciones. Los dos tipos de definición: la *sinonímica*, en metalengua de contenido, y la *explicativa*, en metalengua de signo, están presentes en el *Diccionario del español actual*, pero siempre diferenciadas tipográficamente: la definición sinonímica va siempre en redonda, mientras que la explicativa va siempre en cursiva. Cuantitativamente, constituyen una gran mayoría las definiciones sinonímicas, que hemos utilizado siempre que ha sido posible.

En cuanto a las explicativas, nos hemos servido de ellas en las palabras funcionales, como preposiciones, conjunciones, artículos y pronombres. También en las interjecciones. Y en bastantes fórmulas oracionales.

No hemos vacilado en usar la explicación en casos en que el sentido concreto por definir se realiza normalmente en construcciones de escasa estabilidad formal que —por razones no solo de economía de espacio, sino de economía mental para el lector— no vale la pena definir separadamente. Tal es, por ejemplo, el caso del empleo del nombre *pajarito* cuando se trata de hacer una foto a una o varias personas (no necesariamente a un niño): “Cuidado, que va a salir un pajarito”; “Atención al pajarito”; “Mira el pajarito”. Nos ha parecido suficiente presentar, con la etiqueta de “coloquial”, una acepción explicada así: “Se usa en construcciones como *mirar al pajarito*, o *que sale el pajarito*, para reclamar atención e inmovilidad al que posa para una fotografía”.

También nos ha sido útil la forma “explicación” cuando hemos recurrido a lo que llamamos *definiciones comprensivas*. Ocurre, por ejemplo, que un mismo nombre aparece empleado para designar distintas especies de aves que tienen algunos caracteres comunes, y cuya diferenciación no interesa demasiado al lector normal, aunque ocasionalmente se mencionen con acompañamiento de algún especificador. Así, *paíño*, que explicamos —no definimos— en esta forma sumaria y a nuestro juicio plenamente suficiente para el hablante común: “Se da este nombre a varias aves marinas que suelen seguir a los barcos y solo vienen a tierra para reproducirse”. Y a continuación damos las denominaciones complejas de algunas especies: *paíño común*, *paíño de Leach*, *paíño de Wilson*, *paíño de Madeira*. En un caso como este, una definición unitaria por perífrasis sinonímica hubiera sido imposible. La única alternativa al método adoptado por nosotros hubiera sido presentar en forma enciclopédica cada una de las especies que tienen el nombre vulgar *paíño*; práctica que se saldría de nuestro concepto de diccionario de lengua.

A propósito de este punto, y también a propósito de este ejemplo, quiero precisar que nuestras definiciones de voces designadoras de realidades que son o pueden ser objeto de estudio por parte de especialistas, o nuestras definiciones de tecnicismos, no están redactadas precisamente para servir a los especialistas o a los técnicos, sino a los hablantes comunes, a quienes basta saber qué clase de ser es el *paíño* y qué rasgos lo diferencian entre los de su clase, a fin de no confundirlo con el buitre ni con la gaviota; o qué clase de entidad es el metacrilato y cuál es su principal utilidad, con las notas empíricas

o intuitivas suficientes para que no se confunda con otras sustancias afines. A veces damos algún detalle técnico, como es el nombre zoológico o botánico en todas las entradas de animales y plantas: dato útil para quien desee buscar una identificación más precisa fuera del *Diccionario*. Pero renunciamos decididamente a estos datos adicionales cuando hay que plasmarlos en fórmulas matemáticas o químicas.

Como he dicho, hemos hecho uso de la definición sinonímica siempre que nos ha sido posible. En bastantes casos el sistema no ofrece dificultades: *panglosiano*, "que se caracteriza por la creencia de que la humanidad vive en el mejor de los mundos posibles"; *pánico*, "terror repentino y extremo, generalmente irracional"; *ser más papista que el papa*, "mostrar en un asunto más celo o interés que las personas directamente implicadas".

El trabajo empieza a complicarse cuando se entra en el mundo de los verbos transitivos. Para definir el verbo transitivo *pasar* en un ejemplo como *pasar una hebra por el ojo de la aguja*, la Academia no se plantea ningún problema: "introducir una cosa por el hueco de otra". Definición que estructuralmente no se diferencia de la definición de una acepción o de un verbo intransitivos; por ejemplo, *ladrar*, "dar ladridos". Ambas definiciones, transitiva e intransitiva, coinciden en incluir un complemento directo del verbo definidor: *una cosa* y *ladridos*, respectivamente. No se presta atención al hecho de que en un caso, el de la acepción intransitiva, el complemento directo del verbo definidor forma parte sustancial de la definición, mientras que en la acepción transitiva el complemento directo del verbo definidor ("introducir *una cosa*") forma parte del contorno de la definición. Incluso en la misma definición hay otro elemento del contorno, que es el complemento adverbial "por el hueco de otra". La definición presentada por la Academia no define, en realidad, el verbo *pasar*, sino toda la frase del ejemplo, *pasar / una hebra / por el ojo / de la aguja*, a la que estructuralmente se superpone palabra por palabra: "introducir / una cosa / por el hueco / de otra".

Muchos diccionarios modernos, en sus definiciones de verbos transitivos, separan de la definición el complemento directo de contorno, utilizando diversos procedimientos. Nosotros también lo hacemos, valiéndonos de corchetes. Así, para definir el verbo *pasar* de frases como *pasaron miedo*, *pasaron hambre*, *pasaron humillaciones*, nuestra definición dice: "Experimentar [una sensación o situación, generalmente desagradable]".

Pero, como hemos visto en el ejemplo del ojo de la aguja, el texto definidor puede llevar un contorno que no se limita a un complemento directo. Los otros elementos del contorno también deben

presentarse debidamente marcados como tales. Es más: no solo en definiciones de verbos transitivos es posible hablar de contorno. En una definición académica como la de *nadar* leemos: "Trasladarse una persona o un animal en el agua, ayudándose de los movimientos necesarios, y sin tocar el suelo ni otro apoyo". En esta definición de verbo intransitivo, tan precisa, hay un elemento de contorno, que no es complemento del verbo definidor, sino su sujeto: "una persona o un animal". Es lógico que también este elemento de contorno se exprese aislándolo gráficamente, de algún modo, del resto de la definición. Y así lo hacemos nosotros, dando con ello un paso más en la formalización del esquema definitorio.

Los casos en que la mención del contorno se hace necesaria en una definición no son escasos, y además no se limitan a los verbos. El nombre al que se adjunta un adjetivo es contorno de ese adjetivo. Y hay ocasiones en que es preciso indicar a qué clase de nombres se limita el margen de aplicación del adjetivo. Por ejemplo, nombres que designan persona, o que designan animal. La Academia, para esta precisión, recurre a la definición en metalengua de signo: "dícese del individuo que...", "dícese de la persona que...", "dícese del caballo que...". Por ejemplo, en *saduceo*: "dícese del individuo de cierta secta de judíos que..." (etc.). En *tordo*: "dícese del caballo o yegua, o del mulo o mula, que tiene el pelo mezclado de negro y blanco". El *Diccionario del español actual* emplea para estos casos el mismo recurso que para el contorno de la definición verbal: exponer entre corchetes el nombre designador de la categoría de seres a la que es aplicable el adjetivo; para el caso de *tordo*, sería: "[Caballería] que tiene el pelo mezclado de negro y blanco".

La definición, en nuestro diccionario, va acompañada de las precisiones sintagmáticas necesarias, incluidas, siempre que sea posible, en las precisiones del contorno. Por ejemplo, en el verbo *pagar* tenemos, entre las acepciones transitivas, una de "recompensar", que se formula así: "recompensar [a alguien con algo]". Aquí el contorno está constituido por un complemento directo de persona y un complemento de cosa con la preposición *con* (*le pagó con un buen cargo*). En cambio, en la construcción *pagarlas*, la inconstancia del contorno (compl. indir. pers.) y la necesidad de una indicación pragmática nos inclinan a esta solución: *pagarlas*: "sufrir el castigo o las consecuencias correspondientes a un comportamiento. *Frecuentemente en fórmulas de amenaza y con complemento indirecto de persona*". Ejemplo: "Ya me las pagarás".

La colocación es otra información poco frecuente en los diccionarios españoles, y a veces confundida con las locuciones. En la

voz *país* hay una acepción definida como "región geográfica más o menos netamente delimitada y considerada sobre todo en su aspecto físico". (Su ejemplo es *el país comprendido entre los dos ríos está sembrado de frutales*). Pero a continuación se indica: "Frecuentemente en la locución DEL PAÍS" (por ejemplo, *vino del país*). Otra muestra: en un artículo *pago*, una acepción dice: "En plural: Tierra o lugar. Generalmente en la construcción POR ESTOS PAGOS". Otra más: *al pairo*, locución adverbial: "Sin cuidado. Con el verbo TRAER".

No puedo extenderme más. He rebasado ya mi tiempo y la paciencia de ustedes. Naturalmente, quedan sin exponer y comentar bastantes aspectos más de esta obra que aspira a abrir una ventana nueva en el recinto de la lexicografía española. Pero, con lo dicho aquí, creo que hay trazos suficientes para vislumbrar con alguna nitidez su fisonomía. Espero que no tarden en poder examinarla en toda su integridad y que ese examen abra paso a la aprobación de todos, especialmente de los doctos, perseguida por nosotros con ahínco durante tan largos años.

LES DOMAINES RESPECTIFS DE L'EXEMPLE ET DE LA CITATION DANS LES DICTIONNAIRES DE LANGUE ACTUELS

JOSETTE REY-DEBOVE
Dictionnaires Le Robert
Paris

Introduction

Je n'envisage pas, dans cet exposé, d'étudier les dictionnaires actuels dans l'utilisation qu'ils font des exemples forgés et des citations. D'abord parce que chaque dictionnaire monolingue manifeste certes des intentions, mais subit des contraintes qui ne concernent pas la linguistique. Ensuite parce que dans l'ensemble de la lexicographie, et ceci depuis longtemps, il existe une constante dans ce domaine où il y a eu **peu d'innovations**; en gros, on trouve deux types de dictionnaires de langue monolingues, ceux qui présentent des exemples forgés, et ceux qui présentent des citations signées. Un troisième type cumule les deux premiers modèles. Un quatrième type, très pauvre, n'utilise ni l'un ni l'autre, et ne saurait constituer une option théorique.

On verra que ces quatre types ne constituent qu'un classement superficiel. Les vraies différences entre le premier et le second type touchent la codification: il existe un continuum entre l'entrée, les sous-entrées et les exemples, ce **continuum** représentant une seule matière lexicale travaillée par la **codification** plus ou moins grande des unités.

Je me propose donc, dans un premier temps, d'établir le **statut théorique** de l'exemple forgé et de la citation afin de pouvoir présenter des différences d'ordre sémiotique, en abandonnant l'aspect historique, culturel et littéraire.

Dans un second temps, je mettrai l'accent sur une double opposition encodage / décodage. D'abord l'opposition traditionnelle

de l'encodage d'un énoncé (sa production) et du décodage d'un énoncé (sa compréhension) par un utilisateur quelconque de la langue, hors dictionnaire. Ensuite la nécessaire opposition entre l'encodage du texte de dictionnaire par le lexicographe et le décodage du même texte par un lecteur quelconque. L'importance de la seconde opposition, tout entière dans le métalangage, m'apparaît plus nettement aujourd'hui que naguère.

Enfin, je tenterai d'expliquer quel parti la lexicographie moderne peut tirer de ces distinctions théoriques pour une meilleure description de la langue, en défendant la thèse de la nécessaire complémentarité de l'exemple forgé et de la citation.

1. Le texte du dictionnaire au décodage

La notion d'exemple lexicographique englobe la citation, dans son sens général. C'est ainsi qu'on parlait et qu'on peut encore parler d'exemples signés et d'exemples forgés.

1.1. Statut de l'exemple

-Lemmes et formes

Dans le cas le plus général l'exemple est un énoncé, phrase ou syntagme, qui contient le mot-entrée. Cette entrée est délemmatisée et toutes les formes courantes codées peuvent y figurer (formes féminines, plurielles, conjuguées, etc.). Il arrive que le mot qu'illustre l'exemple soit formellement identique à l'entrée (verbe à l'infinitif, nom au singulier), mais cette entrée est également délemmatisée puisque l'entrée métalinguistique construite comme unité de langue devient un mot ordinaire du discours qui est perçu comme tel par le lecteur. Ainsi, pour l'entrée *manger*, verbe, l'exemple: "*Ils ne peuvent pas manger à leur faim*" ou l'exemple: "*Manger à sa faim*", exemple matriciel sur lequel on reviendra, mais qui est perçu aussi comme du discours naturel bien qu'il ne forme pas de phrase (imaginez *Manger à sa faim* sur des écriteaux, dans une manifestation).

-Discours primaire devenu autonome

Que l'exemple soit forgé ou signé cet énoncé a le même statut sémiotique pour le lecteur de l'article. C'est un discours rapporté qui a été produit par un autre que lui et qui ne lui est pas destiné. Dans ce discours secondaire rapporté, ni le lexicographe ni l'écrivain ne s'adresse au lecteur, alors que c'est le cas pour les autres énoncés

métalinguistiques. A preuve un énoncé comme *Vous mangez trop vite*, qui n'est pas un reproche fait au lecteur. L'exemple forgé ou signé est un discours qui a le statut de l'énoncé rapporté en style direct, séquence autonome qui n'a ni synonyme ni traduction dans une autre langue à l'intérieur du texte; et cette information sur sa forme obligée est présente dans son signifié.

1.2. *Le bon exemple*

Ce qu'on appelle *exemple* garde d'autre part son sens étymologique de "modèle". Le **bon exemple** est un énoncé qui offre **une grande information sur le mot** en montrant quelles positions autorisent la **syntaxe** et quelles formes sont déterminées par la **morphosyntaxe**. Ici la catégorie du mot est très importante; je n'entrerai pas dans le détail que vous connaissez, sinon pour rappeler que le nom est la partie du discours la plus indépendante et que, de plus, tout est susceptible d'être nominalisé en discours. L'information recherchée par le lecteur dans ce domaine concerne donc plutôt le verbe et l'adjectif. Le bon exemple est aussi un énoncé qui, pour le **contenu**, apparaît comme en accord avec la définition, mais plus encore, dont le contexte apporte une forte présomption sur le signifié du mot sans jamais être lui-même une définition. *Tu manges à la faim* informe plus sur le signifié de *manger* que *il ne faut pas manger trop vite*; c'est facile à démontrer.

1.3. *L'exemple signé vu par le lecteur*

Le bon exemple signé doit présenter les qualités dont j'ai parlé. S'y ajoute au décodage une fonction d'**attestation**, où le lecteur voit que le lexicographe n'a pas inventé toute cette histoire. C'est un point important pour la fiabilité, mais cela ne concerne que le lecteur, et la simple attestation peut n'avoir aucune valeur informative. Le mot simplement attesté est même en mesure de donner une **contre-information**; ainsi à *bleu*, "*La terre est bleue comme une orange*" (Éluard). On a en effet appris au lecteur à tenir le lexicographe en piètre estime depuis que les dictionnaires de langue sont anonymes, présentés comme dépourvus d'auteurs. Même s'il s'agit d'un quelconque extrait de presse, lorsqu'il est signé, il bénéficie d'une présomption de véridicité et de compétence langagière alors que le lexicographe suscite toutes les méfiances. D'où les dictionnaires entièrement construits à partir d'un **corpus**, comme

le *Cobuild* de Collins où l'auteur John Sinclair s'est complètement effacé.

Enfin lorsque la citation est signée par un écrivain (**autoridad** comme disent les Espagnols) apparaît le prestige de la littérature. Mais c'est d'ordre social et esthétique. Le lecteur ne considère plus guère le langage pour lui-même; c'est pourquoi il s'indigne si une citation du genre "*Il mangea sa soupe et alla se coucher*" est signée par Victor Hugo, car on ne dérange pas un écrivain pour une phrase qui n'est pas au-dessus du commun.

2. Le texte du dictionnaire à l'encodage

Les choses sont à la fois plus complexes et plus claires à l'encodage, ceux qui écrivent des dictionnaires le savent déjà.

D'abord l'exemple forgé ne saurait aucunement avoir le statut sémiotique de l'exemple signé. Pour le lexicographe ce sont deux types d'informations théoriquement irréductibles.

2.1. Production d'un énoncé hors situation

L'exemple forgé, c'est la simulation personnelle du lexicographe. Il se trouve obligé de construire des énoncés **hors situation**, par rapport au reste de son discours qui possède une énonciation véritable sur le langage et s'adresse au lecteur. Mais, dira-t-on, l'écrivain ne produit-il pas des énoncés hors situation? N'est-ce pas justement son activité principale? Non, l'écrivain a différents modes d'expressions; quand il produit un dialogue entre des personnages il est dans la situation de ses personnages et non hors situation. Par ailleurs, son discours peut être narratif: soit en racontant un univers qui n'est pas le sien, soit en restant dans son propre univers quel que soit le sujet abordé. Dans tous les cas, il s'adresse au lecteur de son livre pour les contenus qu'il exprime, il s'adresse à lui en **langage primaire**, alors que l'exemple du lexicographe est choisi de façon **métalinguistique** pour apprendre au lecteur quelque chose sur le signe.

2.2. Double niveau de la citation

Après avoir comparé l'activité du lexicographe et celle de l'écrivain par rapport à son lecteur, revenons au texte du dictionnaire où les choses deviennent plus compliquées. En effet, ce qu'a écrit un auteur est cité dans le texte du lexicographe, et en tant que tel,

devient autonome comme l'exemple forgé. Les lecteurs du dictionnaire ne doivent pas être confondus avec les lecteurs de l'écrivain. La citation, comme l'exemple, est là pour informer le lecteur sur le signe et non sur l'univers du discours de son auteur.

Mais la différence sémiotique persiste d'une autre manière. Alors que l'exemple autonome est produit d'un **premier jet** par le lexicographe (tout comme l'exemple du grammairien), la citation est un énoncé **rendu autonome** pour les besoins lexicographiques et présente deux niveaux: 1) le choix de son énoncé par l'écrivain qui le produit. 2) le choix du lexicographe d'insérer cette citation plutôt qu'une autre.

Certains **conflits** qui naissent parfois entre le lecteur de dictionnaire et le lexicographe expliquent bien ce double niveau. Ainsi le mot *intérimaire* était illustré dans le Nouveau Petit Robert par cette citation de Remo Forlani: "*sa seule et unique chance de ne pas finir intérimaire mal payée dans des jobs miteux*". Aussitôt plusieurs sociétés d'intérimaires protestaient; elles savaient bien, disaient-elles, que l'exemple n'était pas du lexicographe, mais elles réaffirmaient la responsabilité du lexicographe pour le choix de la citation. Même problème, mais concernant la langue cette fois, et non le référentiel: pourquoi avoir donné une citation de Le Clézio, qui parlant d'une voiture, écrit "*regarder dans le capot*" alors que la construction normale est *sous le capot*?

Il est évident que le lexicographe n'aurait produit aucun de ces deux exemples: le premier est violent et peu objectif, le second est fautif. Finalement on rapprochera le cas de la citation et celui du mot-entrée: ce sont les deux séquences langagières autonomes fondamentales que le lexicographe ne produit pas, qui surgissent du sein de la société. Tout comme il arrive que le lecteur se scandalise du contenu d'une citation, il arrive qu'il souhaite la disparition d'une entrée, souvent pour de nobles raisons (grammaticales, sociales, idéologiques) mais qui n'ont pas vocation à nier la réalité lexicale. La locution et le proverbe sont dans le même cas (*saoul comme un Polonais - souvent femme varie, bien fol est qui s'y fie*). Il ne s'agit rien moins que d'expurger le dictionnaire.

Ceci nous amène, par des anecdotes, à articuler l'encodage d'un article de dictionnaire d'après trois types de critères: 1) le **métalangage** qui décrit le signe et s'adresse au lecteur (par ex. les datations, la transcription phonétique, la définition) 2) **Les séquences autonomes** qui montrent des signes. 3) Parmi les séquences autonomes, celles qui sont produites par le lexicographe (exemples) opposées à celles qui ne sont pas son oeuvre (citations, entrées —ainsi

que synonymes et contraires—, étymons, sous-entrées, locutions, proverbes).

Cette structure d'encodage n'est pas perçue par le lecteur moyen et c'est dommage, car bien des dialogues de sourds seraient évités.

Néanmoins la difficulté est grande, puisqu'on observe que l'exemple forgé comme la citation présentent des côtés faibles, opposés, mais également critiquables.

L'exemple forgé assène le code avec toutes ses bêtises socialisées; l'exemple cité transgresse le code par son absence d'objectivité et ses fantaisies personnelles. C'est **entre ces deux excès** que le lexicographe trouve son véritable rôle, qui ne consiste pas seulement à décrire la norme linguistique, comme on le verra.

2.3. La fonction de l'exemple forgé et de la citation

L'exemple est proche des codes.

Si nous avons été conduite à rapprocher l'entrée, l'exemple, la sous-entrée, la locution et le proverbe comme unités autonomes, on voit donc que seul l'exemple est formé par le lexicographe, autrement dit non codé. Je ne traiterai pas ce sujet du syntagme, sinon pour faire remarquer que la différence entre l'exemple et l'unité codée de plusieurs mots n'est pas toujours claire.

En fait l'exemple du lexicographe n'est pas entièrement libre: s'il était libre, il serait parfois mauvais, selon les critères évoqués tout à l'heure, il pourrait aussi être faux pour le contenu sans être inacceptable au sens linguistique; or le lexique ne peut être décrit sans tenir compte des valeurs de vérité d'une société, fondements de la sémantique lexicale. L'exemple forgé est **déjà codé** à un faible niveau, c'est une phrase attendue, une phrase vraie, une collocation fréquente dont l'aboutissement dans la codification sera le jugement sémiotique (lieu commun, énoncé tautologique) la locution, la sous-entrée syntaxique et le terme. Par exemple pour l'entrée *manger*, cet aboutissement serait: *Il faut manger pour vivre, Manger comme un oiseau, Étoffe mangée aux mites, Salle à manger*. La préférence du lexicographe va aux unités codées ou en voie de codification situées sur un continuum; les unités codées sont de la nature du mot, du syntagme ou de la phrase et toutes relèvent de la description du dictionnaire de langue, si l'on considère que la langue est formée d'une grammaire et d'un lexique. Il existe évidemment des dictionnaires spéciaux de locutions et de proverbes, mais ni plus ni

moins qu'il existe des dictionnaires de prononciation ou d'étymologie. Cependant, le **dictionnaire général** dont le programme d'information est **maximal** doit évidemment traiter toutes les unités codées qui sont des **signes**: morphème lexical libre ou lié comme intégrant de mot (radicaux et affixes); mot graphique et lexie (dont les termes) comme intégrants de phrase; locutions, allusions et proverbes comme intégrants du discours ou du contexte situationnel (la notion d'*intégrant* est empruntée à Benveniste). Rien de tout cela ne relève de la grammaire.

L'exemple matriciel

Enfin, j'envisagerai l'exemple matriciel à part, pour deux raisons. D'abord parce que dans l'encodage du texte il est investi d'une fonction qu'il n'a pas au décodage. L'exemple matriciel présente au moins deux types: 1) la locution avec variables grammaticales (*Manger à sa faim* → "*nous mangeons à notre faim*") 2) le syntagme généralisant avec variables sémantiques (*Manger des choses non comestibles* → "*le bébé a mangé de la terre*").

Dans les deux cas il s'agit d'une matrice donc d'une structure codée, et non d'une séquence codée en bloc formellement; c'est un modèle que le lexicographe propose pour que le lecteur le fasse varier. Ce modèle est néanmoins acceptable tel quel comme simple discours (sa forme est attestée).

La seconde raison qui me fait distinguer l'exemple matriciel, c'est qu'il n'est pas représenté dans l'exemple signé, l'écrivain n'ayant pas du tout cette relation avec son lecteur. Dans ce cas, l'exemple forgé est tout à fait irremplaçable: ni la littérature, ni un corpus de textes ne peuvent offrir cette séquence qui n'est ni du langage primaire ni du métalangage et qui ne peut être produite que dans un texte métalinguistique très pointu.

En résumé la vocation de l'exemple forgé est une tentative de décrire la **langue** plutôt que le discours, et par voie de conséquence de déployer les valeurs de vérité d'une société à une époque donnée; c'est bien le lexique qui est décrit, et non pas n'importe quels mots. La définition des mots est en prise directe sur l'exemple forgé, et non pas sur l'exemple signé; nous ne sortons jamais du **général**.

La citation se libère des codes.

A contrario, la citation signée, surtout la littéraire, est un discours personnel qui se moque des valeurs de vérité, qui repousse

les lieux communs et qui affectionne les collocations imprévisibles. C'est l'irruption de la liberté de l'individu qui fait à la fois un **usage optimal de la langue** (le style) et qui **déstabilise les lieux communs** (le contenu). Il constitue forcément un facteur d'évolution de la langue et de l'idéologie. Et il est normal que le lexicographe, dans sa description, laisse ce rôle à l'écrivain. L'écrivain n'a pas besoin de s'appuyer sur des valeurs socialisées, à la différence du lexicographe qui observe la société dans son expression langagière. D'où ce discours singulier et créateur de mondes qui ouvre le code sur les codes nouveaux de la pensée et de l'expression.

Je vous rappellerai la citation célèbre de Voltaire à propos du premier dictionnaire de l'Académie française (1694): "Il me semble aussi qu'on s'était fait une loi de ne point citer; mais un dictionnaire sans citation est un squelette (Correspondance, 11 août 1760). Comme le dictionnaire de l'Académie ne contient que des exemples forgés, ce *squelette* serait donc aussi fait d'exemples forgés et représenterait le code, alors que la citation est explicitement, pour Voltaire, d'une autre nature.

Nous ne voulons pas, en opposant l'exemple forgé et la citation littéraire caricaturer la situation. Il arrive qu'une citation exprime un lieu commun sans apparence de style, et un exemple forgé peut retenir l'attention pour des raisons de contenu ou de forme; mais c'est à la limite de la tolérance. Pour ma part je considère que ce sont de mauvaises citations et de mauvais exemples dont le lexicographe porte la responsabilité, puisqu'il choisit tout. Cette distorsion du texte par des "contre-emplois", comme on dit au cinéma, nuit à sa fonction pédagogique en provoquant à la fois la gêne et la dérision du lecteur.

3. Conclusion: les normes

Il me semble donc que la citation signée relève du singulier et l'exemple forgé du général et que les deux types soient complémentaires: le lecteur a besoin de connaître **le code**, mais aussi d'admettre tout **ce qui est compatible avec le code**. C'est l'aspect créatif, ludique et évolutif de la langue. Néanmoins l'exemple forgé et notamment l'exemple matriciel est irremplaçable et je maintiens qu'un **corpus** n'est pas de nature, en soi, à distinguer ce qui est codé de ce qui ne l'est pas.

Il est finalement légitime de se demander si l'exemple forgé ne représenterait pas la norme linguistique dans la triple opposition **système, norme et discours** (*sistema, norma y habla* de Coseriu), la

norme linguistique venant recouper le système dans ce que j'ai appelé le code. Mais la notion de norme linguistique est beaucoup trop étroite s'agissant du dictionnaire de langue comme objet social qui renvoie une image de l'homme. L'exemple tient compte aussi de la norme sociale, de la norme normative qui en France est la norme scolaire; il crée aussi des normes qui essaient d'arbitrer entre l'individu et la société, norme scientifique, éthique, esthétique. Le lexicographe est un observateur mais aussi un modérateur et un pédagogue. Il y a encore un bel avenir pour le lexicographe entre l'ordinateur et le lecteur de dictionnaire.

Presentazione

Il Tesoro della lingua italiana della moglie (TELO) è un progetto di vocabolario scritto dalla lingua italiana dalle origini fino al 1975 (data della morte di Botticelli), che costituisce la prima parte del vocabolario scritto della lingua italiana con l'Accademia della Crusca consociato e lavorato nel 1985. Dal 1985 il lavoro è passato al Centro del vocabolario italiano (CIVI), un Centro di studi del Consiglio Nazionale delle Ricerche (CNR) che opera presso l'Accademia, ma è distinto e indipendente da essa.

Il risultato raggiunto fino a questo momento è una base di dati dell'italiano scritto di circa 10 milioni di occorrenze (84 mila i testi più grandi per l'impaginato Novati) di cui circa 5 milioni hanno formato, funzionando nella rete informatica interna dell'ENVI in DSI (il software lexicografico di Eugenio Padoa-Schioppa dell'Istituto di Linguistica Computazionale del CNR), lo stesso con cui funziona la LE di Pasquale Stoppelli e della rete (Cicchi), e porta alla consultazione degli italiani, per ora presso la sede di Firenze.

Gli sviluppi successivi, con l'ENVI ora lavorante, sono: il completamento della base di dati per circa altri 5 milioni di occorrenze prevedibili, e parallelamente la sua messa a disposizione degli studiosi in rete internet, attraverso il progetto NetNet, in collaborazione con la comunità dell'ARTS (Università di Chicago) e la Università di Notre Dame e di Nanning (responsabile tecnico Mark Olsen), e possibilmente anche attraverso altri programmi, con diverse modalità di consultazione, quindi la redazione del ZEO propriamente detto, in forma di vocabolario scritto (carta) o una necessariamente a stampa) in base ai dati automaticamente disponibili.

NOTIZIE DALL'OPERA DEL VOCABOLARIO ITALIANO: IL TESORO DELLA LINGUA ITALIANA DELLE ORIGINI

PIETRO G. BELTRAMI
Università degli studi di Siena

Presentazione

Il *Tesoro della lingua italiana delle origini* (TLIO) è un progetto di vocabolario storico della lingua italiana dalle origini fino al 1375 (data della morte di Boccaccio), che costituisce la prima parte del vocabolario storico della lingua italiana cui l'Accademia della Crusca cominciò a lavorare nel 1965. Dal 1985 il lavoro è passato all'Opera del vocabolario italiano (OVI), un Centro di studi del Consiglio Nazionale delle Ricerche (CNR) che opera presso l'Accademia, ma è distinto e indipendente da essa.

Il risultato raggiunto fino a questo momento è una base di dati dell'italiano antico di circa 12 milioni di occorrenze (14 con i testi già pronti per l'implementazione), di cui circa 8 milioni lemmatizzate, funzionante nella rete informatica interna dell'OVI in DBT (il software lessicografico di Eugenio Picchi, dell'Istituto di Linguistica Computazionale del CNR, lo stesso con cui funziona la LIZ di Pasquale Stoppelli e dello stesso Picchi), e aperta alla consultazione degli studiosi, per ora presso la sede di Firenze.

Gli sviluppi successivi, cui l'OVI sta lavorando, sono: il completamento della base di dati (per circa altri 4 milioni di occorrenze prevedibili), e parallelamente la sua messa a disposizione degli studiosi in rete Internet, attraverso il progetto *ItalNet*, in collaborazione con la struttura dell'*ARTFL* (Università di Chicago) e le Università di Notre Dame e di Reading (responsabile tecnico Mark Olsen), e presumibilmente anche attraverso altri programmi, con diverse modalità di consultazione; quindi la redazione del TLIO propriamente detto, in forma di vocabolario storico (anche se non necessariamente a stampa); infine, se sarà economicamente possibile,

l'impostazione della seconda parte del vocabolario, dal 1375 ai giorni nostri.

Cenno storico

Nella preistoria dell'impresa c'è l'interruzione della quinta edizione del *Vocabolario degli Accademici della Crusca* nel 1923, alla fine della lettera O, per una decisione governativa, che assegnava all'Accademia il compito di pubblicare i testi italiani antichi (cfr. Peri 1990); l'opzione di completarla non fu presa seriamente in considerazione nel successivo dibattito intorno ad un nuovo vocabolario, prospettato in termini molto diversi (cfr. Barbi 1935, Pasquali 1941, Nencioni 1955). L'esigenza era quella di un vocabolario al servizio della storia della lingua, diverso perciò dal *Vocabolario degli Accademici*, che nelle sue cinque edizioni (Crusca 1612, 1623, 1691, 1729-38, 1863-1923) era anche un vocabolario storico, in quanto fondato su citazioni d'autore, ma soprattutto normativo, in quanto proponeva un particolare tipo di lingua a scapito di altri (rimando solo ad alcuni dei numerosi studi che ne parlano: Parodi 1974 e 1983, Vitale 1959, 1966 e 1971, Sessa 1991). Nondimeno, quando il CNR decise di finanziare un nuovo vocabolario, fu proprio la Crusca a riceverne il compito, con una convenzione che assicurò le risorse fino al 1980; direttori dell'impresa furono prima Aldo Duro, poi d'Arco Silvio Avalle. Dopo un periodo di crisi, all'inizio degli anni '80, l'impresa fu separata dall'Accademia e divenne come ho detto un Centro di studi del CNR, voluto da una legge del 1983 e operativo dal 1985; direttori prima Carlo Alberto Mastrelli, poi il sottoscritto (dal 1992).

La decisione di dividere il progetto complessivo in due sezioni cronologiche, e di concentrare tutto il lavoro sulla prima, fu presa nel 1972, in seguito a una più attenta valutazione dei tempi di lavoro prevedibili in relazione ai mezzi a disposizione. Sarebbe però errato considerare il solo aspetto negativo di tale decisione, in quanto riduzione di programma; in realtà il risultato è stato di identificare un progetto scientifico necessariamente distinto da quello del vocabolario fino ai giorni nostri, avente come caratteristiche (sulle quali tornerò più oltre) la possibilità di uno spoglio almeno tendenzialmente esaustivo dei testi e il plurilinguismo, rispetto alla necessaria drastica selettività e al monolinguisimo del progetto formulabile per la sezione 'moderna'. Quanto poi all'idea di procedere per sezioni cronologiche, gioverà ricordare che in Francia, dove la lessicografia della lingua nazionale gode di ben altro sostegno

economico e organizzativo, si è rivelata vincente la scelta di dividere il *Trésor de la Langue Française*, elaborando dapprima la parte relativa agli ultimi due secoli, senza pregiudizio per le altre sezioni, cui ora si lavora affrontando il *Dictionnaire du Moyen Français* con altrettante sicure prospettive di successo. Che poi, com'è logico in Francia affrontare un vocabolario diviso per fasce cronologiche cominciando dagli ultimi secoli, sia logico in Italia affrontarlo partendo dalle origini, è cosa ovvia: non solo perché il periodo delimitato comprende i tre classici fondatori della lingua italiana, ma anche perché questa prima impresa è destinata a colmare il vuoto più rilevante fra gli strumenti disponibili agli studiosi di storia della lingua e di filologia italiana (non manca di ricordarlo Max Pfister, lamentando che l'impresa peraltro vincente del *Lessico Etimologico Italiano* non abbia alle spalle il sostegno di un Godefroy o Tobler-Lommatzsch dell'italiano antico: cfr. per es. Pfister 1994, pp. 14-15).

Il vocabolario storico della lingua italiana: aspetti innovativi

Nel 1965, quando fu avviato il progetto del vocabolario storico, il *Grande Dizionario della Lingua Italiana (GDLI)*, l'unico grande vocabolario storico moderno dell'italiano, oggi felicemente avviato verso la conclusione, stava muovendo i primi passi (erano usciti i primi tre volumi). Il progetto dell'Accademia era qualcosa di più e di diverso; per quanto riguarda ciò che ancora conta nel procedere dell'impresa, gli aspetti più innovativi si possono distinguere in tre punti essenziali.

Il primo consiste nell'idea di non fare un nuovo vocabolario della sola lingua letteraria, o comunque dedicato a questa in misura prevalente, nel solco della tradizione lessicografica italiana (cui questa caratteristica è stata giustamente rimproverata, per es. da Zolli 1989), ma un vocabolario della lingua in tutta la varietà dei suoi usi, non solo letterari, ma anche scientifici, giuridici, tecnici e di ogni altro tipo. Non a caso l'impresa del vocabolario storico della lingua italiana è stata inizialmente congiunta con quella del vocabolario giuridico (cfr. Fiorelli 1979). Questo scrupolo di eshaustività per la parte antica si estende, come si vedrà più avanti, alla documentazione di tutti i tipi di lingua scritta nei primi due secoli.

Il secondo aspetto è l'impegno ad un rigoroso controllo filologico dei testi da utilizzare, messo in pratica non solo esaminando sempre le edizioni dal loro interno, valutandone i criteri e verificando la coerenza con cui questi sono stati applicati, ma anche, in molti casi, procedendo a controlli sulle fonti e a vere e

proprie revisioni testuali, e ciò non solo nel caso dei testi antichi, e in particolare di quelli documentari, per i quali si sono collazionati nuovamente i manoscritti, ma anche nel caso di testi moderni, fin quando l'impresa se ne è potuta occupare (si veda in proposito De Robertis 1985). Si tratta, come appare chiaro, di una scelta molto dispendiosa, e al limite paralizzante, rispetto alle due possibilità che si offrono più semplicemente al lessicografo: o utilizzare i testi così come sono editi (scegliendo le edizioni migliori quando ce n'è più d'una), lasciando al futuro i problemi testuali, oppure selezionare drasticamente, limitandosi ai testi editi in modo soddisfacente, e affidando l'allargamento dei materiali a chi in futuro disporrà di nuove edizioni attendibili per i testi scartati. Ma è anche una scelta che rende il vocabolario profondamente innovativo, soprattutto per la parte relativa alla lingua antica, che è l'oggetto specifico del lavoro attuale.

Il terzo aspetto è la scelta informatica, che oggi sarebbe ovvia (sebbene ancora oggi il *GDLI* proceda con spogli manuali: cfr. Barberi Squarotti 1994), ma non era affatto tale nel 1965, anzi diede luogo ad un lavoro pionieristico e allo sviluppo di metodi allora nuovi di spoglio informatizzato presso la sezione linguistica del C.N.U.C.E. di Pisa, poi divenuta l'Istituto di Linguistica Computazionale del CNR, sempre sotto la direzione di Antonio Zampolli.

Accennerò solo brevemente al fatto, importante per la storia dell'impresa, che lo sviluppo vertiginoso dell'informatica nei trent'anni trascorsi ha comportato lunghe e complesse fasi di riconversione dei materiali. Il progetto iniziale, messo in pratica fino all'inizio degli anni '80, prevedeva di codificare i testi su nastro; dai nastri si producevano per ogni testo le concordanze per forme grafiche, su tabulato; sul tabulato si procedeva alla lemmatizzazione, selezionando al tempo stesso un limitato numero di contesti da associare ai lemmi; con un successivo passaggio su nastro, si dovevano ottenere le concordanze lemmatizzate, di nuovo su tabulato; alla fine (ma a questa fase non si giunse se non in qualche limitato esperimento), dall'insieme delle concordanze lemmatizzate si dovevano ottenere le cosiddette schede-contesto, cioè le schede da fornire ai redattori, paragonabili per forma a quelle tradizionali, ma frutto di uno spoglio esaustivo. A parte il fatto che stava diventando obsoleto, con passaggi continui dai ricercatori che preparavano i testi ai tecnici che li dovevano elaborare, questo metodo produceva un'enorme quantità di materiale cartaceo, che lasciava prevedere tempi pressoché infiniti per la redazione. Con un lavoro

imprevedibilmente lungo e difficoltoso, ma finalmente compiuto, tutto il materiale è stato perciò riconvertito in forma adatta al lavoro interattivo con il personal computer (singolo o in rete) con i moderni programmi lessicografici (quale appunto il DBT di Picchi).

È un effetto dell'evoluzione dell'informatica il fatto che lo spoglio elettronico dei testi non è più soltanto un mezzo per fornire materiali alla redazione del vocabolario, concepito ancora in forma tradizionale, ma produce basi di dati interrogabili in modi sempre più complessi, che tendono a diventare potenti strumenti di ricerca lessicografica e storico-linguistica alternativi al vocabolario, anche se a mio parere non sostitutivi (tornerò su questo punto alla fine del mio intervento).

Il corpus del TLIO

Nonostante le possibilità di gestione dei dati fornite dall'informatica siano incomparabilmente superiori a quelle concepibili con schedature manuali, un vocabolario dell'italiano fino ai nostri giorni non può basarsi altro che su una selezione di testi, per quanto vasta essa sia; tale è stato perciò il programma di lavoro, fin quando questo è stato affrontato su tutto l'arco cronologico della lingua italiana (sebbene non si sia arrivati a definire il corpus nel suo insieme), e così tornerà ad essere quando sarà possibile riprendere in mano la sezione successiva al 1375. La sezione precedente, invece, è tale che, se si limita com'è logico lo spoglio ai testi editi, questi sono in numero relativamente gestibile. Per questa sezione, dunque, il programma è sempre stato quello di uno spoglio esaustivo, che prendesse in considerazione, con le parole della *Bibliografia* edita alla fine della direzione di Carlo Alberto Mastrelli (OVI 1992, p. IX)

tutti quanti i testi, letterari o documentari di qualsiasi genere, originali o volgarizzati dal latino o da altra lingua, toscani o di qualsiasi altra regione d'Italia (non però i sardi né i ladini), di cui si possa dare per certa o per altamente probabile la composizione in una data non posteriore alla morte del Boccaccio e di cui si possieda un'edizione a stampa.

La *Bibliografia* citata elenca perciò 2208 testi, per circa 20 milioni di occorrenze; avendo carattere ancora provvisorio (ne è infatti ora in corso una rielaborazione), cita anche testi inediti, facendo riferimento a edizioni annunciate che sono però probabilmente lontane dalla pubblicazione, e testi editi, ma in edizioni inattendibili, che perciò devono essere considerati come

inediti; di questi ultimi potranno restare nel corpus solo quelli tramandati da ms. unico per i quali ci sia convenienza scientifica a procedere a una collazione (si tratterà tipicamente di testi documentari). La *Bibliografia* non cita invece edizioni posteriori al 1987.

Una valutazione di massima fa pensare che il corpus definitivo per la redazione si possa aggirare intorno ai 18 milioni di occorrenze; ma soprattutto si terrà conto in futuro del fatto che l'esaustività *assoluta*, nel costituire un corpus lessicografico, è uno scopo fuorviante: quella che conta è l'esaustività *tendenziale*, col criterio di assicurare la presenza di tutto ciò che è significativo (per data, per genere, per lingua), e di ampliare il corpus intorno a questo nucleo quanto più è ragionevolmente possibile.

Plurilinguismo e ordinamento dei dati

Al di là dei numeri, è essenziale notare che il corpus del *TLIO* è plurilingue; l'obiettivo è un tesoro della lingua italiana delle origini piuttosto che delle origini della lingua italiana, che dia documentazione lessicografica di tutte le lingue che hanno raggiunto la dignità della scrittura in Italia prima della costituzione di una lingua letteraria nazionale e accanto ad essa, mentre è evidente che la sezione successiva, dal 1375 in poi (fermo restando che il preciso limite cronologico è convenzionale), non potrà essere che monolingue. La stessa definizione di un italiano 'monolingue' è tutt'altro che semplice anche per il presente e per la stessa tradizione letteraria tra il Cinquecento e la fine dell'Ottocento; meno pacifico ancora è indicare come dovrebbero essere selezionati i testi dei primi secoli classificabili come fase antica della lingua nazionale: la scelta dei soli fiorentini sarebbe troppo ristretta (come escludere, per es., i due grandi aretini, Guittone per la poesia e Restoro per la prosa scientifica?); l'allargamento ai soli toscani introdurrebbe varietà linguistiche che dalla lingua nazionale sono poi rimaste fuori, comportando però ancora esclusioni poco ammissibili (che dire di Iacopone?). Se le cose si vedono non dall'oggi verso il passato, ma nel loro procedere dalle origini alla fine del Trecento, è logico che il *TLIO* sia plurilingue, perché plurilingue è la cultura che si esprime in testi scritti nei primi secoli, 'romanza' all'inizio e solo progressivamente 'italiana'. L'obiezione ricavabile da Lepschy 1979, p. 145 (che non si riferiva direttamente al *TLIO*):

è difficile pensare a buoni argomenti linguistici a favore di un vocabolario che mescoli insieme le parole milanesi di Bonvesin, quelle romane della *Vita di Cola*, e quelle siciliane di *Pir meu cori*

vale piuttosto a ricordare che un vocabolario plurilingue dell'italiano antico presenta problemi particolari, diversi da un lessico etimologico (punto di riferimento per l'italiano, oggi, il *LEI* di Pfister) che ordina i dati per basi etimologiche. Ma quanto al milanese Bonvesin, si noti che il *GDLI* lo cita, come cita anche, per es., il poemetto veronese *Della caducità della vita umana*, 'mescolando' effettivamente, per usare l'espressione di Lepschy, le forme milanesi e veronesi a quelle italiane. Si veda sotto *radice*, dove le forme *radici*, *radicie*, *radisce*, *radise*, *radixe*, *radize*, *raìs*, *raìse*, *raìxe*, *reìxe* sono tutte insieme qualificate come 'antiche', ma sono in parte forme appartenenti a lingue diverse dal fiorentino e dal toscano, e dove si trovano per es. citazioni da Bonvesin (*De Maria Egyptiaca*, v. 380, ed. Contini 1960; la parentesi sostituisce il segno di espunzione):

D'erb(e) crude e de radise la vita sustentava

e dalla *Caducità* (vv. 33-35, ed. Contini 1960):

Mo qual sia la raìs e la somença
là o' la toa miseria se comença
eo te l'ò dir...

Non è il caso, naturalmente, di creare appositi lemmi lombardi e veneti, da accostare ai lemmi toscani, perché ciò significherebbe lavorare alla compilazione di molti vocabolari 'dialettali' anziché ad un solo vocabolario della lingua delle origini; poiché questo è pensato come la prima parte di un vocabolario fino ai giorni nostri, tutte le forme corrispondenti per etimologia a forme della lingua moderna sono riportate al lemma moderno, mentre il lemma antico toscano o il lemma non toscano è riservato ai casi in cui non ci sia corrispondenza con la lingua moderna (è previsto naturalmente un indice delle forme, sostituibile con la consultazione per forme della base di dati). Forme come *radise* e *raìs* sono lemmatizzate sotto *radice* anche nel *TLIO*, ma affrontando il problema di fornire al lettore un orientamento preciso attraverso la pluralità delle lingue nella bibliografia, nella base di dati e nelle voci progettate.

Nella bibliografia ogni testo è caratterizzato con due indicazioni linguistiche, di area specifica e di area generica, per lo più regionale (rispettivamente, nel caso di Bonvesin, mil[anese] e lomb[ardo]), che occupano due campi della scheda informatizzata.

Nei casi in cui un'indicazione di area specifica non può essere data, entrambi i campi contengono l'indicazione di area generica (per es. i volgarizzamenti di Andrea da Grosseto sono classificati come toscani in entrambi i campi). Nella forma prevista (non ancora implementata) la base di dati può essere interrogata nel suo insieme, oppure selezionando i soli testi che rispondono a una stessa area specifica o generica; in questo secondo caso essa si scinde virtualmente in più basi di dati relative a singole aree linguistiche (si aggiunga che è sempre possibile consultare anche un solo testo).

Nel modo di funzionamento previsto come normale (per *default*), la base di dati fornisce tutti i contesti in cui compare una forma, in ordine cronologico, senza distinguere le aree linguistiche; nella consultazione per lemmi (non prevista in *ItalNet*, ma nelle altre versioni in cui i materiali saranno resi disponibili e sono già ora disponibili in sede), tutti i contesti selezionati per le forme riportate a un dato lemma, di nuovo senza distinzione di area linguistica. Ad ogni contesto viene però associata un'abbreviazione bibliografica che contiene la data del testo e l'indicazione dell'area linguistica specifica (che può, come ho detto, coincidere con quella generica); in questo modo la consultazione rende subito evidente la distribuzione cronologica e areale delle forme e dei lemmi.

Il sistema di abbreviazioni che si va elaborando contempla i seguenti tipi:

1. «autore, [opera,] data (lingua)». L'indicazione dell'*opera* può essere omessa quando non sono possibili ambiguità. I volgarizzatori (dal latino) e i traduttori (dalle lingue contemporanee) sono considerati autori (e come tali figurano col nome nel relativo campo della scheda bibliografica informatizzata), in quanto la ragione per cui il testo compare nella bibliografia dei citati è il fatto che esso è tradotto; l'autore del testo originale figura nel campo testo (anch'esso raggiungibile dall'interrogazione). I numeri romani designano come d'uso il secolo, con «sec.» sottinteso (così anche in tutti gli esempi successivi).

Bono Giamboni, *Vizi e Virtudi*, a. 1292 (fior.)

Bono Giamboni, *Il Libro de' Vizi e delle Virtudi*, in *Il Libro de' Vizi e delle Virtudi e Il Trattato di Virtù e Vizi*, a c. di C. Segre, Einaudi, Torino 1968, pp. 3-120.

Brunetto Latini, *Retto.*, c. 1260-61 (fior.)

Brunetto Latini, *La Rettorica*, a c. di F. Maggini, Le Monnier, Firenze 1968.

Panuccio del Bagno, XIII s.m. (pis.)

Panuccio del Bagno, *Rime*, ed. critica a cura di F. Brambilla Ageno, Accademia della Crusca, Firenze 1987.

2. «*opera, data (lingua)*» (per le opere di anonimo o anonimi)

Laudario S.M. d. Scala, XIII ex./XIV p.q. (tosc.)

Laudario di Santa Maria della Scala, ed. critica a cura di R. Manetti, Accademia della Crusca, Firenze, 1993.

Sonn. ann. Vat.Lat. 3793, XIII /XIV (tosc.)

Sonetti anonimi del Vaticano Lat. 3793, ed. critica a cura di P. Gresti, Accademia della Crusca, Firenze 1992.

3. Per i testi documentari (*Doc.*) e legislativi (*Stat.*) e per le lettere di carattere privato (*Lett.*), i cui autori non possono essere confusi con autori di testi letterari, si ritiene pertinente per l'utente, al momento della consultazione, sapere che si tratta di un testo di questo tipo, di una certa lingua o area linguistica e di una certa data; di quale testo effettivamente si tratti si potrà sapere consultando la lista delle abbreviazioni o richiamando la scheda bibliografica. Quando *lingua* e *data* coincidono per più testi, si aggiungerà in parentesi un numero d'ordine, col solo scopo di identificare il testo nella lista (senza applicare criteri rigidi, che renderebbero poi problematico l'inserimento di nuovi testi nella base di dati; non bisogna dimenticare che questa è sempre aperta). Fanno eccezione alcuni testi documentari più opportunamente citabili per nome d'autore, come per es. il senese *Libro di Mattasalà di Spinello* (per il quale si potrà fruire 'in anteprima' del testo critico di Arrigo Castellani), che si citerà «Mattasalà, 1233-43 (sen.)». Di regola, perciò:

«*Doc. lingua, data [(numero d'ordine)]*» per tutti i tipi di testi documentari (che nella scheda bibliografica sono poi distinti: documenti in genere = doc., documenti giuridici = doc. giur., documenti mercantili = doc. merc.):

Doc. aret., 1240

Affitti della badia di Santa Fiora d'Arezzo, in *La prosa italiana delle origini: I, Testi toscani di carattere pratico*, a cura di A. Castellani, Pàtron, Bologna 1982, pp. 159-62.

Doc. amiat., 1363

Contratti per vendita d' un «terrato» e d' una casa, in *Documenti inediti dell' amiatino tre-quattrocentesco*, a cura di S. Sbarra, in SFI, XXXIII (1975), pp. 15-188 [testo a pp. 96-97].

«Stat. lingua, data [(numero d'ordine)]» per tutti i tipi di testi legislativi:

Stat. pis., 1304

Breve dell' Arte della lana di Pisa, in *Statuti inediti della città di Pisa dal XII al XIV secolo*, a c. di F. Bonaini, vol. III, Vieuvesseux, Firenze 1857, pp. 647-49, 651-739. [Spoglio: pp. 651-739].

Stat. fior., a. 1364

Statuto dell' Arte dei vinattieri, in *Statuti delle Arti dei fornai e dei vinattieri di Firenze*, a c. di F. Morandini, Olschki, Firenze 1956, pp. 49-172. [Spoglio: pp. 49, 54-172].

«Lett. lingua, data [(numero d'ordine)]» per le lettere che non abbiano lo statuto di opera letteraria (i nomi dei mittenti si troveranno comunque nel campo «autore» della bibliografia informatizzata, ma tra parentesi graffe):

Lett. pist., 1321

Lettera di Amato da Empoli a Lazzarino, in L. Chiappelli, *Un carteggio di parte nera*, in BISI, XLIII (1925), pp. 1-74 [testo a pp. 39-40].

Lett. friul., 1358

Lettera di Mainardo di Villalta ad Artusino di Cividale sull' incendio della villa di Villalta fatto ai Signori di Uruspergo, in *Testi inediti friulani dei secoli XIV al XIX*, raccolti e annotati da V. Joppi, in AGI, IV (1878), pp. 185-342 [testo a p. 329].

Struttura delle voci del TLIO

Nella struttura delle voci, che si va definendo ora che la base di dati ha raggiunto una 'massa critica' sufficiente per una redazione almeno sperimentale, i contesti saranno ordinati per significati, e al

di sotto di ogni significato in ordine cronologico; la provenienza linguistica di ogni esempio sarà esplicitata dall'abbreviazione bibliografica, che la contiene, come detto, insieme con la data. I necessari chiarimenti sulla distribuzione geolinguistica di ogni forma saranno contenuti nell'intestazione premessa alla voce. Questa avrà la forma seguente:

ENTRATA, categoria grammaticale.

La categoria grammaticale sarà indicata con le sigle della *Grammatichetta della lingua italiana ad uso del calcolatore* di Piero Esperti (edita in Avalle 1979).

0.1. Varianti grafiche e/o fonetiche dell'entrata.

L'elenco completo di tutte le varianti grafiche, senza riguardo al loro effettivo valore, servirà all'utente per interrogare la base di dati, raggiungendo l'intero spoglio per forme, in tutti i casi in cui sia desiderabile una ricerca che vada al di là dei dati contenuti nella voce.

0.2. Etimologia.

Il *Tesoro* non è pensato come un lessico etimologico, ma l'etimologia funge da principale elemento di distinzione fra entrate lessicali omografe; sarà sufficiente il rinvio ai principali strumenti (primo fra tutti il LEI), e solo in casi particolari sarà opportuno aggiungere una breve nota.

0.3. Prima attestazione.

L'indicazione prima della voce è opportuna, perché l'articolazione per significati potrebbe non far cadere la prima attestazione al primo posto fra i contesti citati.

0.4. Note sulla distribuzione geolinguistica.

Stante l'articolazione per significati e in subordine cronologica, si indicheranno qui le prime attestazioni di ogni forma per ogni tipo linguistico, nel caso di voci di una certa ampiezza, o, secondo i casi, indicazioni riassuntive (del tipo: 'attestazioni solo toscane') ed eventuali altre note.

0.5. Note linguistiche.

Si indicheranno qui soprattutto le costruzioni sintattiche, in tutti i casi in cui la voce sia complessa, o se la sua articolazione impedisca di cogliere unitariamente particolarità sintattiche, o anche se in una voce breve sia opportuno mettere in evidenza una qualche particolarità.

0.6. Rinvii e altre osservazioni.

Si inserirà qui ogni altro tipo di osservazioni, e in particolare (almeno nella prima stesura) si faranno presenti le integrazioni

di rilievo, soprattutto retrodatazioni, rispetto agli strumenti d'uso. Sempre nella prima stesura questo è il luogo in cui saranno evidenziati dubbi e problemi.

0.7. Struttura della voce (per voci complesse, o anche solo 'lunghe').

Ad eccezione di quanto può avvenire per le voci di scarsa attestazione, per le quali si potrà dare la documentazione completa, per tutte le altre si dovrà procedere ad una selezione dei contesti da citare, che in certi casi potrà essere drastica. Tale selezione (controllabile sempre dall'utente attraverso la ricerca delle forme nella base di dati) seguirà alcuni criteri di massima, che si preciseranno meglio nel corso del lavoro.

(1) Si citerà sempre la prima occorrenza del lemma (non necessariamente di ogni forma riconducibile al lemma) in ogni area linguistica.

(2) Si eviterà di regola di citare, per lo stesso significato o accezione, più di un esempio tratto dallo stesso testo, a meno di particolari ragioni di interesse linguistico (ma senza la necessità di documentare, dallo stesso testo, forme diverse riconducibili allo stesso lemma).

(3) Si eviterà, per lo stesso significato o accezione, di citare esempi tratti da più testi dello stesso genere e della stessa area linguistica contemporanei o quasi contemporanei (stabilendo in linea di massima cadenze da 10 anni circa per i lemmi relativamente meno attestati fino a un massimo di 25 anni per quelli più attestati).

(4) A parità di altre condizioni, si darà la precedenza ai testi non letterari rispetto a quelli letterari, e a quelli classificabili in un'area linguistica specifica (per es. fiorentini) rispetto a quelli classificabili solo in un'area linguistica regionale (per es. toscani).

Vocabolari e basi di dati: un problema (non solo) del *TLIO*

L'aspetto per il quale il *TLIO* è una base di dati dell'italiano antico, non ancora completa ma in buona parte realizzata, prevale per ora su quello per cui il *TLIO* sarà o dovrà essere un vocabolario storico, conformemente al progetto iniziale: è in forma di base di dati che esso, da ora e per lungo tempo, renderà il suo servizio agli studiosi. Questo fatto è certo, prima di tutto, una conseguenza dell'andamento dei lavori, poiché la base di dati è necessaria alla redazione del vocabolario, e solo ora essa comincia a rispondere ai requisiti di dimensione e di consultabilità per poterla affrontare. C'è

però anche una ragione non contingente, legata ai metodi e agli strumenti di lavoro, per la quale un'impresa nata per produrre un vocabolario con l'aiuto dell'informatica si è trasformata abbastanza presto in un'impresa volta di fatto a produrre una base di dati, sulla quale si pone ancora il problema di produrre il vocabolario, e si è anche posto (qualcuno lo pone ancora) il problema se il vocabolario sia necessario, opportuno, o anche solo possibile. Una riflessione a questo proposito era già in interventi raccolti in un volume del 1979 con il quale d'Arco Silvio Avalle, allora direttore, voleva fare il punto della situazione dell'impresa (Avalle 1979). Si vedrà lì, per es. nella discussione seguita ad una presentazione del *TLIO* alla «Seconde table ronde de lexicographie historique» (Leida, 10-12 maggio 1977), come Avalle vedesse il vocabolario tradizionale messo in crisi dalle grandi masse di dati rese disponibili dall'informatica, e che solo gli strumenti informatici permettono di consultare vantaggiosamente, e come Bernard Quemada opponesse già allora una *dictionnaire*, finalizzata a produrre vocabolari, a una *lexicographie* non rivolta necessariamente al vocabolario, da giudicarsi implicitamente più moderna, più potente e scientificamente più interessante.

Il fatto è che la scelta informatica dava in primo luogo la possibilità, e perciò almeno apparentemente anche il dovere, di avviare l'impresa del nuovo vocabolario ripartendo dalle fondamenta, costituendo un grande corpus dal quale, attraverso il lavoro di lemmatizzazione, doveva scaturire un nuovo lemmario, fondato sui testi e indipendente dalla tradizione lessicografica. Dubito molto che per la parte successiva al 1375 si possa fare un vocabolario senza utilizzare, almeno per confronto, quelli esistenti (ma il progetto non avanzò abbastanza perché questo problema dovesse essere considerato); quanto al *TLIO*, questo metodo aveva (e ha) una forte giustificazione nella mancanza, alle spalle, di un vero vocabolario dell'italiano antico, che non risulti dall'estrazione delle citazioni antiche dai vocabolari generali, oltre che nella possibilità materiale di raccogliere all'incirca tutto l'edito.

In secondo luogo, la scelta informatica spinge irresistibilmente contro la selezione; e il fatto che nei vocabolari tradizionali questa fosse materialmente inevitabile non toglie che essa sia comunque indispensabile per qualunque tipo di vocabolario. Nei lavori del *TLIO* è stata attuata fin dall'inizio una strategia di riduzione, in fase di lemmatizzazione, consistente nell'associare ai lemmi, testo per testo, solo una limitata scelta di contesti, ma essa è stata ridotta da una strategia di documentazione, per la quale i lemmatizzatori sono stati tenuti a selezionare almeno un contesto per ogni forma da ogni

testo, producendo così, ad onta della riduzione, imponenti archivi di contesti lemmatizzati per forme correnti e ricorrenti, spesso anche banali; con un risultato senz'altro positivo, ma nella logica di una base di dati.

In questa logica, il *TLIO* è, forse involontariamente rispetto alle prime intenzioni, al passo con i tempi, nei quali l'evoluzione e la diffusione dell'informatica stanno rendendo le basi di dati sempre più appetibili, e la loro produzione sempre più rigogliosa. Ma, anche a rischio di parere autolesionistico, come direttore di un'impresa che è avanti su questo fronte e indietro sull'altro, devo rilevare che la lessicografia non si può esaurire nelle basi di dati, che danno enormi possibilità di analisi, ma lasciano la sintesi e l'interpretazione al lavoro dell'utente. È, piuttosto, importante prospettare una sinergia tra vocabolario e base di dati: l'accessibilità di questa libera quello da molti problemi di documentazione, dal timore di lasciar cadere esempi importanti, se questi sono comunque recuperabili, e permette di concentrare gli sforzi sulla definizione, l'articolazione e la storia dei significati, sull'interpretazione di cui l'utente ha bisogno per non rifare il lavoro ogni volta che accede agli ingenti materiali con cui la base di dati risponde ogni volta alle sue domande. Il *TLIO* al quale si dovranno ora rivolgere gli sforzi dell'OVI sarà perciò molto più 'leggero' rispetto alla forma che avrebbe dovuto avere senza la base di dati che ora lo rende possibile e poi lo affiancherà; sia che esso debba o possa avere la forma tradizionale del volume a stampa, sia che si presenti anch'esso come uno strumento informatico, dal quale il passaggio alla base di dati potrà essere anche automatico.

Ciò appartiene al futuro, del quale è piacevole, ma anche troppo facile parlare, tanto più che esso dipende in larga parte dal modo in cui saranno risolti i problemi di risorse che condizionano l'attività dell'OVI; potrà essere un argomento di discussione utile e istruttiva con gli studiosi non appena sarà disponibile la base di dati in rete, che appartiene anch'essa al futuro, ma è un programma di lavoro ormai prossimo alla realizzazione.

Bibliografia

- AA.VV. 1957 - *Per un grande vocabolario storico della lingua italiana*, Firenze, Sansoni.
- ARTFL - *American and French Research on the Treasury of the French Language*. A Cooperative Project, CNRS, Institut National de la Langue Française - The University of Chicago, Divisions of the Humanities and Social Sciences.

- Atti Crusca* = *La Crusca nella tradizione letteraria e linguistica italiana*. Atti del Congresso Internazionale per il IV centenario dell'Accademia della Crusca, Firenze, Accademia della Crusca, 1985.
- Atti TLIO* = *Atti della giornata di studi su "Il Tesoro della lingua italiana delle origini e le imprese lessicografiche in area italiana e romanza"*, Firenze, Villa di Castello, 9 giugno 1993, in «Bollettino» dell'Opera del vocabolario italiano, I, 1994, pp. 3-76.
- Avalle 1979 = d'Arco Silvio Avalle, *Al servizio del vocabolario della lingua italiana*, Firenze, Accademia della Crusca.
- Barberi Squarotti 1994 = Giorgio Barberi Squarotti, *Il Grande Dizionario della Lingua Italiana*, in *Atti TLIO*, pp. 5-11.
- Barbi 1935 = Michele Barbi, *Crusca Lingua e Vocabolari*, in «Pan», III, 9, pp. 13-24 (poi in AA.VV. 1957).
- Contini 1960 = *Poeti del Duecento*, a cura di Gianfranco Contini, Milano-Napoli, Ricciardi, 2 voll.
- Crusca 1612 = *Vocabolario degli Accademici della Crusca*, prima ed., Venezia.
- Crusca 1623 = *Vocabolario degli Accademici della Crusca*, seconda ed., Venezia.
- Crusca 1691 = *Vocabolario degli Accademici della Crusca*, terza ed., Firenze.
- Crusca 1729-38 = *Vocabolario degli Accademici della Crusca*, quarta ed., Firenze.
- Crusca 1863-1923 = *Vocabolario degli Accademici della Crusca*, quinta ed., A-O, Firenze.
- De Robertis 1985 = Domenico De Robertis, *L'ufficio filologico dell'Opera del vocabolario, il suo impianto, il suo lavoro*, in *Atti Crusca*, pp. 443-51.
- Fiorelli 1979 = Piero Fiorelli, *L'Accademia della Crusca per il Vocabolario giuridico italiano*, in «Studi di Lessicografia Italiana», I, pp. 55-81.
- GDLI* = Salvatore Battaglia, *Grande Dizionario della Lingua Italiana*, Torino, UTET, I A-BALB 1961, II BALC-CERR 1962, III CERT-DAG 1964... XVII ROBB-SCHI 1994.
- Lepschy 1979 = Giulio C. Lepschy, *Lessico*, in *Enciclopedia Einaudi*, VIII, pp. 129-51.
- LIZ* = *Letteratura Italiana Zanichelli. CD-ROM dei testi della letteratura italiana*, a cura di Pasquale Stoppelli e Eugenio Picchi, Bologna, Zanichelli, 1993.

- Nencioni 1955 = Giovanni Nencioni, *Relazione all'Accademia della Crusca sul vocabolario della lingua italiana*, in «Studi di Filologia Italiana», XIII, pp. 395-420 (poi in AA.VV. 1957).
- OVI 1992 = *Bibliografia dei testi in volgare fino al 1375 preparati per lo spoglio lessicale*, Firenze 1992.
- Parodi 1974 = Accademia della Crusca, *Gli atti del primo vocabolario* editi da Severina Parodi, Firenze, Sansoni.
- Parodi 1983 = Severina Parodi, *Quattro secoli di Crusca, 1583-1983*, Firenze.
- Pasquali 1941 = Giorgio Pasquali, *Per un Tesoro della lingua italiana*, in «Atti della R. Accademia d'Italia. Rendiconti della Classe di scienze morali e storiche», s. 7a, II, pp. 490-521 (poi in AA.VV. 1957).
- Peri 1990 = M[aria] S[erena] P[eri], *Crusca (Accademia della)*, in *Letteratura italiana* dir. da Alberto Asor Rosa, *Gli Autori. Dizionario bio-bibliografico e Indici*, I, Torino, Einaudi.
- Pfister 1994 = Max Pfister, *Il LEI (Lessico etimologico italiano): una sfida scientifica e operativa*, in *Atti TLIO*, pp. 13-26.
- Sessa 1991 = Mirella Sessa, *La Crusca e le Crusche: il vocabolario e la lessicografia italiana del Sette-Ottocento*, Firenze, Accademia della Crusca.
- Vitale 1959 = Maurizio Vitale, *La prima edizione del Vocabolario della Crusca e i suoi precedenti teorici e critici*, in *I primi vocabolari delle grandi lingue europee*, a c. di Antonio Viscardi, I. *Le lingue romanze*, Milano-Varese, pp. 27-74 (ora in Vitale 1986).
- Vitale 1966 = Maurizio Vitale, *La III edizione del «Vocabolario della Crusca» Tradizione e innovazione nella cultura linguistica fiorentina secentesca*, in «Acme», XIX, pp. 109-53 (ora in Vitale 1986).
- Vitale 1971 = Maurizio Vitale, *La IV edizione del Vocabolario della Crusca. Toscanismo, classicismo, filologismo nella cultura linguistica fiorentina del primo Settecento*, in *Studi di filologia romanza offerti a Silvio Pellegrini*, Padova, Liviana, pp. 675-704 (ora in Vitale 1986).
- Vitale 1986 = Maurizio Vitale, *L'oro nella lingua. Contributi per una storia del tradizionalismo e del purismo italiano*, Milano-Napoli, Ricciardi.
- Zolli 1989 = Paolo Zolli, *Points noirs dans la lexicographie des langues romanes: l'italiano*, in *Actes du XVIII^e Congrès International de Linguistique et Philologie Romanes* (Trèves 1986), IV, Tübingen, Max Niemeyer, pp. 28-32.

CINCUENTA AÑOS DE DICCIONARIOS MONOLINGÜES VOX

MANUEL ALVAR EZQUERRA
Universidad de Málaga

Uno de los diccionarios señeros del español contemporáneo es el *Diccionario General Ilustrado de la Lengua Española* (Vox)¹, desde que apareció en 1945². Se cumplen, pues, ahora los cincuenta años de su publicación. No es ese diccionario una obra que haya permanecido sin cambios a lo largo de tanto tiempo, pues ha tenido varias ediciones, supervisadas por don Samuel Gili Gaya³ en 1953⁴ y en 1973 -con varias reimpressiones-, y otra más que ha aparecido en 1987, realizada bajo mi dirección⁵.

¹ Citado en adelante como DGILE.

² Editado en Barcelona por Publicaciones y Ediciones Spes, s. a.

³ La figura de don Samuel es de sobra conocida en la filología española. De todas maneras, véase Ángel López García, «Gili Gaya: un eslabón en la historia de la lingüística española», apud Jesús Costa (coord.), Ángel López y Dolors Sistac, *Samuel Gili Gaya. Vida y obra (1892-1976)*, Texto e Imagen, Lérida, 1991, págs. 63-82; y, para otros aspectos, Dolors Sistac y Sanvicén, *Samuel Gili i Gaya. Notes per a una bibliografia*, IEI, Lérida, 1976, y de la misma autora, «Notes biogràfiques», apud Jesús Costa (coord.), Ángel López y Dolors Sistac, *op. cit.*, págs. 7-44.

⁴ A partir de esta edición lo publica Bibliograf, s. a.

⁵ Además de las ediciones reseñadas, el diccionario ha tenido numerosas reimpressiones, algunas de las cuales se han anunciado por la editorial como ediciones diferentes. Sobre la salida de 1987, véanse las reseñas de Marta C. Ayala Castro, *Revista de Filología Española*, LXVIII, 1988, págs. 169-172; Ramón Cerdá, «Vox. Diccionario General Ilustrado de la Lengua Española», *Boletín de la Asociación Europea de Profesores de Español*, 36-37, 1989, págs. 209-214; David Mighetto, «BIBLOGRAF S. A., *Diccionario General Ilustrado de la Lengua Española* (VOX87)», *Moderna Språk*, LXXXII-2, 1988, págs. 175-177, y *Revista Española de Lingüística*, 19-1, 1989, págs. 191-194; Ignacio Ahumada, «El nuevo diccionario VOX (dgile)», *Revista española de lingüística aplicada*, V, 1989, págs. 197-202; María José Quilis Sanz, *Revue de Linguistique Romane*, 209-210, enero-junio 1989, págs. 225-228; y Ángel Manteca Alonso-Cortés, «Las palabras en el Vox», *Libros*, n° 25, marzo 1984, pág. 9.

La autoría del DGILE ha sido, y es, prácticamente desconocida para los especialistas⁶, pues nunca se ha anunciado públicamente. Y no resulta fácil buscar por unos archivos editoriales que han padecido el paso de los años y los cambios de domicilio. He conseguido averiguar que la editorial poseía un diccionario que había encargado a algunos profesores de enseñanza media. Para revisar su contenido los editores se dirigieron a don Ramón Menéndez Pidal, quien no pudo hacerse cargo de la tarea -luego escribiría el prólogo-, pero recomendó que lo hiciese Gili Gaya, a quien, finalmente, se encargó la corrección. Don Samuel finalizó su tarea un año antes de lo que se había comprometido con el editor, si hacemos caso a los datos que constan en un contrato firmado con fecha de doce de abril de 1944, en el que se dice que estaría lista antes de finales de 1946, a la vez que se obligaba también a redactar un diccionario para la Escuela Media antes del 31 de enero de 1945. El *Diccionario General* apareció en 1945, pero el otro vino más tarde, en 1954. Ésa es la razón por la que el nombre de don Samuel nunca ha figurado como autor ni en la portada de la obra, ni en los diversos contratos que suscribió, porque el original no era suyo, por más que después quedara de forma permanente su huella en el diccionario y en la editorial. Sí se reconocía, por el contrario, que era el autor de la revisión, además de firmar las «Características del Diccionario VOX», enunciado en el que quiero ver cierta distancia con respecto a la obra. Es más, cuando se preparaba la tercera revisión del DGILE, pareja a otra del *Diccionario Manual*, Gili Gaya redactó unas «Bases de revisión i ampliació dels diccionaris Vox General i Manual»⁷ en las que puede leerse: «El *Manual* no me'l caldrà revisar personalment; però la quantitat susdita comporta l'autorització a la casa editora per que'l meu nom figuri al *Manual* com a revisor, i per a reproduir el pròleg amb els retocs que calguin per a actualitzar-lo».

El DGILE va precedido de un prólogo de don Ramón Menéndez Pidal, «El diccionario que deseamos», el único trabajo teórico de don Ramón dedicado a la lexicografía, y que ha sido examinado en varias ocasiones por lo que supone de contribución para el progreso de la disciplina tanto desde el punto de vista teórico como del práctico⁸. Si-

⁶ Cfr., por ejemplo, Manuel Seco, «Medio siglo de lexicografía española (1930-1980)», *Revista de Bachillerato*, 10, abril-junio de 1979, págs. 2-7, después recogido en los *Estudios de lexicografía española*, Paraninfo, Madrid, 1987, págs. 194-220, y en especial la pág. 203.

⁷ Con fecha de 31 de enero de 1965.

⁸ Por ejemplo, David L. Gold, «New Perspectives in Spanish Monolingual Lexicography», *Babel. Revue internationale de la traduction*, 21, 1975, págs. 69-77; Gregorio Salvador, «Don Ramón Menéndez Pidal y el Diccionario», *Boletín de la Real*

guen a esas páginas otras de Gili Gaya en las que se explican las «Características del Diccionario VOX»⁹. El diccionario se completa con una lista de nombres geográficos y gentilicios y otra de voces y locuciones latinas y extranjeras.

Durante sus cincuenta años de existencia, el diccionario ha sufrido algunos cambios o modificaciones en su forma, como el del apéndice de americanismos que había en la primera entrega, y que a partir de la segunda se incorporó al cuerpo de la obra en un proceso similar al que sucedió con el *Vocabulario de xerigonza* de Juan Hidalgo en el *Tesoro* de César Oudin. En esa segunda edición no sólo se incluyeron los americanismos, sino que se añadían nuevas voces americanas, haciendo del DGILE el diccionario general más rico en términos y acepciones propios de aquel continente. En su conjunto, el DGILE ha pasado de algo más de 58000 entradas en la primera edición, a algunas más de 72000 en la segunda, a las aproximadamente 80000 de la tercera, y a las 89165 de que consta ahora, el mayor número en un repertorio de sus características.

Manuel Seco¹⁰ ha explicado con gran claridad cómo es el contenido de la obra: «*Vox* ha tratado con bastante libertad la nomenclatura académica: ha eliminado numerosas voces anticuadas y, en contrapartida, ha incorporado numerosas voces vivas, parte de ellas americanismos, parte palabras procedentes de diversos ámbitos científicos y técnicos (medicina, biología, química, filosofía, mecánica, industria, etc.) que circulan hoy de manera generalizada en la lengua culta. En cambio [...] ha sido bastante sobrio en la distribución de acepciones de cada artículo: no desmonta ni desdobra las acepciones académicas, limitándose a abreviar y modernizar -con discreción y acierto notables- los enunciados clásicos; lo cual no impide que añada, cuando procede, acepciones modernas no académicas, y, sobre todo, que introduzca la novedad de colocar los sintagmas binarios y locuciones a continuación de la acepción con la que por lógica están emparentados, y no almacenarlos alfabéticamente a la cola del artículo, según la práctica tradicional.» No es de extrañar, pues, la actualidad y la implantación de la obra a pesar de los años transcurridos desde su primera salida.

Entre las innovaciones lexicográficas, tal vez la más importante sea la del empleo de unos corchetes para indicar el contorno de la defi-

Academia Española, LXVIII, 1988, págs. 369-373; y Manuel Alvar Ezquerro, *Lexicología y lexicografía. Guía bibliográfica*, Almar, Salamanca, 1983, págs. 25-29.

⁹ A partir de la tercera edición se llamarán «Características de este diccionario».

¹⁰ «Medio siglo de lexicografía», págs. 203-204.

nición, de todo aquello que sin pertenecer a ésta es necesario para que pueda ser entendida, es decir, la parte de la definición que resulta necesaria en el predicado definicional para hacerla comprensible y exacta, pero que no se puede emplear al sustituir el definido por la definición, esto es, lo que no son semas de la voz en cuestión¹¹. La indicación del ámbito de empleo no es un hallazgo de los diccionarios Vox, aunque sí su extensión y explotación sistemática. Los antecedentes hay que buscarlos ya en el *Dictionnaire général de la langue française* de Hatzfeld y Darmesteter¹² según señaló Rey-Debove¹³, y para lo que respecta a la lexicografía del español coetáneo, e incluso anterior, en el *Diccionario de construcción y régimen*¹⁴ de Rufino José Cuervo como han expuesto Porto Dapena¹⁵ e Ignacio Ahumada¹⁶. Para nuestra lengua, según este último investigador, la información del contorno no fue atendida hasta que Gili Gaya se hizo cargo de la revisión del DGILE¹⁷, si bien Humberto Hernández apunta que la lexicografía escolar española utilizó la indicación del contorno en el *Compendio del diccionario de la lengua castellana, basado en el de la Real Academia, para uso de la infancia*¹⁸ de Aladro Hernández de Padilla¹⁹. Es necesario decir, por otra parte, que para el catalán ya señaló el contorno, entre paréntesis, Pompeu Fabra

¹¹ Véase a este propósito Manuel Seco, «El «contorno» en la definición lexicográfica», en *Homenaje a Samuel Gili Gaya (in memoriam)*, Bibliograf, Barcelona, 1979, págs. 183-191, ahora recogido en los *Estudios de lexicografía española*, págs. 35-45; del mismo autor, cfr. «Medio siglo de lexicografía española», pág. 204.

¹² 2 vols., París, Delagrave, 1889-1900.

¹³ Josette Rey-Debove, *Étude linguistique et sémiotique des dictionnaires français contemporains*, Mouton, La Haya-París, 1971, pág. 211.

¹⁴ 2 vols. (hasta la letra D), París 1886-1893. Ha habido una reimpresión hecha por el Instituto Caro y Cuervo, Bogotá, 1953. La obra ha continuado publicándose por fascículos hasta su culminación y presentación oficial en 1995.

¹⁵ José Álvaro Porto Dapena, *Elementos de lexicografía*. El Diccionario de construcción y régimen de R. J. Cuervo, Instituto Caro y Cuervo, Bogotá, 1980, pág. 64.

¹⁶ Ignacio Ahumada Lara, «Contorno de la definición verbal y régimen lexemático: su indicación formal en la lexicografía hispánica», en *Amistad a lo largo. Estudios en memoria de Julio Fernández-Sevilla y Nicolás Marín López*, Universidad de Granada, Granada, 1987, págs. 13-25, y en especial la pág. 17 y sigtes. Es más, para Ahumada «la indicación formal del contorno de la definición o régimen lexemático tiene sus orígenes en la *Muestra [de un Diccionario de la lengua castellana]* (1871) de Cuervo y Manrique» (pág. 24).

¹⁷ Cfr. Ignacio Ahumada Lara, «Información gramatical implícita en la definición lexicográfica», *Thesaurus*, XLIII, 1988, págs. 81-94, en especial la última página del trabajo.

¹⁸ Madrid, 1911.

¹⁹ Cfr. Humberto Hernández, «¿Son escolares los diccionarios escolares?», *Apuntes de educación. Lengua y literatura*, 39, octubre-diciembre 1990, págs. 10-12, y en especial las págs. 10-11.

en el *Diccionari General de la Llengua Catalana*²⁰, publicado por vez primera en 1932 -aunque comenzado a redactar varios años antes-, seguramente por haber tenido delante el de Hatzfeld y Darmesteter, según consta en el prefacio de la primera edición. Estoy convencido de que esa marcación del diccionario de Fabra tuvo mucho que ver, en sus orígenes, con la que hay en el DGILE.

Otra de las características del DGILE ha sido la gran atención prestada a los usos regionales, y muy en especial los del español americano, como ha quedado expuesto más arriba. Yo mismo escribí hace algún tiempo que

El tratamiento de los dialectalismos y americanismos ha sido distinto en las diversas ediciones de la obra, pues en la primera se prescindió de las palabras y acepciones de uso local en España, mientras que los americanismos figuraron en un apéndice. A partir de la segunda edición se da cabida a muchos dialectalismos y se incorporan los americanismos, en unas cantidades nada despreciables, y en ocasiones superiores al diccionario oficial de la lengua, prescindiendo de las voces o acepciones de reducida extensión geográfica, y las vulgares, jergales e incorrectas, pues favorecen la fragmentación del idioma.²¹

En la última edición de la obra, como explicaré más adelante, todavía se ha prestado una mayor atención a esta parcela del vocabulario, de modo que «ningún otro repertorio léxico se nos muestra con mayores méritos en el tratamiento de lo diatópico del español que el *Diccionario general ilustrado de la lengua española*»²².

En el prólogo que puso Gili Gaya al diccionario explicaba, entre otras cosas, que se prescindía de los arcaísmos. La decisión es consecuente con el carácter moderno del DGILE. Sin embargo, se mantienen las palabras que ya han desaparecido y tenían un uso frecuente en la literatura clásica, lo que hace que no sean raras las voces o acepciones con la marca que señala su carácter anticuado.

El DGILE es un diccionario rico en nombres propios, cuyos artículos, en ocasiones, proporcionan informaciones enciclopédicas, sin que se pueda afirmar que estamos ante un diccionario enciclopédico, por más que se cumpla alguna de las características de estos. La presen-

²⁰ Cfr. Germà Colon y Amadeu-J. Soberanas, *Panorama de la lexicografia catalana. De les gloses medievals a Pompeu Fabra*, Enciclopèdia Catalana, Barcelona, 1986, pág. 224.

²¹ Manuel Alvar Ezquerro, *Lexicología y lexicografía. Guía bibliográfica*, Almar, Salamanca, 1983, pág. 191.

²² José Luis Aliaga Jiménez, «Reflexiones sobre el caudal léxico de un diccionario de lengua: notas a propósito de los regionalismos en los diccionarios de la lengua española», *Stvdium. Filología* (Teruel), 7, 1991, págs. 25-38; la cita es de la pág. 34.

cia de muchos nombres propios se debe a la necesidad de explicar expresiones que los contienen y que se consignan en la obra, pero otras veces figuran por seguir una larga tradición en la confección de diccionarios.

Representó una novedad cuando apareció el DGILE la inclusión de prefijos y sufijos, y de elementos compositivos, como formas combinatorias de muy fecundo papel en la creación de léxico culto, especialmente científico²³. Aunque inicialmente las palabras formadas con ellos aparecían como subentradas del encabezamiento, a partir de la edición de 1987 constituyen entradas independientes tanto los prefijos y sufijos como las voces construidas a que dan lugar. El aumento de estos elementos y de las voces formadas con ellos ha sido continuo desde la edición de 1945, habiéndose pasado de los 727 formantes consignados entonces a los 757 (922 en el DALE, del que me ocuparé enseguida), mientras que el DRAE registraba 24 en 1947, 115 en 1984, y 188 en 1992. El número de palabras construidas con esos formantes es de 10457 en el DGILE (12401 en el DALE), frente a 6340 en el DRAE de 1984 y 7088 en la edición de 1992²⁴.

El aspecto didáctico de este diccionario se deja ver en la inclusión de sinónimos -y en menor medida de antónimos- explicados²⁵ e informaciones gramaticales de diversa índole en algunos artículos, hecho muy poco frecuente en nuestra lexicografía, por más que la tendencia se esté invirtiendo en los últimos tiempos, tal vez por influencia del DGILE, tal vez porque el interés didáctico de los diccionarios se está afianzando. Además, en el diccionario que nos ocupa, se abordan no pocas cuestiones de metalenguaje en los cuadros que hay distribuidos a lo largo de la obra. Esos cuadros que eran de variada índole en las primeras ediciones lo son casi exclusivamente de aspectos lingüísticos en la última, pues se ha ido prescindiendo de lo superfluo y extralingüístico a lo largo de las salidas del DGILE (quedan los cuadros dedicados a la poesía y la versificación clásica, así como al año, la Biblia, la Geología, la Historia y las Medidas, de entre los 41 que contiene; han desaparecido, por ejemplo, el del Automóvil con las contraseñas de las matrículas y el de la Familia).

Más rara es, incluso, la presencia de ilustraciones en nuestros diccionarios generales, de las que hace buen acopio el DGILE. Como

²³ Cfr. Manuel Seco, «Diccionarios: la realidad y el deseo», *Saber Leer*, 47, Agosto-Septiembre 1991, págs. 6-7; la cita procede de la pág. 6.

²⁴ Cfr. Manuel F. Pérez Lagos, *La formación de palabras: la composición culta en los diccionarios (DRAE-Vox)*, tesis doctoral presentada en la Universidad de Málaga, 1995, págs. 961 y 974-975.

²⁵ Manuel Seco, «Diccionarios. la realidad y el deseo», citado, pág. 6.

en otras partes de la obra, el contenido no es anecdótico o puramente ornamental: está destinado a satisfacer las necesidades del usuario y a guiarlo hacia lo que busca o precisa; por ello, las ilustraciones del DGILE son de carácter sistemático, y no meramente decorativas. En este sentido es necesario advertir que de una edición a la siguiente se han ido quitando unas ilustraciones, se han sustituido otras, algunas han sido modernizadas, añadidas otras, adecuando la obra a la imagen real, actual, de lo representado. En ocasiones, el cambio intenta sistematizar lo ilustrado, ofreciendo al que acuda a la obra simultáneamente la posibilidad de ir desde la cosa nombrada a la palabra que sirve para designarla, aunque, es cierto, en un número muy limitado de ocasiones, ya que esta función pertenece a otro tipo de obras.

El DGILE, en sus versiones anteriores y en la actual, «como la gran mayoría de sus congéneres, está concebido únicamente para los hablantes nativos o al menos para quienes ya dominan muy a fondo las picardías de la lengua. Para ser más exactos habría que decir que está redactado desde la perspectiva del hablante nativo culto que rara vez se pone en el lugar del observador externo o ingenuo.»²⁶

El carácter didáctico de la obra queda patente por ese «empleo funcional de la imagen por el sistema de la ilustración en constelación, como medio no sólo de precisar, por contraste, la comprensión de los nombres de objetos materiales, sino también de enriquecer las asociaciones semánticas en la mente del usuario.»²⁷

Son las características expuestas, y el análisis efectuado sobre la obra, las que me llevaron a afirmar antes de iniciar la última revisión del diccionario que se trataba de una de las mejores y más completas obras lexicográficas del español actual, y que «la presencia de los cuadros gramaticales, las explicaciones de los términos metalingüísticos, la inclusión de nombres propios, paradigmas y otras cuestiones gramaticales, y las listas finales de nombres geográficos y gentilicios así como de voces y locuciones latinas y extranjeras, junto a la abundancia de ilustraciones, hacen de él algo más que un diccionario.»²⁸

Una obra que se hubiera dormido en los laureles pasados habría tardado muy poco en envejecer. Por ello, la editorial que publica el DGILE se ha preocupado en irlo actualizando y adecuando a la realidad y necesidades de cada momento. La última revisión apareció en

²⁶ Ramón Cerdá, «Vox. Diccionario General Ilustrado de la Lengua Española», citado, pág. 211.

²⁷ Manuel Seco, «Diccionarios: la realidad y el deseo», citado, pág. 6.

²⁸ Manuel Alvar Ezquerro, *Lexicología y lexicografía. Guía bibliográfica*, citado, pág. 197.

1987 realizada bajo mi responsabilidad. Se había hecho necesario acometerla porque no estaban consignados muchas palabras y significados nuevos, y porque los estudios sobre la lengua y sus variedades regionales, los progresos científicos y técnicos, la renovación del lenguaje juvenil, etc., habían sufrido cambios de los que era obligado dejar constancia. Ahora, todas las modificaciones producidas en su interior nos proporcionan una obra nueva, aunque siguiendo los principios que ha mantenido durante cincuenta años, y conservando mucho de lo que tiene desde entonces, pues lo bueno bien está y no hay por qué cambiarlo.

En otro lugar²⁹ he expuesto en qué consistió la revisión y aumento del DGILE para la salida de 1987, en los que se tuvieron bien presentes los principales rasgos del diccionario que lo habían convertido en una pieza valiosa en el interior de nuestra lexicografía. Es una obra de carácter selectivo, no total, que pretende reflejar la Lengua Española de los días que nos ha tocado vivir, y que está concebido por y para hablantes nativos³⁰. Por el respeto a la tarea realizada, y teniendo en cuenta las propiedades del diccionario, es por lo que en su estado actual no hay arcaísmos, voces anticuadas, desusadas, etc., que no figurasen ya en anteriores salidas, aunque sí recoge voces que designan objetos, conceptos, etc., caídos en desuso: la cosa nombrada puede haber desaparecido, pero no el término que sirve para nombrarla. Tampoco tiene variantes fonéticas, sin interés léxico, pues su estudio y recapitulación pertenece a los tratados de fonética, a los atlas lingüísticos, o a las monografías sobre hablas vivas.

En la última entrega del diccionario figuran con la forma femenina también las designaciones de muchas actividades y profesiones que sólo aparecían en masculino, pues hasta época muy reciente la mujer no había tenido acceso a ellas: las circunstancias ya no son las mismas, y el diccionario no puede permanecer ajeno a los cambios sociales que se producen en su entorno. De este modo figuran tanto en masculino como en femenino *biólogo*, *-ga*, *gemólogo*, *-ga*, *tecnólogo*, *-ga*, *urólogo*, *-ga*, y tantas otras.

Es una empresa imposible lograr que en un repertorio léxico estén consignadas todas las innovaciones que se van produciendo en la lengua, pues el léxico está en continuo movimiento mientras que el

²⁹ Manuel Alvar Ezquerro, «Un diccionario moderno del español: DGILE», en mi *Lexicografía descriptiva*, Bibliograf, Barcelona, 1993, págs. 309-312, del que apareció una versión abreviada casi con el mismo título en *Leer*, 26, diciembre de 1989, págs. 58-59. Resumo lo dicho allí.

³⁰ Véase lo que dice a este propósito Ramón Cerdá, «Vox. Diccionario General Ilustrado de la Lengua Española», pág. 214.

diccionario permanece sin cambios hasta la siguiente edición. En esa entrega del DGILE se consignaron muchas voces que estaban en la lengua desde hacía algún tiempo, aunque sin registrar en nuestros diccionarios, como *biodegradable*, *computerizar*, *contactología*, *contracultura* o *uperizar*. Entre esos términos unos eran -lo siguen siendo- barbarismos no admitidos en los repertorios más normativos, aunque sean de uso generalizado: *break*, *catering*, *jazz* (el DRAE escribe *yaz*), *jet-set*, *ketchup*, *lichi*, *rack*, *strip-tease*, etc.

Otro grupo de entradas y acepciones nuevas estaba constituido por vulgarismos flagrantes (por ejemplo, *cojonudo*, *gilipollas*, *taxi 'prostituta'*, etc.), términos del lenguaje de los jóvenes (como *carroza*, *dabuti*, *flipar*, *tronco*, etc.), de la droga (*camello*, *chocolate*, *nieve*, *porro*, *viaje*, etc.), etc., en una actitud más permisiva de lo habitual en los diccionarios generales de la lengua, por el deseo de ser descriptivo y no solamente prescriptivo.

Y del mismo modo, la última salida del DGILE registraba voces propias de la ciencia y de las técnicas tal vez con mayor generosidad que otros repertorios de características similares, pues un diccionario que se precie de moderno ha de acoger la terminología especializada para atender las exigencias de un gran número de usuarios, tanto especialistas de las distintas ramas del saber, como de estudiantes y personas cultas en general. El incremento de tecnicismos³¹ pretendía satisfacer las necesidades del hablante que pudiera encontrarse con ellos en alguna ocasión, y del que, interesado por acrecentar sus conocimientos, tuviera un trato más o menos asiduo con ellos. Por ello aparecían términos como *esferometría*, *fotocinesis*, *fotocoagulación*, *hidrogeología* o *hemocromatosis*.

Para esa salida de 1987 se reexaminaron los regionalismos de una manera tan pormenorizada que voces o acepciones que figuraban en la edición anterior como propios de algún país americano también eran localizados en España (por ejemplo, *boconada*, *busfar* o *conmuta*). Una atención especial mereció el vocabulario dialectal del centro peninsular, pues los diccionarios suelen atender las hablas laterales del español olvidándose del castellano. Esas hablas quedaron representadas por voces y acepciones como *garullo*, *garrucha*, *germanía*, *lombricero*, *llanquete* o *tarriza*. Los dialectalismos, regionalismos y localismos no llegaron de

³¹ Para la presencia de la marca de los tecnicismos en este diccionario en comparación con el DRAE, debe verse Dolores Azorín Fernández, «El Diccionario General de la Lengua frente a los vocabularios científicos y técnicos», en *Euralex '90 Proceedings. Actas del IV Congreso Internacional. Benalmádena (Málaga), 28 agosto-1 septiembre 1990*, Bibliograf, Barcelona, 1992, págs. 445-453.

una manera caprichosa: su difusión fue un factor considerado para traerlas, pero también el interés que representan para la historia de la lengua, la ayuda que prestan para entender ciertos fenómenos, para explicar la forma o el significado de voces emparentadas o próximas por cualquier motivo.

En la última edición del DGILE también se corrigieron muchas definiciones que era necesario modificar o actualizar. Y con el fin de completar las de los reinos animal y vegetal, se hizo un gran esfuerzo para proporcionar los nombres científicos de todos los animales y plantas, añadiéndose los que faltaban -constaban muchos en ediciones anteriores, lo que no se ha hecho ver con frecuencia a pesar de tratarse de una importante innovación en nuestra lexicografía-, siempre que se lograba averiguar qué era lo definido, tarea nada fácil con las designaciones vulgares acogidas en las ediciones precedentes del diccionario (por ejemplo, *blanquita, gallarito, genipí, gitano* o *tartajal*), y que en más de una ocasión sólo son las de la especie, pues en el uso diario que hace el hablante de su lengua no necesita precisar con el mismo rigor que lo hace el científico. Del mismo modo, buscando la precisión y comodidad del usuario, siempre que figura en la definición un nombre propio de persona se pusieron junto a él las fechas de su nacimiento y muerte, o la época en que vivió con el fin de acabar con muchas pistas perdidas.

Se transformó otro grupo de definiciones, las del tipo 'acción y efecto de', corrientes en otros diccionarios, y que aparecen en el DGILE desdobladas siguiendo los consejos que expuso hace años Julián Ribera³². En el intento de romper con las fórmulas anquilosadas en la lexicografía, se procuró que todas las definiciones fueran no sólo equivalentes semánticos de lo definido, sino también sintácticos, y así ya no aparecen definiciones introducidas por 'dícese de', 'aplicase a', frecuentes en otros diccionarios, lo cual ha llevado a desarrollar la indicación del contorno de la definición, y que en las ediciones anteriores afectaba fundamentalmente a los complementos verbales³³. Dice Ignacio Ahumada que «el DGILE, que fue precursor en España de esta indicación tan necesaria desde el punto de vista lingüístico, debe extender esta indicación formal a todas sus funciones. Tanto en la edición de 1973

³² Julián Ribera, «Los nombres de acción en el Diccionario actual de la Academia», *Boletín de la Real Academia Española*, V, 1918, págs. 281-297.

³³ El contorno en el DGILE fue estudiado por Manuel Seco, «El "contorno" en la definición lexicográfica», citado. También se dio cuenta de la importancia del contorno Edwin B. Williams, «The Translator and the Bilingual Lexicographer», *Babel. Revue internationale de la traduction*, 21, 3, 1975, págs. 126-127. Para el contorno en el DGILE, véase lo que expone Ignacio Ahumada, «El nuevo diccionario VOX (dgile)», citado, págs. 199-202.

como en la nueva edición de 1987 así lo viene haciendo; ahora bien, debe quedar expreso en el prólogo y debe sistematizarse a todas las definiciones que lo requieran, y que no son en exclusiva las de verbos transitivos, sino que afecta también a intransitivos y pronominales, así como a adjetivos, nombres y adverbios.»³⁴

El DGILE en su salida actual tiene un gran número de artículos con contenido gramatical, en la intención de cubrir la mayor parte posible de requisitos del usuario de la obra, siempre con el ánimo de servirle en las necesidades cotidianas del conocimiento de la lengua. De esta manera se indican usos incorrectos en artículos como *afirmativo*, *brevedad*, *inclusive*, *otro*, *-tra*, etc.

Después de haberse publicado esa versión del DGILE, tuve conocimiento de las bases de revisión de la obra que dejó escritas don Samuel³⁵, con las que coincide plenamente el trabajo desarrollado. Valga como muestra el párrafo en el que escribe:

Cal mantenir també l'estil del diccionari, és a dir, la mateixa concisió definidora. Lo'ampliació del *Vox gran* es refereix principalment a afegir articles i accepcions noves, no pas a allargar les definicions quan no sigui necessari. Es clar, però, que els articles existents han d'esser allargats o escurçats, quan sentim la definició actual com insuficient, poc clara o amb detalls balders. Convé admetre sense por una quantitat considerable de neologismes, i adhoc de barbarismes d'ús generalitzat. Haurem d'admetre també els tecnicismes que s'hagin incorporat a la llengua viva general, deixant fora els mots tècnics que restin confinats als especialistes. Els americanismes han de tenir un tractament extens i acurat. Un diccionari particular no té la responsabilitat normativa de l'Acadèmia, i per consegüent pot i deu obrir lliberalment les portes a totes les innovacions.

Constituyó una novedad en el proceso de redacción la asistencia de ordenadores electrónicos, como se hace en cualquier empresa lexicográfica de envergadura, por más que entonces fuese insólito en España³⁶. Con un diccionario estructurado en un soporte informático

³⁴ Ignacio Ahumada, «El nuevo diccionario VOX (dgile)», citado, pág. 200.

³⁵ Citadas más arriba.

³⁶ Del uso de ordenadores electrónicos en lexicografía ya había hablado en mi *Proyecto de lexicografía española*, Planeta, Barcelona, 1976. He de advertir que la revisión del DGILE nada tuvo que ver con lo expuesto en ese libro. Para esta otra tarea véase Manuel Alvar Ezquerro, «Una base de datos lexicográficos del español», en *Computational lexicology and lexicography. Special issue dedicated to Bernard Quemada*, I, Giardini editori e stampatori, Pisa, 1990, págs. 33-41, y Manuel Alvar Ezquerro, «Desarrollos actuales en lexicografía automatizada del español», en José Vidal Beneyto (dir.), *Las industrias de la lengua*, Fundación Germán Sánchez Ruipérez, Madrid, 1991, págs. 219-240.

parecía fácil la tarea de sacar provecho de lo que ya se tenía, como estaba previsto, pero no fue así debido a una serie de hechos que no vienen al caso ahora. El DGILE se halla bajo una nueva forma electrónica después de haber pasado por otras etapas de codificación³⁷, hasta llegar a la forma definitiva de base de datos con múltiples campos relacionados entre sí, y no de manera jerárquica. El cambio de estado no ha sido traumático al disponer de una organización sistemática y coherente de los datos y de una gramática interna de la obra³⁸, por lo que ha podido ir cambiando de formatos sin grandes inconvenientes. De esta manera es posible tener total y permanentemente actualizada la obra, tanto en la corrección de errores y erratas, como en la incorporación de palabras, acepciones, giros, y cuantos elementos puedan pertenecer al contenido de un diccionario, así como la extracción de las informaciones contenidas en el diccionario con los más diversos fines³⁹. La actualización de la obra se consigue gracias a los datos que se recogen en el continuo análisis de fuentes de información sobre la lengua, y de los medios que la emplean, especialmente los de comunicación. De este modo, podemos decir que el diccionario está al día, aunque la forma impresa que se puede encontrar en las librerías sea más anticuada.

Se ha conseguido tener un gran diccionario de nuestra lengua, como lo demuestran sus cerca de cien mil artículos, y la reconocida calidad que poseía desde su primera salida, pero eso no es suficiente después del trabajo realizado y de los medios empleados. Ahora el texto del DGILE, una vez corregido y actualizado constituye una fuente de datos lexicográficos que sirve como base para la modernización de la colección de diccionarios Vox. De este modo, en 1990 la editorial publicó una nueva versión del diccionario, sin ilustraciones y con el formato que poseía antes de la nueva ampliación, bajo el título de *Diccio-*

³⁷ Para los antecedentes y primeros pasos de esa versión automatizada, véase Manuel Alvar Ezquerro, «Una base de datos lexicográficos del español», citado, así como «Desarrollos actuales en lexicografía automatizada del español», también citado.

³⁸ Véase Manuel Alvar Ezquerro, María José Blanco Rodríguez, Juan Manuel López Guzmán y M. Fernando Pérez Lagos, «La gramática interna de los diccionarios Vox», en *Euralex '90 Proceedings. Actas del IV Congreso Internacional. IV International Congress. Benalmádena (Málaga), 28 agosto-1 septiembre 1990*, Bibliograf, Barcelona, 1992, págs. 427-435. De la gramática interna del diccionario se han ocupado, por otras razones, M. Antonia Martí e Irene Castellón, «Gramática para el análisis del diccionario VOX», *Boletín de la Sociedad Española para el Procesamiento del Lenguaje Natural*, 9, marzo de 1991, págs. 124-143.

³⁹ Véase, por ejemplo, Alicia Ageno, Sivar Cardoce, Irene Castellón, Ma. Antonia Martí, Germán Rigau, Horacio Rodríguez, Mariona Taulé y Felisa Verdejo, «Análisis de definiciones del diccionario VOX», *Boletín de la Sociedad Española para el Procesamiento del Lenguaje Natural*, 11, diciembre de 1991, págs. 139-154.

nario Actual de la Lengua Española (DALE)⁴⁰, cuyo contenido, corregido y aumentado, supera en número de entradas y acepciones al DGILE, alcanzando las 93465⁴¹ entradas y las 169293 definiciones, lo que supone una proporción de 1.8 acepciones por entrada, cifra muy elevada para un diccionario de la extensión de éste, lo que, por otra parte, nos habla de la riqueza de su contenido.

Dolores Azorín Fernández⁴² señala como elementos que diferencian al DALE del DGILE su carácter onomasiológico estricto al prescindir de las ilustraciones, la acentuación de su orientación descriptiva y su formato más reducido y manejable en el aspecto externo; en lo que se refiere al aspecto interno destaca la utilización del orden alfabético internacional (la *ch* y la *ll* son tratadas como dos letras), tal y como proponía Menéndez Pidal en el trabajo que servía de prólogo al DGILE, orden recientemente adoptado por la Real Academia Española. Tras la muestra analítica que realiza esta investigadora concluye que la ampliación del DALE (más de 4000 artículos más que su predecesor; 4300 según Medina Guerra⁴³) se efectúa, sobre todo, en el ámbito de las voces de uso científico y técnico⁴⁴, así como en el léxico temático de ámbito restringido si bien sin marca de especialidad, aunque también aparece un elevado número de elementos formadores de palabras⁴⁵. De los nuevos términos que figuran en el DALE, según Medina Guerra⁴⁶ muchos son de nueva creación (*animalario, circadiano, degradable, delfinario, ecu, farmacológicamente, hipocalórico, inmunodeficiencia, intoxicador, lycra, macroeconómico, parapente, peatonalizar, preeuropeo, preolímpico, redimensionamiento, robótica, sobreventa, uetista, ultraligero, videoedición, vietnamizar, vocacionalidad*) o barbarismos de reciente incorporación al idioma (*clínic, coequipier, eagle, hall, hobby, living, parking, pub, punk, rappel, rol, rough, squash, stick*), aunque todavía faltan algunos como *look* o *pressing*.

Otro grupo importante de las nuevas incorporaciones lo constituyen los americanismos (*jachuchuy!, afió, alguita, cajú, cardosanto, carocito, coupí, guajaba, guayabita, huizache, majagüillo, maruguito, medio-*

⁴⁰ Sobre la obra, véase Manuel Seco, «Diccionarios: la realidad y el deseo», citado; y las reseñas de Hugo Sonnevile, *Romaneske*, XV-1, 1990, pág. 70; Dolores Azorín Fernández, *Estudios de Lingüística. Universidad de Alicante*, 7, 1991, págs. 277-281; y Antonia M^a Medina Guerra, *Revista de Filología Española*, LXXIV, 1994, págs. 196-198.

⁴¹ Según expone Antonia M^a Medina Guerra, en la reseña recién citada, pág. 196.

⁴² Cfr. Dolores Azorín Fernández, en la reseña citada, págs. 278-279.

⁴³ Véase su reseña, pág. 196.

⁴⁴ Me remito a lo dicho más arriba sobre la inclusión de afijos de carácter culto y las voces formadas con ellos.

⁴⁵ Más arriba expuse las cifras al hablar del DGILE.

⁴⁶ Medina Guerra en su reseña, pág. 196.

pelo, michay, nené, pantaloneta, pepinito, purgo, querequere, quinilla, recursivo, -va, sacarracacha, zapupe, zubinché), junto a las voces peninsulares de difusión geográfica restringida (*alpendada, alvejuela, bogalla, mostolilla, mule, villería*), y términos del habla familiar (*bibe, depre, facul, mani, minina, naturaca, poli, presi, profe, propi, ridi*).

Hay otros cambios no menos significativos tanto en la macroestructura de la obra como en el contenido del artículo, pues en unos se eliminan acepciones (como en *báltico, benedictino, -na, calcáreo*), en otros se añaden (*acapulco, ácido, almendrón, bandeja, cachet, paquete*), y en otros se modifican las definiciones.

«La aparición del *Vox actual* -en palabras de Manuel Seco⁴⁷ - escribe, por el momento, el último capítulo en la biografía de un diccionario que ha sabido superarse constantemente a lo largo de los años.»

El DALE sufrió un ligero aumento y correcciones en su contenido, aunque no han afectado a la totalidad de la obra, para presentarse como CD-ROM bajo la forma de libro electrónico⁴⁸, presentación que no ha trascendido mucho al gran público.

En el mes de septiembre de 1995 se ha finalizado una nueva corrección del contenido del diccionario, con la correspondiente incorporación de voces y acepciones nuevas, y los cambios precisos en el interior del artículo para darlo a la luz en formato CD-ROM estándar de manera inmediata, manteniéndose en la vanguardia de forma y contenido que siempre ha caracterizado a los diccionarios Vox.

Hasta aquí he hablado de una sola obra en sus diversas ediciones y presentaciones. Es el repertorio léxico Vox, el banco de informaciones de que dispone esta colección de diccionarios, y del que, necesariamente, estando actualizado y organizado en una forma adecuada, se derivan otros diccionarios.

Al mismo tiempo que se trabajaba en el DALE, se efectuó una reducción para crear un nuevo *Diccionario Manual Ilustrado de la Lengua Española*⁴⁹ para cuya realización no se suprimieron artículos o acepciones sin más, ni se eliminaron informaciones porque sí, ni se cortaron las existentes, sino que se hizo en función del público al que va destinado, poniéndose el máximo cuidado para que en el proceso de reducción no quedasen definiciones con palabras que a su vez no estuviesen definidas dentro del diccionario, lo cual obligó a un proceso de

⁴⁷ Manuel Seco, «Diccionarios: la realidad y el deseo», citado, pág. 7.

⁴⁸ Sony-Biblograf, Barcelona, 1992.

⁴⁹ 14ª ed., Biblograf, Barcelona, 1990.

análisis y remodelación que alteró en muchos lugares el contenido del diccionario que servía de punto de partida. De este modo se logró un diccionario con unas 37600 entradas y 74884 acepciones (lo que supone dos acepciones por entrada, de promedio).

En el proceso de modernización de los diccionarios Vox, gracias también a la base de datos lexicográficos, al DGILE, al DALE y al *Diccionario Manual*, ha podido crearse un *Diccionario Práctico de la Lengua Española*⁵⁰ destinado a un público escolar, con un total de 23806 entradas. Debido a sus características, es esta obra la que concede un mayor espacio, a través de las ilustraciones, a la realidad extralingüística. Por otra parte, presenta la novedad de ofrecer los modelos de la conjugación verbal en los márgenes laterales de la página, y la de añadir en otras una última línea de información normativa con el fin de corregir los errores más comunes, o para hacer reflexionar al usuario sobre algunos aspectos (ortográficos, morfológicos, semánticos) en los que no siempre repara, y que pueden contribuir a mejorar su dominio de la lengua.

En 1992, la editorial Biblograf tomó la decisión de realizar un diccionario de enseñanza de la lengua, del cual carecíamos, pese a existir ya varios en otras lenguas. Para realizarlo se buscó la colaboración de un grupo de expertos de la Universidad de Alcalá de Henares. La idea inicial era partir del *Diccionario Manual* para adaptarlo a las nuevas necesidades; sin embargo, por otras determinaciones que se tomaron (por ejemplo, limitar el número de definidores, evitar las definiciones sinónimas, etc.), añadido al hecho de contar con un corpus sobre el que fundamentar el contenido y de basarse en recuentos estadísticos del léxico, el diccionario que debía ser la base de la nueva obra se quedó tan solo como guía. Hablo del *Diccionario para la enseñanza de la lengua española* (DiPELE)⁵¹.

La macroestructura de este repertorio (unas 22 000 entradas) se ha seleccionado a partir de los recuentos léxicos de que disponemos, de las frecuencias que ofrece el corpus Vox-Biblograf (se utilizó cuando contaba con cinco millones de palabras), de la distribución sociolingüística de los elementos, y de las necesidades culturales de los usuarios. Su contenido ha sido seleccionado y estructurado en función de lo que precise el que acuda a él.

La característica más sobresaliente del DiPELE es la inclusión de ejemplos, rompiendo con la pertinaz ausencia de ellos que hay en nues-

⁵⁰ También publicado por Biblograf, Barcelona, 1990.

⁵¹ Biblograf, Barcelona, 1995.

tra producción lexicográfica, además de ser ilustrado. Ahora bien, los ejemplos no figuran en todos los artículos, ni en todas las acepciones, pues se ha pretendido evitar que sean una repetición de lo que se dice en la definición; por ello, no figuran los que no aportan nada nuevo, ya que su inclusión en un sintagma simple como muestra de uso no permite deducir al usuario cuáles son las posibles restricciones en el empleo de la voz. De esta manera se evitan los ejemplos inútiles y repetitivos.

Las acepciones aparecen ordenadas con arreglo a su frecuencia de empleo en la lengua, por más que eso pueda desestructurar en ocasiones el contenido, o pueda sorprender al usuario habitual de diccionarios. Se ha puesto un especial cuidado en la selección de la macroestructura y de la microestructura, teniendo bien presente el público al que van destinados.

El DiPELE no sólo está concebido para atender las necesidades de un usuario escolar español, sino también para el estudiante que desea aprender el español como segunda lengua, lo que ha llevado a incluir la indicación de la pronunciación en todas las entradas. Además, junto al lema se señala la división silábica de las palabras, para despejar algunas de las consultas que pueden hacer quienes se inician en el conocimiento de la lengua.

La microestructura de este diccionario posee, por vez primera en nuestra lexicografía, una serie cerrada de definidores (dos mil; se enumeran en un apéndice), de manera que ningún usuario se vea en dificultades para entender lo que se dice en las definiciones (las dificultades las han tenido los redactores para establecer la lista y para hacer las definiciones).

El ejemplo en este diccionario es algo necesario, no sólo como muestra del uso de las palabras, sino también como complemento de la definición. El lector encontrará en él multitud de informaciones adicionales, de la definición y sobre lo nombrado. Así por ejemplo, la definición de *natillas*, 'Dulce hecho con yemas de huevo, azúcar y leche', resulta imprecisa debido al tipo de obra que se ha redactado, pero con la ejemplificación se ayuda a la comprensión: «para hacer ~ tienes que mezclar los ingredientes mientras se cuecen lentamente; el niño ha tomado ~ en el postre».

Dado su carácter didáctico, presenta abundantes notas y observaciones de carácter normativo, sinónimos, antónimos, voces relacionadas, etc. «Ofrece toda la información ortográfica, fonética, gramatical, semántica y pragmática necesaria para garantizar un uso de la lengua correcto y adecuado a un contexto determinado». En ese sentido,

proporciona como apéndices, según es característica de los diccionarios Vox, los modelos de conjugación irregular.

Como los demás diccionarios de la colección Vox, el DiPELE es puramente descriptivo de lo que se halla en la lengua, por lo que en él se encuentra una gran cantidad de voces del habla juvenil y familiar, evitando el divorcio que pueda existir entre la lengua culta que se enseña y el habla cotidiana con la que se tropiezan los estudiantes de español.

Paralelo a la serie de diccionarios Vox, aunque no ajeno a ella, hay que mencionar el *Diccionario escolar de la lengua española*⁵². Se trata de un diccionario especial dentro de la serie de los Vox, pues en el contenido de algunos de sus artículos hay largas explicaciones enciclopédicas cuyo fin es el de acrecentar los conocimientos del usuario, con lo que se convierte, sin serlo, en un diccionario enciclopédico, en una obra de consulta permanente. El enciclopedismo de su contenido se pone de manifiesto, por otra parte, en el no muy largo número de entradas, si bien son abundantes las voces consignadas ya que al final de una gran cantidad de artículos figura un apartado con los derivados del término de la entrada. Para conceder un mayor espacio al contenido enciclopédico dentro de los límites de un diccionario destinado al público escolar, se prescindió de las voces de uso más frecuente, o bien conocidas, como *carretera, finca, ventana, jardín, jamón*, etc. El contenido del diccionario siempre ha sido el mismo hasta que en 1986 se le añadieron unos centenares de entradas más, entre las que no se encontraban las voces a que acabo de aludir.

Como consecuencia del proceso de renovación de todos los diccionarios Vox emprendido por la editorial hace unos años, así como consecuencia del acopio de conocimientos efectuado, esta obra ha sido sometida a una nueva redacción, y en 1992 apareció la 14ª edición -también realizada bajo mi dirección-, en la que se incorporan las voces más frecuentes o conocidas que estaban ausentes de las anteriores salidas, se incluyen neologismos, se corrigen las definiciones de la terminología científica y técnica y se le da una nueva apariencia formal renovando la tipografía y la ilustración. En definitiva, se ha hecho un diccionario nuevo manteniendo la estructura anterior de los artículos.

En la primavera de 1995 ha aparecido un nuevo diccionario ideológico de nuestra lengua que viene a romper la atonía existente en

⁵² Según datos que me facilitan en la editorial, es una obra que se encargó a Raimon Galí antes del año 1955.

este tipo de obras. La editorial disponía de un *Diccionario temático de la lengua española*⁵³ que no gozaba de los favores del público. En diversas ocasiones se pensó en actualizar su forma y su contenido, hasta que, finalmente, se optó por acometer una obra nueva, el diccionario ideológico, que ha venido a enterrar definitivamente al temático.

Con el nuevo *Diccionario Ideológico* se ha pretendido atender a las necesidades de las personas que acuden a los diccionarios ideológicos, que se pueden resumir en los siguientes puntos:

- a) empleo como diccionario de sinónimos.
- b) empleo como diccionario de términos relacionados, que no siendo sinónimos, y por tanto estando fuera de la mayor parte de los diccionarios que los recogen, presentan un cierto número de coincidencias semánticas; sólo en un diccionario de estas características podrán encontrarse juntas palabras como *tubo* y *manga* o *lectura* y *fichero* (bajo el epígrafe de *mensaje*).
- c) búsqueda de una palabra cuya forma se desconoce, y cuyo significado se conoce, aunque sólo sea por aproximación: si se ignora el nombre con que se designa al madero grueso propio para ser serrado bastará con consultar el apartado *medios* y el subapartado *madero*, donde encontraré *cachizo*.
- d) búsqueda del vocabulario agrupado bajo un concepto o una idea de carácter más o menos amplio. De este modo, en el apartado *anotación* encontramos distintos subapartados encabezados por *nota*, *cita*, *filiación*, *asiento*, etc., donde se enumeran las palabras relacionadas con esas ideas.

Además, no se han querido dejar al margen otros empleos de los diccionarios, como es el habitual para consultar el significado de las palabras.

Por las razones recién expuestas, el diccionario se puede emplear como diccionario de sinónimos y de voces relacionadas, pues como en el de Casares aparecen agrupadas las palabras en series, constituyendo bloques independientes según su proximidad significativa, y la categoría gramatical que desempeñen. Así, por ejemplo, son voces relacionadas

⁵³ Barcelona, 1975, con prólogo de Juan Alcina Franch; para su contenido, véase lo que escribí en «Diccionarios ideológicos», *Libros*, nº 24, febrero 1984, págs. 14-18, luego recogido con apuro crítico bajo el título de «Los diccionarios ideológicos del español» en mi *Lexicografía descriptiva*, págs. 289-301; del *Diccionario temático* me ocupo en las págs. 299-300.

entre sí, y, en cierta manera, tienen una parte de su significación muy próxima, las denominaciones de las distintas embarcaciones (encabezadas por *barco* encontramos *embarcación, nave, bastimento, nao, navío, vapor*, etc., junto a otro subapartado para *carguero, motonave, petrolero, urca, chalana, rompehielos*, etc., así como para *acorazado, portaaeronaves, portaaeriones, portabelicópteros, destructor, fragata, corbeta, torpedero, dragaminas*, etc.), las que designan la *imprudencia* (a continuación de *imprudencia*, encontramos *intrepidez, irreflexión, inconsideración, temeridad, alegría, aventurismo*, etc., le sigue un subapartado para *locura, desafuero, disparate, cabildada, atrocidad, barbaridad*, etc. y otro de *precipitarse, atropellarse, desmadrarse, embalsarse, excederse, desafinar*, etc.), o todos los del campo designativo que gira en torno a la idea del *descanso* (tras *descanso* encontramos *respiro, receso, respiradero, oasis, tregua*, junto a otro subapartado para *sueño, entresueño, adormecimiento, duermevela, siesta, canóniga*, y otro para *somnolencia, cabezada, soñera, sopor, soñarrera, pesadez*, etc.).

Para encontrar una palabra que se desconoce pueden emplearse dos vías:

1.- Partiendo del árbol clasificatorio general, y guiados por el índice numérico de cabeceras que contiene cada cuadro, se llega a la palabra o grupo de palabras que se busca. Así, por ejemplo, para hallar *amor propio*, buscaré en el cuadro general y encontraré como cuadro más apropiado el de *sentimiento*; una vez consultado el listín de cabeceras de este cuadro, comprobaré que lo más apropiado es buscarla en el apartado *orgullo*; llegar a continuación a la página donde están listadas las denominaciones de la cabecera resulta fácil al estar guiados por su marcador numérico.

2.- Si busco una palabra que no conozco, o no me acuerdo, puedo acudir al índice alfabético para buscar alguna próxima que se me venga a la memoria, desde allí se me remitirá al cuadro preciso. Así, por ejemplo, si deseo encontrar algo parecido a *telenovela*, el índice me remite al cuadro *cinematografía* donde se hallan *culebrón, serial, serie*, etc.

Para saber el significado de las palabras registradas, se dispone de la última parte del diccionario que incluye por orden alfabético, y con sus significados, todas las voces consignadas en las partes anteriores, y que han servido de fuente para el trabajo, el *Diccionario Manual Vox* del que acabo de hablar.

Este diccionario ideológico no es un diccionario de sinónimos, sino un diccionario de ideas afines que se aleja de la idea que tenemos del repertorio de Casares y se aproxima a los que existen para otras lenguas de nuestro entorno; se trata de una obra en la que las palabras aparecen juntas por las ideas que nosotros tenemos de las cosas.

Esto puede causar un gran inconveniente, pues estamos acostumbrados a buscar las palabras en los diccionarios de manera formal y no conceptual. Por otra parte, no todos los hablantes tienen las mismas ideas de las cosas, lo que, con seguridad, producirá fracasos en las búsquedas. Para evitar que el usuario no encuentre lo que desea, se ha incorporado, además del cuadro clasificatorio general, un índice con todas las palabras recogidas en la obra, de modo que se pueda acceder a aquello que se busca con sólo conocer las palabras próximas o que designan realidades cercanas por cualquier motivo.

La clasificación por la idea que tenemos de las cosas no es de ningún modo un ordenamiento de la realidad extralingüística. De ahí que, por ejemplo, la *tinta* aparezca incluida en el campo *tinta*, la *pluma* con el *lápiz*, y el *papel* encabece su propio apartado. El medio para establecer estas relaciones ha sido la agrupación de términos a partir del análisis de los componentes de su significado y de la imagen que poseemos de la realidad designada. Por ejemplo, los *turdetanos*, los *celtas* y los *iberos* son pueblos diferentes, pero quedan agrupados porque pertenecen a nuestra antigüedad; en cambio es la funcionalidad lo que hace que se agrupen la *tenaza* con la *pinza* o el *fórceps*, o el *martillo* con el *batán*.

Como se puede apreciar por los ejemplos aducidos, no se trata de una enciclopedia que clasifique las cosas, ni un repertorio de sinónimos, sino un diccionario en el que las palabras se agrupan porque sus significados se agolpan en nuestra mente asociados por las ideas que tenemos de las cosas.

Han sido estos cincuenta años de diccionarios monolingües Vox muy fructíferos, gracias al empeño de muchas personas que han puesto todas sus ilusiones en estas obras, y al frente de todos nosotros al maestro don Samuel Gili Gaya, sin cuyas enseñanzas no habríamos llegado hasta aquí. El de los diccionarios Vox es un capítulo ciertamente importante, quizás el más importante, de la lexicografía española no académica. Han sabido atender las necesidades de los que desean una descripción de la lengua sin las limitaciones estrechas de la norma que debe observar la Academia, por más que intente liberarse ahora de ellas; y también ha querido atender a los públicos más variados, teniendo siempre presente el carácter didáctico y escolar de los diccionarios. En

los últimos años se han abierto nuevos horizontes con la renovación tecnológica, y ahí están en la vanguardia de las técnicas de redacción de diccionarios, pero también en lo más avanzado de la producción en formato electrónico. Pero ello no es todo: son numerosísimos los diccionarios bilingües y enciclopedias que llevan la marca Vox, de los que no me ocupo ahora por falta de tiempo.

Antes de acabar quiero recordar con toda sencillez el más viejo de los diccionarios de la casa, una joya que sigue brillando hoy, el *Diccionario ilustrado latino-español español-latino*, que se publica desde 1943, sin el cual, tal vez, no habría venido nada de lo que hoy tenemos.

Profesor Emile Litvinovitch de la Sorbona

Vous pouvez légitimement être surpris que l'on ait entrepris un dictionnaire de mots de haute fréquence, alors que l'on a part, les mots de haute fréquence sont généralement ceux qu'on aime très bien connaître et qu'on se cherche pas dans le dictionnaire, et l'autre part que les dictionnaires recherchent d'ordinaire l'élégance.

Mon objectif est double, théorique et pratique.

Sur le plus théorique, il s'agit pour moi de revenir à l'origine, sur un nombre de mots assez considérable pour être significatif, la théorie de la polysémie, d'inspiration gollaudienne, que j'ai exposée dans mes *Directeurs sémantiques de langues françaises* (Paris, Nathan 1984) et de voir quelle peut être sa fécondité.

Sur le plus pratique, il s'agit de donner aux enseignants de français langue maternelle ou seconde un instrument de travail qui les aide à traverser le choc d'un enseignement symbolique, et une occasion de langage. Le dictionnaire a été conçu et organisé en une filière de la langue française qui permet d'observer les problèmes de l'Afrique. Je pense tout particulièrement à l'Université africaine d'Abidjan, avec lequel nous le publions en France, nous être un lien qui confirme, d'ailleurs, le français à des côtés interchangeables de pour être avec moi pour voir de manière continue et plus un nous affaiblit de voir. Il s'agit de l'appeler un instituteur abidjanaise "Moussé" et je réclame de lui être une marchandise dans le monde français présente le mot d'usage possible.

Je voudrais que mon dictionnaire lui soit utile dans son travail qui est à améliorer le vocabulaire et les performances d'Abidjan pour qui le français est une langue seconde mais non maternelle. Il veut donc le plus grande clarté et la plus grande simplicité possible.

Au cours des années, deux collaborateurs se sont joints à moi: Marie-Luce Homère, maître de conférences à Saint Etienne et Jean-

... y en el momento de la escritura, el autor se encuentra en un estado de tensión y de concentración que le permite captar y registrar los detalles más insignificantes de la vida cotidiana. Este estado de ánimo es el que le da a su obra una fuerza y una profundidad que no se encuentran en la obra de un escritor que simplemente quiere contar una historia.

Este estado de ánimo es el que le da a su obra una fuerza y una profundidad que no se encuentran en la obra de un escritor que simplemente quiere contar una historia. Este estado de ánimo es el que le da a su obra una fuerza y una profundidad que no se encuentran en la obra de un escritor que simplemente quiere contar una historia.

La obra de un escritor que simplemente quiere contar una historia es una obra que carece de fuerza y de profundidad. Este estado de ánimo es el que le da a su obra una fuerza y una profundidad que no se encuentran en la obra de un escritor que simplemente quiere contar una historia.

Este estado de ánimo es el que le da a su obra una fuerza y una profundidad que no se encuentran en la obra de un escritor que simplemente quiere contar una historia.

Este estado de ánimo es el que le da a su obra una fuerza y una profundidad que no se encuentran en la obra de un escritor que simplemente quiere contar una historia.

LEXICOGRAPHIE THÉORIQUE ET APPLIQUÉE: UN DICTIONNAIRE DES MOTS DE HAUTE FRÉQUENCE

JACQUELINE PICOCHÉ

Professeur émérite à l'Université de Picardie

Vous pouvez légitimement être surpris que j'aie entrepris un dictionnaire des mots de haute fréquence, alors que d'une part, les mots de haute fréquence sont généralement ceux qu'on croit très bien connaître et qu'on ne cherche pas dans le dictionnaire, et d'autre part que les dictionnaires recherchent d'ordinaire l'exhaustivité.

Mon objectif est double, théorique et pratique.

Sur le plan théorique, il s'agit pour moi de mettre à l'épreuve, sur un nombre de mots assez considérable pour être significatif, la théorie de la polysémie, d'inspiration guillaumienne, que j'ai exposée dans mes *Structures sémantiques du lexique français* (Paris, Nathan 1986) et de voir quelle peut être sa fécondité.

Sur le plan pratique, il s'agit de donner aux enseignants de français langue maternelle ou étrangère un instrument de travail qui les aide à trouver la voie d'un enseignement systématique, et non accidentel du lexique. Ce dictionnaire a été conçu au moment où mon Histoire de la langue française me faisait découvrir les problèmes de l'Afrique. Je pense tout particulièrement à l'instituteur africain chargé, sans jamais avoir mis le pied en France, sans être un liguiste confirmé, d'enseigner le français à des classes surchargées de petits élèves assez mal pourvus de matériel scolaire et plus ou moins affamés de savoir linguistique. J'appelai cet instituteur stéréotypique "Moussa" et je résolus de lui offrir une marchandise dont le mode d'emploi présente le moins d'obscurité possible.

Je voudrais que mon dictionnaire lui soit utile ainsi qu'à tous ceux qui ont à améliorer la compétence et les performances d'élèves pour qui le français est une langue usuelle mais non maternelle. Il vise donc la plus grande clarté et la plus grande simplicité possible.

Au cours des années, deux collaborateurs se sont joints à moi: Marie-Luce Honeste, maître de conférences à Saint Etienne et Jean-

Claude Rolland chargé d'études au Centre International d'Etudes Pédagogiques de Sèvres.

A la question "par où commencer?", j'ai trouvé une réponse dans la structure pyramidale du lexique, bien mise en lumière par Etienne Brunet dans son *Vocabulaire français de 1789 à nos jours*, exploitation du *Dictionnaire des fréquences* du TLF qui repose sur un corpus littéraire compris entre 1789 et 1965 qui a fourni 70.317.234 occurrences, pour 70.000 vocables. De ses calculs, il ressort que les mots de fréquence supérieure à 7.000, jusqu'au rang 907, représentent plus de 90% de l'ensemble des occurrences de n'importe lequel des textes dépouillés. Viennent ensuite environ 5.800 mots de fréquence supérieure à 500 qui en représenteraient 8% et (si l'on néglige quelque 20.000 hapax) 40.000 vocables d'autant plus riches d'information qu'ils sont plus rares, représentant environ 2% du corpus.

Là se trouvait de toute évidence le bout du fil d'Ariane permettant de se repérer dans le labyrinthe du lexique.

De plus c'est dans cette partie du lexique que se trouvent ce que j'appelle les "grands polysèmes": les mots qui ont une polysémie très complexe et qui entrent dans un grand nombre de locutions figées. Donc, mes deux préoccupations convergeaient vers le traitement de cette partie du lexique.

Ces chiffres doivent être considérés avec une certaine circonspection, pour diverses raisons: il y a parfois plusieurs mots sur un seul rang; les homonymes ne sont pas distingués sauf s'ils sont de catégorie grammaticale différente. Les infinitifs (sous lesquels sont classées les formes conjuguées), les participes présents et les participes passés des verbes sont systématiquement dégroupés, sans qu'il soit possible de distinguer parmi eux les emplois adjectivaux des emplois participiaux. Il est certain que, s'ils étaient regroupés, une plus grande prédominance de verbes apparaîtrait dans le début de la liste.

Quoique, par l'ampleur sans pareille des dépouillements effectués, la liste de fréquences du TLF, soit plus fiable que toutes les précédentes, il nous a paru utile de la comparer:

- 1) à celle de Juilland (corpus littéraire compris entre 1920 et 1950), qui repose sur 464.134 occurrences, pour 5.082 vocables
- 2) à celle du Français Fondamental (corpus oral de 1956), qui repose sur 312.135 occurrences, pour 7.995 vocables
- 3) à celle d'Engwall (corpus littéraire compris entre 1962 et 1968) qui repose sur 500.096 occurrences pour 4.853 vocables

Les listes provenant des corpus littéraires (TLF et Juilland et Engwall) sont communes à 80% alors que la liste du Français Fondamental n'est commune avec le TLF qu'à 65%. Reposant sur un corpus oral, elle s'écarte considérablement des deux autres, notamment à cause d'un nombre non négligeable d'interjections et de vulgarismes dont l'étude ne présente pas un grand intérêt dans notre perspective. Restent tout de même 159 mots du TLF qu'on ne retrouve dans aucune autre liste, courants dans les oeuvres romanesques du XIXe s. et qui se raréfient aujourd'hui. Il est donc légitime de ne pas se rendre esclave d'une seule liste et de pondérer l'une par l'autre. Telles qu'elles sont, ces listes fournissent tout de même une base de départ non arbitraire pour l'exploration du vocabulaire français. Très usuels, les mots de haute fréquence sont évidemment fondamentaux pour l'enseignement d'une langue, et très polysémiques, ils sont un bon matériel d'expérimentation théorique.

J'envisagerai d'abord la structure des articles, ensuite celle des définitions et enfin, j'essayerai de vous donner une idée des applications pédagogiques envisageables.

I. La structure des articles

1) Je préfère un grand article subdivisé à une série d'homonymes.

Je ne nie pas le phénomène de l'homonymie; il y a des homonymes dans mon dictionnaire (ex. 2 articles BORD, 2 articles ETAT, un indépendant, et un autre rattaché au verbe *être* à titre de dérivé). Je ne nie pas qu'il y ait des degrés dans la cohérence des signes linguistiques en général, lexicaux en particulier, et que la question "polysémie ou homonymie" appelle souvent une réponse en termes de "plus" ou de "moins" plutôt qu'une réponse en termes de "oui" ou de "non". Et même dans le cas où la réponse est indubitablement "oui", je sais qu'il peut y avoir intérêt pour certaines raisons pratiques, à adopter un traitement homonymique.

Mais je commence toujours par faire l'hypothèse de la polysémie, qui est un phénomène très général qui relève de la nature même du langage, alors que l'homonymie est une sorte d'accident de la langue. Je suis tout à fait opposée au dégroupement systématique des polysèmes en homonymes imposé naguère par Jean Dubois à un certain nombre de dictionnaires Larousse encore en usage. Il y a chez moi un seul article COEUR, alors qu'on en dénombre trois dans le DFC (éd. de 1967) et quatre dans le Lexis (éd. de 1975).

Personnellement, si j'avais appliqué sa méthode, j'en aurais dénombré sept. Quoique j'admire beaucoup certains aspects du *Dictionnaire explicatif et combinatoire* d'Igor Melc'uk, en particulier ses "fonctions lexicales", je suis tout à fait opposée à son parti-pris de traiter isolément les locutions figées, de telle sorte qu'on trouve à la tête, au pied à la lettre A et les articles TETE et PIED respectivement à la lettre T et à la lettre P. Bien au contraire, dans la mesure où elles ne sont pas complètement démotivées, c'est une navette entre les emplois libres du mot vedette et les locutions figées où aucune substitution n'est possible qui me révèle des traits sémantiques cachés et d'autant plus fondamentaux qu'ils sont généralement inconscients. Ainsi, dans l'étude des parasyonymes BORD et COTÉ, c'est en réfléchissant sur les locutions *au bord des larmes, au bord de la tombe, au bord de la catastrophe* et sur l'impossibilité de dire **au bord du rire, au bord du succès*, que nous avons découvert que le *bord* est un endroit éminemment dangereux par risque de basculement du stable dans l'instable, ce qui n'est pas du tout le cas de *côté*.

2) Si l'on admet que les mots lexicaux sont des objets linguistiques et pas des étiquettes collées au petit bonheur la chance sur des classes de référents, ils doivent avoir, tant dans leur organisation interne que dans leurs relations les uns avec les autres, un caractère plus ou moins systématique qu'il y a intérêt à découvrir. J'appelle *signifié de puissance* le principe en synchronie d'unité et de cohérence des polysèmes qui relève forcément de cet inconscient linguistique créateur dans lequel sont "en puissance" les discours appelés à s'actualiser. Presque tous les mécanismes du langage sont inconscients, et c'est justement le rôle du linguiste de les faire affleurer à la conscience: personne n'a présentes à l'esprit les constructions syntaxiques ni les oppositions phonologiques quand il parle. Elles n'en sont pas moins réelles. Pourquoi exigerait-on des oppositions lexicales qu'elles soient conscientes?

3) J'attache la plus grande importance à l'ordre des acceptions et aux relations d'engendrement qui existent de l'une à l'autre. mais je ne travaille pas en diachronie. Mon objectif est de décrire la langue moderne la plus courante de la façon la plus intelligible possible., en montrant par quelles étapes l'esprit doit de préférence, ou obligatoirement passer pour penser un ensemble d'emplois et de sens comme constituant un unique vocable, ce que j'appelle avec Guillaume "chronologie de raison". Je ne cherche nullement à retrouver un ordre historique. Il se peut que je rencontre la diachronie et qu'il y ait coïncidence entre la chronologie historique et la chronologie de raison, mais ce n'est pas mon problème, du

moins dans ce dictionnaire. Toutefois, j'introduis parfois la mention "vieux" quand il m'arrive d'introduire une locution un peu archaïque utile pour faire comprendre la cohérence de l'ensemble (ex. *courir le cerf* dans l'article *courir* qui permet de comprendre qu'on puisse dire *courir les filles, courir les honneurs, courir un danger*)

4) Les articles sont fondés sur la structure interne des mots plutôt que sur la description de leurs référents. Une telle exigence serait impossible à satisfaire en traitant de mots techniques et monosémiques, ayant plus ou moins le statut de "termes": Ex. *amibe* "protozoaire des eaux douces et salées, qui se déplace à l'aide de pseudopodes, pourvu d'un noyau et se reproduisant par division indirecte"- *ion* "atome ou groupement d'atomes portant une charge électrique"- *ioniser*: "transformer en ions"- (d'où *ionisation, ionisant* etc.)- adj. *hygrophobe* "qui craint l'humidité". Des mots de ce genre sont étroitement liés au sujet traité; on ne peut rien prédire de leur fréquence, et le linguiste sort en quelque sorte de son domaine en les traitant, obligé qu'il est de se référer aux ouvrages des spécialistes.

Par contre, elle est possible et parfois obligatoire pour les polysèmes comportant un bel éventail de locutions et de distributions, ce qui est le cas de la plupart des mots qui arrivent en tête des listes de fréquences.

5) les divisions de l'article sont fondés prioritairement sur des critères sémantiques et non syntaxiques. Naturellement, j'exploite, comme tout le monde, les faits de syntaxe dans l'analyse du sens; c'est un bon outil de recherche, mais ce n'est généralement pas un bon outil d'exposition. Voyez ce que serait devenu mon article INTERET (donné en annexe) si je l'avais construit sur les oppositions syntaxiques! L'expérience prouve que la priorité donnée aux structures syntaxiques éparpille les sens et rend l'article incohérent, notamment en ce qui concerne les verbes. Il me semble évident qu'une opposition comme concret/abstrait est au moins aussi parlante que transitif/intransitif. Les développements concernant la syntaxe se trouvent toujours à une place secondaire et signalés comme tels par un signe spécial.

6) J'introduis dans mes articles, à leur place logique, des dérivés, synonymes et antonymes usuels qui n'entrent pas dans le peloton de tête, d'une part parce que ce sont, eux aussi, des outils très importants de l'analyse du sens, et, d'autre part, dans l'espoir de faciliter un enrichissement progressif et rationnel du vocabulaire des apprenants. Je ne crains pas la notion de "dérivé sémantique" (et pas uniquement morpho-sémantique), non seulement dans le domaine savant (on ne peut pas laisser l'apprenant ignorer que les dérivés de

eau se forment sur les bases *hydr-* et *aqu-*) mais même dans le domaine populaire: je présente *chute* comme dérivé de *tomber*, *coup* comme dérivé de *frapper*

7) Il résulte de tout cela qu'il existe deux grands types de polysémies: celles de type métaphorique, qui comportent un mécanisme de subduction et celles de type métonymique qui n'en comportent pas, sans exclure que certains mots compliqués (ex. CARTE, et, plus fondamentaux, ETRE, AVOIR) combinent les deux mouvements de pensée.

A) Polysémie avec un mécanisme de subduction

Ce terme dénote, dans la terminologie guillaumienne, le mouvement de pensée (ou *cinétisme*) qui va de l'acception plénière la plus riche de sèmes, la plus concrète quand le mot réfère à un être concret, jusqu'à l'acception sémantiquement la plus pauvre, donc la plus abstraite et la plus générale.. L'acception, ou les acceptions les plus pauvres, car les mots lexicaux étant sémantiquement beaucoup plus complexes que les mots grammaticaux, il arrive que plusieurs cinétismes convergent vers l'acception plénière (v. en annexe le mot COEUR). C'est le cas de nombreux substantifs concrets, et même de verbes mettant en jeu, par ex. une gestuelle, et, dans leur cas on a souvent affaire à toute une constellation de locutions figées où se reflète l'imaginaire d'une communauté linguistique. La *métaphore* est un cas particulier de ce type qui existe aussi en l'absence de toute métaphore: On peut donner comme exemple l'ensemble INTERET, INTERESSER, INTERESSANT (v. en annexe) ou le verbe DEVOIR dont tant de linguistes isolent les emplois "modaux": Or, il me paraît facile d'inscrire tous ses emplois sur un cinétisme unique, en commençant obligatoirement par la dette d'argent et en finissant obligatoirement par la probabilité dans le présent ou le passé, en passant par dette de reconnaissance, causalité, nécessité, et futur probable. Je pense que c'est faire violence à la nature des choses et se condamner à ne pas comprendre le fonctionnement de cette "machine sémantique" que de procéder autrement.

Il est des cas très fréquents de "métaphores spatiales", où l'emploi "figuré" n'a rien de stylistique, est la manière normale de parler, notamment quand il s'agit de verbes de mouvements comme *prendre*, *mettre*, *pousser*, *entraîner*, *retenir*: qu'est-ce qui oblige à penser l'espace en premier? Doit-on suivre Bernard Pottier qui, dans sa récente *Sémantique Générale*, met exactement sur le même plan les sens spatial, temporel et notionnel d'un mot polysémique, comme

les différentes branches d'un éventail ou les diverses voies d'un carrefour ? S'il a raison, l'ordre des sens, dans un dictionnaire, est sans importance. Qu'on essaye donc de construire un article *prendre* où on trouvera, dans l'ordre, 1. *Jean prend l'avis de Paul* - 2. *Ce travail prend beaucoup de temps à Paul* - 3. *Jean prend un livre à Paul*, le geste de la main et le déplacement de l'objet apparaissant en 3. comme des détails accessoires. L'envers ne vaut pas l'endroit. Il est à remarquer de plus que le temps ne se mesure que par des aiguilles en mouvement sur des cadrans, ne se figure que par des lignes, et que B. Pottier lui-même, quand il ne recourt pas aux mots, recourt à des schémas pour figurer les concepts les plus abstraits. Nous sommes donc de ceux qui croient que l'espace est l'image du temps et que "*nihil est in intellectu quod non fuerit prius in sensu*". Les métaphores spatiales, tout inconscientes qu'elles sont, n'en structurent pas moins notre imaginaire et, comme l'ont bien mis en valeur Lakoff et Johnson (1980), sont devenues pour nous une sorte de "milieu de vie".

B) Polysémie sans mécanisme de subduction

a) certaines reposent (en ce qui concerne les verbes et les noms abstraits qui les nominalisent) sur des combinaisons actantielles différentes: réduction du nombre des actants exprimés, interversions des places; c'est ainsi qu'on peut traiter les verbes dits "symétriques" et, par ex. les deux acceptions principales du verbe *défendre*

b) d'autres résultent de "spécialisations contextuelles": un noyau sémique, généralement simple et vague, attaché à un signifiant (ex. *hôtel* "bâtiment important, en agglomération, servant à abriter des êtres humains") ne sert qu'à propos de certaines classes de référents (*hôtel de tourisme, hôtel particulier, divers locaux administratifs*).

c) D'autres, enfin, relèvent d'un mouvement de pensée inverse de la subduction, pour lequel on pourrait proposer, par contraste, le nom de *transduction*, qui consiste à réutiliser tout ou partie d'un sémème A dans un sémème B éventuellement aussi riche, ou plus, que le précédent, cette opération pouvant facilement se répéter plusieurs fois et constituer une chaîne de transformations. C'est celui dans lequel s'inscrit la *métonymie*. Le *verre* nous offre un bon exemple de métonymies en chaînes aboutissant à un sémème terminal qui n'a plus un seul sème en commun avec le sémème initial: sémème A: *verre*: «substance dure, cassante et transparente fabriquée à partir de silicates alcalins» -sémème B: «récipient

individuel fait de verre, destiné à boire des liquides froids» - sémème C: «le même récipient, fait d'une substance ressemblant au verre» sémème D: «contenu d'un tel récipient» -sémème E: «quantité de liquide pouvant être contenue dans un tel récipient de taille normale, soit 20 cl.»

L'ordre des acceptions s'impose aussi fortement que dans le cas de subduction. Il y a des polysémies intelligibles si on les traite comme des métaphores et absurdes si on les traite comme des métonymies. Il y a des chaînes de métonymies qui se déroulent sans difficulté si on les prend par le bon bout et qui se mettent à grincer et à coincer si on les prend à l'envers. Essayez un peu d'engendrer, à partir de la notion de "20 cl", celle de "substance dure, cassante et transparente fabriquée à partir de silicates alcalins".

II. La formulation des définitions

1) les définitions ont l'ambition d'être plus faciles à comprendre que le défini, de n'utiliser, dans la mesure du possible que des mots figurant dans la liste de base: Un exemple simple nous est fourni par le mot *coin* dont les trois subdivisions sont les suivantes: A. "espace étroit formé par deux côtés qui se rejoignent" -B. "espace étroit" -C. "espace peu visible": *côté* et *espace* sont dans la liste de base et *étroit* se trouvera sous *large* à titre d'antonyme et *rejoindre* sous *rencontrer* à titre d'homonyme. J'introduirai le mot *angle*, plus rare et plus savant, comme synonyme d'une des acceptions de *coin*, mais je ne l'utiliserai pas dans la définition.

Le métalangage est réduit au minimum et n'utilise que les mots les plus usuels de la grammaire traditionnelle.

2) elles ont l'ambition d'être cohérentes: Je ne me contente pas de définir les diverses acceptions d'un polysème en ne me souciant que de ses oppositions avec ses parasyonymes. J'essaie de prendre une vue globale de l'ensemble de la polysémie, et je cherche s'il est possible de rédiger les définitions des diverses acceptions en utilisant les mêmes définisseurs (non certes, sans les combiner différemment ni sans en éliminer en cours de route).

3) Elles ont également l'ambition de ne pas être des trompe-l'oeil: Elles évitent des circularités artificielles (définir *défendre* par *interdire* et *interdire* par *défendre*), mais pas une circularité conforme à la nature des choses (*voir* figure dans la définition de *oeil* et *oeil* dans la définition de *voir*). Le métalangage est réduit au minimum et n'utilise que les mots les plus usuels de la grammaire traditionnelle

4) les exemples, forgés, non littéraires, appartiennent au vocabulaire le plus courant, le plus moderne, et le plus indispensable. L'archaïsme, l'argot, l'excessive technicité sont évités. Les niveaux de langage sont signalés. Les locutions, toutes les fois qu'elles ne sont pas totalement démotivées, se trouvent à la place où leur justification sémantique apparaît le mieux

5) Enfin, au moins dans le cas des verbes et même dans le cas de certains noms, j'ai décidé de simplifier la vie à Moussa au moins sur un point où les dictionnaires usuels se la compliquent bien inutilement: la définition des verbes, qu'ils rédigent tous à l'infinitif. Or l'infinitif présente un inconvénient majeur: il n'a pas de sujet, ce qui oblige le rédacteur à des contorsions syntaxiques du genre *en parlant de chose bien gênante*, car il existe une relation sémantique évidente entre le sujet du verbe et ses compléments essentiels, de sorte que le sujet commande dans une certaine mesure toute la suite! Selon que vous trouverez, devant une forme conjuguée à la 3e personne du singulier du verbe *porter*, le nom d'un humain, *Jean*, ou d'un objet concret: *le canon*, ou un nom abstrait comme *la colère* la phrase se terminera de façon fort différente: *Jean porte la barbe* - *le canon porte à 10 kilomètres* - *la colère porte Jean à des excès regrettables*. Je pris donc la décision initiale et fondamentale de rédiger mes définitions verbales à la 3e personne du présent de l'indicatif. Elle entraînait, comme conséquence, l'obligation de remplir, dans la définition même, toutes les places qu'un verbe donné offre non seulement à son sujet, mais à tous ses compléments essentiels c'est-à-dire indispensables pour révéler le sens complet du verbe. Le sujet, le complément d'objet et le complément d'attribution sont, dans la terminologie de Tesnière "Prime actant", "second actant" et "tiers actant". Il m'a semblé inutile de me limiter au nombre 3, j'ai pris le parti d'étendre la notion d'actant à l'ensemble des compléments essentiels (dont plusieurs sont comptabilisés par la tradition grammaticale parmi les circonstants.). La distinction entre actants et circonstants sera éclairée par la comparaison des deux types d'exemples ci-dessous où les "actants" sont en gras:

1. *Le match a duré deux heures / longtemps avant la victoire du Paris-Saint Germain* - 1' *Jean a réfléchi deux heures / longtemps avant de se décider* - 2. *Jean va / réside à Paris* - 2' *Jean travaille à Paris* - 3. *Jean est / paraît / semble / devient / se sent vieux !* - 3. *Vieux*, *Jean est moins fou que par le passé* - 4. *Jean est / paraît / semble / devient / se sent à l'aise* - 4'. *Jean est assis à l'aise* - 5. *Jean se porte à merveille, les affaires de Marc vont mal* - 5'. *Jean travaille à merveille, Marc s'occupe mal de*

ses affaires - 6. Eric mène une vie dissolue - 6'. il faut être philosophe, dans la vie - 7. Marc comprend qu'Alice le trompe - 7. Marc comprend, quand on lui explique bien 8. Marc demande à Eric / comment fonctionne le moteur à explosion - 9. Sylvie apprend à tricoter - 10 l'assassin frappe sa victime à la tête avec un marteau - 10' Jean lit un livre dans son bureau avec ses lunettes

Il est clair qu'il n'est pas possible de supprimer un actant sans mutiler la phrase et la laisser en suspens, alors que la suppression d'un circonstant laisse une phrase complète. Une forte isotopie entre un verbe et son complément rend parfois possible la suppression d'un actant en surface; il reste implicite en structure profonde: *Le match a duré, avant la victoire* signifie évidemment "a duré longtemps, plus longtemps que je ne pensais"

En résumé, il n'y a pas une catégorie grammaticale qui, dans un cas ou dans un autre, ne puisse jouer le rôle d'actant, et le nombre des actants est variable, dans certains cas, supérieur à trois de plusieurs unités.

Divers *cas de conscience terminologiques* se posèrent à moi. et ce n'est qu'au bout de multiples tâtonnements que j'ai adopté la solution qui apparaît dans les documents en annexe. J'ai essayé successivement d'éviter les définitions et de remplacer les A par des exemples vraiment exemplaires, mais les exemples sont toujours trop particuliers, donc équivoques, et cela ne permet pas une consultation rapide.

J'ai essayé de les remplacer par des noms de fonctions sémantiques (agent, patient, observateur, interlocuteur etc.) mais la terminologie n'était pas fixée, les flottements dans la désignation des actants étaient inévitables et rendaient cette solution inutilisable.

J'ai essayé de les remplacer par des noms de fonctions grammaticales (sujet, objet, complément). Ce type s'est révélé être le plus catastrophique de tous, à éviter absolument, parce qu'on ne sait jamais si on est dans le domaine syntaxique ou dans le domaine sémantique: *objet* peut signifier "chose concrète de petites dimensions", "finalité d'un acte de pensée" ou "complément d'un verbe"; *sujet* peut signifier "être humain, siège d'affects et de pensées" ou "mot qui détermine la personne du verbe" et souvent deux de ces choses à la fois. Parce que, de plus, en cas de transformation, les différents actants changent de fonction grammaticale; au passif, notamment, l'objet devient sujet et le sujet complément.

Or rien de plus important, sémantiquement, que la stabilité dans la désignation des actants, quelle que puisse être dans les structures de surface leur fonction syntaxique.

Je me suis résolue par élimination des trois précédentes, à adopter la solution qui consiste à désigner les actants par A1, A2, A3 etc. . La méthode de numérotation retenue est de partir de la phrase de base la plus simple et la plus complète possible, de numéroter les actants dans l'ordre où ils se présentent le plus naturellement dans cette phrase et de s'imposer la règle de toujours conserver à cet actant son numéro quelles que soient les transformations possibles et la variété des structures de surface. Ces A numérotés signalant les fonctions sémantiques essentielles sont ensuite spécifiés de la façon sémantique la plus simple et la plus appropriée au mot traité, et ensuite, si nécessaire, défini syntaxiquement avec les termes techniques les plus usuels de cette discipline. J'ai renoncé à désigner le complément de lieu essentiel par L, le complément de temps essentiel par T, à cause des incertitudes de la dénomination des compléments circonstanciels dans la grammaire traditionnelle. D'où des formules comme «A1 humain lit A2 écrit» «A1 est A2, qualité de A1, adjectif attribut» ou «A1 animé, mobile, va du lieu A2 au lieu A3» ou «A1 dure A2, espace de temps» ou «A1 va A4, manière d'être». Donc «A1 voit A2: A1 animé, pourvu d'yeux a conscience de A2 concret, éclairé» est la définition illustrée par l'exemple *Les invités voient la vedette* et «A1 apparaît à A2: A1 concret, éclairé, absent à l'instant A3, et présent à l'instant A4, est soudain vu par A2» On arrive parfois, pour certains verbes à un nombre d'actants diversement spécifiés dépassant quatre ou cinq, ce qui est déjà considérable, pour s'y reconnaître.

Il aurait été bien tentant d'établir une stabilité absolue dans la désignation des actants à travers tout le dictionnaire, décidant que A1 serait toujours un agent, A2 un patient, A3 un destinataire, A4 un lieu, A5 un temps etc. Mais c'était chose impossible pour diverses raisons: Pour le verbe *passer*, par ex. on a besoin de trois lieux A4, A5, A6... Comment faire si j'ai décidé a priori un seul actant spatial? vais-je mettre des A4, A4', A4''? Plus fondamental: regardons les relations sémantiques qui existent entre le verbe *voir* et le verbe *apparaître*: «A1, agent voit A2, patient.» D'accord? Mais si j'ai à traiter l'exemple *La vedette apparut dans une robe de lamé d'or*, vais-je désigner la vedette par A1 ou par A2? Il est bien évident qu'aucun A ne peut *apparaître* s'il n'y a en même temps un autre A pour le *voir*. Qui est agent, qui est patient, qui est destinataire, là-dedans? le A qui *fait son apparition* ou l'autre A qui, même s'il n'est pas exprimé,

invisible en surface mais présent en profondeur, *voit, observe, admire* le premier et qui peut fort bien rester sous-entendu? *Voir* et *apparaître* se trouveront aux deux bouts du dictionnaire. Outre la rigidité insupportable de la formule généralisant la désignations des actants, il serait fort gênant d'avoir à courir d'*apparaître* à *voir* pour essayer de comprendre pourquoi ce ne serait pas *la vedette* qui serait désignée par A1 et rester perplexe tant il est évident que si elle *est vue*, c'est parce qu'elle a fait, de la façon la plus agentive du monde, tout ce qu'il fallait pour *se faire voir*, comme le montre bien la *robe de lamé d'or*, précision circonstancielle, certes, mais non dénuée d'importance.

Il faut savoir limiter ses ambitions ! Etablir une stabilité dans la désignation des actants à l'intérieur d'une seule polysémie, c'est déjà beaucoup

On trouvera en annexe l' article FRAPPER rédigé selon les principes ci-dessus exposés de la "définition actancielle";

L'inconvénient de cette solution est qu'il donne aux articles un faux air algébrique rebutant au premier abord. Mais enfin, c'est de loin la plus pratique parce qu'elle évite toute ambiguïté terminologique et permet de rendre compte clairement des relations fondamentales entre les différents ingrédients sémantiques nécessaires au sens plénier d'un verbe. Il me semble que c'est une solution de bon sens.

Je pense que Moussa, qui n'est peut-être pas un linguiste confirmé, mais qui est doué, comme tout être humain, de la faculté d'abstraction, s'habituerait facilement, si on le lui explique bien en introduction, au système des A numérotés du type IV, alors qu'il aurait eu des problèmes bien pires avec les types I, II et III. Descartes ne nous-a-t-il pas appris que "le bon sens est la chose du monde la mieux partagée?"

Espérons aussi qu'il saura feuilleter *un index*, car si les mots-vedettes sont en très petit nombre, les mots cités dans les articles au titre d'actants spécifiques, de synonymes, d'antonymes et de dérivés foisonnent. Pour que l'utilisateur puisse les retrouver, il faut bien qu'ils se trouvent quelque part en ordre alphabétique. Dans les dernières pages du livre, évidemment!

III. Applications pédagogiques

La *Didactique du vocabulaire français* est, du moins dans sa seconde partie, le mode d'emploi du dictionnaire.

Dans la première partie, je donne des exercices fondés sur les méthodes d'analyse du sens en cours aujourd'hui chez tous les lexicologues (dérivation, structures syntaxiques et analyse sémique). Dans la seconde partie, qui m'est plus personnelle, je fais travailler sur les structures actanciennes pour enrichir leur vocabulaire: dénomination des actants spécifiques à un certain verbe, et qualifications de ces actants. Je fais aussi construire des structures actanciennes autour de certains noms en examinant leurs cooccurrences les plus attendues dans leurs diverses fonctions syntaxiques. Un des intérêts de ces structures actanciennes est de mettre en valeur l'existence de "familles sémantiques" (ex. *voir, clair, lumière, couleur, sombre, obscur*) qui n'ont rigoureusement aucun trait morphologique commun, et qui appartiennent à des catégories syntaxiques différentes, mais qui recouvrent des structures mentales extrêmement cohérentes.

D'autre part, pour leur donner le moyen d'employer le mot juste, et leur donner le sentiment de l'unité du polysème, je les fais travailler sur les locutions figées, le passage du concret à l'abstrait, les métaphores.

Travailler dans ce sens amène à des comparaisons de structures polysémiques révélant que chaque mot a son signifié de puissance bien particulier: jusqu'à *clair* et *clarté* qui ont un dynamisme sémantique différent, le premier beaucoup plus puissant que le second, mais qu'ils présentent aussi des "airs de famille". L'inventaire de leurs ressemblances et de leurs différences me paraît plein de promesses: ainsi *savoir*, orienté vers le *pouvoir* que donne la science et le renseignement et *connaître*, orienté vers l'*avoir* que constitue une expérience et une culture. C'est principalement par ce biais que les apprenants de français langue étrangère pourront découvrir dans cette langue et dans la leur deux "visions du monde" différentes. Le terme de ce travail est donc la prise de conscience des structures mentales sous-jacentes au lexique, qu'elles soient abstraites ou qu'elles comportent toute la symbolique des mots concrets. Ainsi peut-on espérer, pour l'utilisateur, un affinement des instruments lui permettant de penser l'univers.

Bibliographie

- ALFA -1992- vol. 5: *Métaphore, définition, description* - 265 p. [revue de l'Université Dalhousie (Halifax, Nova Scotia, Canada) dirigée par R. Kocourek].

- Bonhomme (Marc) 1987 - *Linguistique de la métonymie* - préfacé par M. Le Guern - Berne - Peter Lang - 324 p.
- Brunet (E.) 1981 - *Le vocabulaire français de 1789 à nos jours d'après les données du "Trésor de la Langue Française"* - Genève-Paris - Slatkine-Champion - vol. I 852 p., vol. II 518 p., vol. III 453 p.
- Dubois (J.) et alii - 1966 - *Dictionnaire du français contemporain* - Paris - Larousse - 1224 p.
- Dubois (J.) et alii - 1975 - *Lexis, dictionnaire de la langue française* - Paris - Larousse - 1950 p.
- Engwall (G.) - 1984 - *Vocabulaire du roman français (1962-1968), Dictionnaire des fréquences* - Stockholm - Almqvist et Wiksell.
- G. Gougenheim et alii - 1956 - *L'Elaboration du français fondamental 1er degré* - Paris - Didier.
- Hausmann (F.-J.) et alii - 1989 - éd. de *Wörterbücher - Dictionaries - Dictionnaires, ein internationales Handbuch zur Lexicographie - An International Encyclopedia of Lexicography - Encyclopédie internationale de lexicographie* - Tomes 1 et 2 - Berlin-New York - Walter de Gruyter.
- INALF - 1971 - *Dictionnaire des fréquences du Trésor de la Langue Française* - Paris - Didier - 6 vol.
- Juilland (A.), Brodin (D.), Davidovitch (C.) 1970 - *Frequency dictionary of french words* - The Hague-Paris - Mouton.
- Lakoff (G.) and Johnson (M.) - 1980 - *Metaphors we live by* - The University of Chicago Press - Chicago and London - 237 p. (trad. française, *Les Métaphores dans la vie quotidienne*, Editions de Minuit, 1985).
- Le Guern (M.) - 1971 - *Sémantique de la métaphore et de la métonymie* - Paris - Larousse - 126 p.
- Mel'uk (I.) - *Dictionnaire explicatif et combinatoire du français contemporain, recherche sémantique* - les Presses de l'Université de Montréal - vol. 1 1984 - vol. 2 1988 - vol. 3 1992.
- Muller (Ch.) - 1973 - *Initiation aux méthodes de la statistique linguistique* - Hachette-Université.
- Muller (Ch.) - 1977 - *Principes et méthodes de statistique lexicale* - Paris - Hachette-Université.
- Muller (Ch.) - 1982 - *Une Nouvelle façon de voir le lexique: le "Brunet", Le Français moderne*, n° 4.
- Picoche (J.) - 1986 - *Structures sémantiques du lexique français* - Paris Nathan 142 p.
- Picoche (J.) - 1993 - *Didactique du vocabulaire français* - Paris - Nathan 206 p.

- Picoche (J.) - 1995 - *Combien y a-t-il de "coeurs" en français?* - dans Langue française n° 105: *Grammaire des sentiments* - pp. 120 à 125 - Paris - Larousse.
- Picoche (J.) - 1995 - *Etudes de lexicologie et de dialectologie* rassemblées par Nelly Andrieux et Geneviève Hasenohr - Conseil international de la langue française - 11 rue de Navarin 75009 Paris - 394 p.
- Picoche (J.) - 1995 - *Définitions actanciennes*. - dans *Cahiers de Lexicologie* - n° 66 pp. 67-76 - Paris - Didier-érudition.
- Picoche (J.) et Honeste (M.-L.) - 1992 - *Un dictionnaire des mots français de haute fréquence - Projet et bilan provisoire* - (Actes du XXe congrès international de Linguistique et Philologie romanes - Zurich - 2-11 avril 1992 - Tome IV - section VI - pp. 269 à 281).
- Picoche (J.) et Honeste (M.-L.) - 1993 - *L'expérience de l'espace et sa symbolisation vue à travers la polysémie des mots bord et côté* - dans la revue *Faits de Langue* n°1 - *Motivation et iconicité* - Paris PUF - pp. 163-173.
- Pottier (B.) - 1992 - *Sémantique générale* - Paris - PUF - 231 p.
- Rey (A.) et Chantreau (S.) - 1979 - *Dictionnaire des expressions et locutions figurées* - Paris - Le Robert - 946 p.
- Ruhl (C.) - 1989 - *On monosemy, a study in linguistic semantics* - Albany, NY - State University of New York Press - 299 p. - reviewed by R. Landheer - *Journal of Pragmatics* 15 (1991, febr.) pp. 210-215.
- Tamba-Mecz (I) - 1981 - *Le sens figuré* - Paris - PUF - 199 p.

ANNEXE

ARTICLES DE DICTIONNAIRE:

on notera que le symbole Σ introduit des considérations syntaxiques, le symbole ¶ une locution figée, et le symbole Δ un dérivé, morpho-sémantique, populaire, comme *frappe* dans l'article FRAPPER ou savant, comme *cardiaque* dans l'article COEUR ou sémantique seulement, comme *coup* - qui ne figure que pour mémoire dans l'article FRAPPER, faisant l'objet d'un article distinct. Π signale un proverbe.

Exemple de mot abstrait avec mécanisme de subduction et un seul cinétisme

INTERET nom masc., INTERESSER v. trans. et pronom., INTERESSANT adj.

I. A1 humain tire un profit financier de A2, affaire d'argent, qui l'intéresse

1. Jean a prêté une certaine somme d'argent à son ami Luc, ou à une entreprise, ou à une banque. Il est convenu que Luc, l'entreprise, la banque, lui rendront cet argent au bout d'un temps convenu, grossi d'un pourcentage convenu. La différence entre la somme prêtée (appelée le *capital*) et la somme rendue c'est l'*intérêt* de ce capital. Cet *intérêt* constitue pour Jean un enrichissement, donc des moyens de vivre. Lorsqu'on prête à une société ou à une banque qui *s'oblige* à rendre au prêteur son capital et à lui en verser en plus un certain pourcentage, elle fournit un reçu appelé OBLIGATION - Ex: "l'*intérêt* de ces obligations est de 10% sur 12 ans" - Dans ce cas, le mot *intérêt* est souvent au pluriel: "Jean a touché *les intérêts* de ses obligations" (Syn.: RENTE (fém.), DIVIDENDE, AGIO (noms masc. souvent plur.)

2. Paul a causé un dommage quelconque à Jean: par ex. il l'a blessé dans un accident et il est resté infirme. Un tribunal oblige Paul à verser à Jean une certaine somme d'argent pour compenser plus ou moins l'appauvrissement qui en résulte pour Jean. Cette somme s'appelle des "dommages et *intérêts*" (plur. obligatoire - Syn.: INDEMNITÉ (fém.) DÉDOMMAGEMENT (masc.)

3. Les *intérêts* de Jean, c'est la bonne gestion de l'ensemble des affaires financières engageant son avenir. Jean a le souci de *ses intérêts*. S'il est orphelin et mineur, un tuteur serait désigné pour veiller sur *ses intérêts* (pluriel obligatoire): Si tout va bien, en arrivant à sa majorité, il ne sera pas plus pauvre et sera même un peu plus riche qu'à sa naissance.

4. a) c'est l'*intérêt* de Jean de placer son argent au Crédit Populaire de Guyenne, il y a *intérêt*: il s'enrichira ainsi plus, ou plus sûrement, qu'en utilisant son argent autrement; il aura grâce à cela des moyens qu'il n'avait pas auparavant. (A partir d'ici et dans les acceptions suivantes, *intérêt* n'apparaît qu'au singulier). Il s'en rend bien compte. Dès lors, Jean *s'intéresse*

à ce placement, *est intéressé* par ce placement. Autrement dit, ce placement *intéresse* Jean; ce placement *a de l'intérêt* pour Jean; c'est un placement *intéressant*.

b) A3 *intéresse* A1 à A2: Marc, patron d'une entreprise, beau-père de Jean, *l'intéresse* à son affaire: il lui attribue une part des bénéfices. De plus, il le prend pour collaborateur, espérant ainsi qu'il y travaillera efficacement, avec une certaine ardeur, et pourra lui succéder. On peut parler de l'**INTÉRESSEMENT** de Jean aux affaires de Marc. Supposons maintenant que Jean ait des dettes: il peut **DÉSINTÉRESSER** ses créanciers en les remboursant.

c) Si Jean, dans toutes ses actions est préoccupé avant tout par l'argent qu'il gagnera, ou du moins ne perdra pas, on dira qu'il est **INTÉRESSÉ**; c'est un trait de son caractère. Si au contraire il fait ce qui lui semble bon sans (trop) songer à l'argent, on dira qu'il est **DÉSINTÉRESSÉ**; qu'un de ses traits de caractère est le **DÉSINTÉRESSEMENT**.

II. A1 humain tire un profit pratique de A2 qui l'intéresse

C'est *l'intérêt* de Jean de faire des mathématiques s'il veut devenir ingénieur - c'est *l'intérêt* de Jean d'épouser Sylvie parce que c'est une fille sérieuse et gentille; Jean *a (tout) intérêt* à faire des mathématiques, à épouser Sylvie: ces diverses actions lui donneront le moyen de mieux réussir sa vie que s'il ne les faisait pas. Autrement dit c'est plus *intéressant* pour Jean de faire des mathématiques que de la sociologie; d'épouser Sylvie que cette folle de Véronique.

III. A1 humain tire un profit intellectuel de A2 abstrait qui l'intéresse

Jean lit un livre sur la préhistoire. Ce livre *intéresse* beaucoup Jean. Il est très *intéressant*, *a de l'intérêt*, *beaucoup d'intérêt*, *un grand intérêt* aux yeux de Jean: il y trouve un savoir qui enrichit son esprit, qui lui donne le moyen de mieux comprendre l'histoire de l'humanité et sa situation propre. Il en éprouve un grand plaisir et a envie d'en savoir davantage. Autrement dit Jean *trouve*, *éprouve un grand intérêt* à la lecture de ce livre; *l'intérêt de ce livre pour Jean* est grand; Jean *s'intéresse* à ce livre; *l'intérêt de Jean pour ce livre* est considérable.

A3 *intéresse* A1 à A2: Un professeur a jadis su *intéresser* Jean à la préhistoire.

IV. A1 humain tire une satisfaction par sympathie de A2 humain qui l'intéresse:

Luc est plus jeune, plus faible, plus pauvre que Jean. Jean veut l'aider parce qu'il reconnaît ses qualités, éprouve de la joie quand il réussit. On peut dire que Jean *s'intéresse* à Luc; que Jean éprouve de *l'intérêt* pour Luc, que

Jean trouve Luc *intéressant*; que Luc *intéresse* Jean. (Syn.: SYMPATHIE, BIENVEILLANCE, PITIÉ, CHARITÉ (noms fém.)

V. A1 humain est concerné par ce qui l'intéresse

Jean reçoit une lettre d'une administration comportant la formule "l'intéressé devra se présenter tel jour à telle heure pour régularisation". C'est lui, l'intéressé. Cette lettre l'intéresse, c'est-à-dire le "concerne" comme elle concerne tous les gens qui sont dans son cas. Même désagréable, elle ne le laisse normalement pas indifférent et il doit y réagir. En ce sens le complément de *intéresser* peut même être non humain: *une blessure qui intéresse le poumon*

VI. A1 humain se désintéresse de A2:

Quel que soit l'objet A2 (financier, intellectuel, humain) auquel Jean s'est *intéressé* un certain temps, le jour où, les circonstances ayant changé, il cessera de s'y intéresser, on dira qu'il *s'en désintéresse*. On peut parler de son **DÉSINTÉRÊT** à l'égard de cet objet.

Exemple de verbe à plusieurs actants

FRAPPER v. trans. et pronom.

I. A1 humain, ou concret en mouvement, *frappe* A2 concret avec A3 (la main, ou un instrument) en A4 (un endroit de A2)

Le boxeur (A1) frappe son adversaire (A2) du gauche (A3) à la poitrine (A4): le boxeur imprime à sa main gauche fermée (autrement dit son poing gauche), en direction de la poitrine de son adversaire, un mouvement rapide et violent au terme duquel le poing entre en contact avec la poitrine de façon soudaine et intense, avec effet douloureux chez l'adversaire. Quand un A1 humain frappe un A2 humain, c'est habituellement avec l'intention de lui faire du mal

Σ: Le résultat de l'action de *frapper* n'est exprimé qu'exceptionnellement, sous forme de complément prépositionnel dans ¶ *frapper à mort*, employée surtout au passif: *la victime est ¶ frappée à mort* et sous forme de complément d'objet direct dans ¶ *frapper de la monnaie*, la monnaie étant le résultat du coup donné sur un lingot de métal. Δ **FRAPPE**, nom fém. 1) action de frapper de la monnaie, ou des médailles: *la frappe des monnaies est un monopole d'Etat* - 2) manière dont cette monnaie est frappée *la frappe de cette médaille est bien nette*.

Diverses transformations sont possibles: *l'assassin frappe sa victime à la tête avec un marteau* - *La victime est frappée par l'assassin, par le marteau de l'assassin* - *le marteau frappe la victime / la tête de la victime*.

Dans un contexte sans ambiguïté, A1 peut être seul exprimé: *L'assassin (A1) a encore frappé !*: Dans ce cas, le sens de *frapper* peut se réduire à: "commettre une action dommageable de façon soudaine et violente" - A 4

ne peut pas exister dans le cas où A1 est un phénomène naturel: *le soleil, la pluie* (A1) *frappent le mur* (A2) - les actants A3 et A4 peuvent facilement n'être pas exprimés: *L'accordeur* (A1) *frappe les touches du piano* (A2) - *la danseuse* (A1) *frappe le sol* (A2) *du pied* (A3) - *Jean* (A1) *se* (A2) *frappe la poitrine* (A4): geste traditionnel de repentance après une action qu'on regrette.

Lorsque A2 n'est pas exprimé, il s'agit le plus souvent d'une porte dont A1 humain demande l'ouverture en signalant sa présence par plusieurs petits coups: *Frappez avant d'entrer* - *On frappe ! va ouvrir!* - Ces coups sont de façon figée représentés par l'onomatopée: *toc, toc* - Avant d'ouvrir, la personne qui est à l'intérieur peut demander: "*Qui est là ?*" - N-B on ne dit pas: A1 **frappe la porte* mais A1 ¶ *frappe à la porte* -

II. A1 humain *frappe un coup* - Δ COUP, nom masc. "action de frapper" (v. ce mot) peut être complément de *frapper*: Dans les théâtres traditionnels, *le régisseur frappe les trois coups*, sur la scène, avec un bâton appelé "le brigadier" pour annoncer le début de la représentation - *frapper un coup décisif*: agir de façon inattendue et déterminante dans une certaine situation -

Σ: *coup* n'est pas un A2 concret extérieur à A1, mais l'action de A1; c'est simplement une nominalisation de *frapper*, ce qu'on appelle un "objet interne".

III. A1, abstrait, circonstance ou phénomène naturel, *frappe* A2 humain

Les impôts qui frappent les contribuables sont excessifs: ces impôts (A1) atteignent certains contribuables (A2) et leur causent un dommage comparable à celui d'un coup.

Quand je suis entrée dans la maison, j'ai vu que le grand tableau du vestibule n'y était plus; cela (A1) *m'a frappé*: la circonstance A1, à laquelle je n'étais pas du tout préparé, et dont j'ai eu connaissance avec une grande soudaineté, m'a causé une surprise comparable à un coup -

Cette circonstance A1 est très souvent une ressemblance qu'on découvre entre deux A: *La ressemblance entre Jean et Luc me frappe* - *c'est une ressemblance FRAPPANTE*.

Dans le cas de ce genre de A1, le passif est très fréquent: *Je suis frappé de la ressemblance de Jean et de Luc*; il est seul possible dans ¶ *être frappé d'apoplexie* - ¶ *être frappé de stupeur* - ¶ *une politique frappée d'impuissance* ¶ *du champagne, du café frappé* ont été refroidis rapidement et intensément, l'effet du froid étant comparé à un coup.

IV. A1, humain, se frappe:

A1, surpris par un changement d'état mauvais et soudain, lui attache une importance excessive, prend peur, vit dans l'inquiétude: *Jean se frappe parce qu'on lui a découvert une petite tumeur ! Ne te frappe pas ! Il faut suivre son traitement sans se frapper* -

Exemple de nom concret combinant métaphore et métonymie

COEUR, nom masc.

1) Le *coeur bat* au centre de notre poitrine aussi longtemps que nous vivons; cet organe profond et essentiel règle la circulation du sang. On pratique aujourd'hui des opérations ¶ à *coeur ouvert*, des transplantations, des greffes du coeur sur les gens qui sont ¶ *malades du coeur*. On ne réchappe pas d'une blessure ¶ *en plein coeur* - On meurt d'un arrêt du coeur - Δ -CARD-, **CARDIO-**, bases de mots savants signifiant "coeur": *cardiaque*: malade du coeur - *cardiologue*: médecin spécialiste du coeur - *péricarde*: enveloppe du coeur etc.

- le coeur de certains animaux est utilisé pour la nourriture: *un coeur de veau aux carottes*

2) Sa forme stylisée est un ensemble gracieux de courbes et de contrecourbes, et elle peut être donnée à toutes sortes d'objets, ainsi, *un ¶ coeur à la crème*, fromage moulu en forme de coeur; *le coeur, rouge*, une des quatre "couleurs" du jeu de cartes: *la dame, le roi, l'as de coeur* - il est *venu me parler*, ¶ *la bouche en coeur*: en arrondissant ses lèvres pour donner à son visage une expression aimable, engageante (ironique); *Marie est ¶ jolie comme un coeur*: extrêmement jolie

3) La partie du corps où il se trouve est particulièrement sensible; elle est proche de l'estomac: *mon déjeuner ¶ m'est resté sur le coeur*: je ne l'ai pas digéré; ¶ *j'ai mal au coeur*, *j'ai ¶ le coeur barbouillé*; *j'ai des ¶ haut-le-coeur*: j'ai des NAUSEES (nom fém.); j'ai envie de VOMIR (v. trans. et intrans.); *cette nourriture ¶ me lève* (ou *me soulève*) *le coeur*: elle me donne envie de vomir; fig.: *Jean s'est conduit de façon dégoûtante; ça me lève le coeur*; *cette injure m'est restée sur le coeur*. Δ -ECOEURER v. trans.: A1 *écoeur* A2 humain: A1 donne des nausées à A2, au propre et au fig. - **ECOEURANT**, adj. qual.: qui donne des nausées

Mais surtout, c'est un centre nerveux dont les sensations sont liées aux états de l'âme (v. *Ame* 4. 2.): c'est là qu'on éprouve

des émotions: *Sylvie, inquiète, attend Jean ¶ le coeur battant* -

des sentiments: on ¶ *prend à coeur* les choses auxquelles on attache de l'importance; elles vous ¶ *tiennent à coeur*, provoquant la tristesse ou la joie: *j'ai le coeur lourd, gros, serré, triste, léger, gai*; *j'ai* (ou, plus souvent, *je n'ai pas*) ¶ *le coeur à rire*; *ça me fait ¶ mal au coeur d'avoir raté cette occasion*; *j'en ai ¶ gros sur le coeur*, *j'ai la rage au coeur*, *le coeur meurtri, blessé*; A1 *fend, perce le coeur* à A2 humain, lui *reste sur le coeur*: "lui fait éprouver une vive douleur, ou des regrets". A1 *gonfler, dilater le coeur* à A2 humain: lui fait éprouver de la joie - l'amour et l'amitié: *Jean a le coeur tendre, épris, ardent, il a donné son coeur à Marie*; *il serre Marie sur son coeur*; *Jean ¶ ne porte pas Eric dans son coeur*: il ne l'aime pas - des ¶ *affaires de coeur*, des ¶ *peines de coeur* sont de nature amoureuse - un ¶ *ami de coeur* est aimé tendrement - mais quand A1

humain ne porte pas A2 humain dans son coeur, c'est qu'il ne l'aime pas - la pitié, la générosité: un blessé, sur la route, a touché le coeur d'un Samaritain, qui ¶ avait bon coeur, ¶ le coeur sur la main, n'a ¶ écouté que son coeur et l'a fait soigner; d'autres passants ne se sont pas arrêtés: ils ¶ n'avaient pas de coeur, ¶ étaient sans coeur - Π Loin des yeux, loin du coeur: on cesse d'aimer les gens qu'on ne voit plus.

C'est là qu'on sent naître des intuitions (citation: le coeur a ses raisons que la raison ne connaît pas); A1 humain a un ¶ coup de coeur pour un A2 concret: il tombe en admiration devant A2, a envie de l'acheter - des questions, des idées mêlées de sentiment: J'ai des soupçons, je veux ¶ en avoir le coeur net: je ne veux pas rester dans une désagréable incertitude; ¶ parler de l'abondance du coeur: discourir avec éloquence, sans préparation, sur un sujet qui vous ¶ tient à coeur - une parole efficace ¶ va droit au coeur de l'auditeur - Jean a le coeur à droite, Marc le coeur à gauche: leurs opinions politiques sont opposées - Lorsqu'on est capable de redire un texte mot à mot, de façon mécanique, sans même y penser, on dit qu'on a appris, qu'on sait, qu'on récite sa leçon ¶ par coeur. C'est dans la région du coeur que naissent les désirs, les décisions, sources d'actes volontaires, éventuellement courageux: Faites-le ¶ si le coeur vous en dit: "si vous le désirez vraiment"; Jean accepte la proposition de Marc ¶ de bon coeur, ¶ de grand coeur, ¶ de tout coeur; il travaille ¶ de tout son coeur; il a ¶ du coeur à l'ouvrage: il le fait bien volontiers, et pas seulement par obéissance; ¶ le coeur n'y est pas: il agit par obéissance, mais ¶ à contre-coeur; ce n'est ¶ pas de gaieté de coeur! - ¶ il n'a de coeur à rien. ¶ Le coeur lui manque; il n'a pas ¶ le coeur de faire cela. Il a ¶ du coeur au ventre (càd., en ancien français, "dans la poitrine"): "il est capable de courage"; un roi d'Angleterre a été surnommé à cause de son courage Richard Coeur de Lion. Citation: Rodrigue, as-tu du coeur? "as-tu du courage?". ¶ Haut les coeurs! "Courage!" (exclamation d'encouragement).

4) Le coeur est, en nous, ce qu'il y a de plus profond et de plus secret (v. Ame 4. 1.); le refuge de notre liberté; tout ce que nous faisons de bien et de mal en provient; ¶ un homme de coeur a toutes les qualités de générosité, dévouement, courage; le coeur s'oppose à la bouche et à la mine qui peuvent mentir; ¶ sonder les coeurs, est un privilège de Dieu; Il est rare qu'un homme puisse ¶ lire dans le coeur d'un autre - ¶ un coeur double est celui d'un traître; Jean remercie Marc ¶ du fond du coeur: avec la plus grande sincérité; C'est ¶ le cri du coeur: un cri absolument sincère qui lui échappe sous le coup d'une grande émotion; entre Jean et Marie, c'est l'union des coeurs: une entente parfaite. Jean ¶ ouvre son coeur à Marie, il lui parle ¶ à coeur ouvert, laisse parler son coeur, ses paroles viennent du coeur: "il lui dit franchement tout ce qu'il pense". Il lui parle ¶ la main sur le coeur pour mieux la convaincre (geste peut-être un peu hypocrite)

5) On emploie donc parfois le mot coeur pour désigner la personne tout entière (v. Ame 5): Pierre ¶ fait le joli coeur: se montre galant, courtise les femmes; mon coeur, mon petit coeur, sont des manières tendres de s'adresser à

un enfant, une jeune fille; *Jean est un coeur d'or*, 'une personne sensible à la pitié et très généreuse'; ¶ *un sans-coeur*: "une personne sans pitié"; *un grand coeur*: "une personne capable de dévouement et de désintéressement"; *un coeur d'enfant*, *un coeur simple*: sans malice ni ambition - *un brave coeur*: "une personne humble, honnête et dévouée"; le *Sacré Coeur*: Jésus-Christ en tant qu' Amour de tous les hommes (et les diverses églises qui lui sont dédiées) .

6) Le centre, l'élément le plus important de quelque chose de matériel ou d'abstrait, peut être appelé son *coeur*: *du coeur de chêne*; *un coeur de laitue*, *d'artichaut*; *être au coeur de l'hiver*, *au coeur de la ville*; *au coeur du problème* . (v. *Ame* 6)

M-L Honeste: Comparaison de plusieurs listes de fréquences:

Celle du TLF repose sur 70.317.234 occurrences, pour 70.000 vocables (corpus littéraire compris entre 1789 et 1965). Nous l'avons comparée: 1) à celle de Juilland (corpus littéraire compris entre 1920 et 1950), qui repose sur 464.134 occurrences, pour 5.082 vocables

2) à celle du Français Fondamental (corpus oral de 1956, enrichi par les auteurs de quelques "mots disponibles" usuels), qui repose sur 312.135 occurrences, pour 7.995 vocables

3) à celle d'Engwall (corpus littéraire compris entre 1962 et 1968) qui repose sur 500.096 occurrences pour 4.853 vocables

Les listes provenant des corpus littéraires (TLF, Juilland et Engwall) sont communes à 80% alors que la liste du Français Fondamental n'est commune avec le TLF qu'à 65%. Elle s'écarte considérablement des deux autres, notamment à cause d'un nombre non négligeable d'interjections et de vulgarismes.

Seuls les 150 premiers mots sont rigoureusement identiques dans les trois listes. La moitié de ces 150 premiers mots sont des lexèmes, alors qu'à partir du rang 150, on a entre 90 et 100% de lexèmes.

A partir du rang 250, 25% de vocables du TLF manquent dans la liste du Français Fondamental, mais le pourcentage de mots manquants dans Juilland et dans Engwall est encore presque nul.

A partir du rang 500, 25% des vocables du TLF manquent dans Juilland et dans Engwall

A partir du rang 550, 50% des vocables du TLF manquent dans le Français Fondamental

A partir du rang 800, 50% des vocables du TLF manquent dans Engwall.

A partir du rang 850, 50% des vocables du TLF manquent dans Juilland et 75% manquent dans le Français Fondamental.

Nous avons complété cette confrontation par un examen statistique de la composition des listes, du point de vue des catégories grammaticales,

comme du point de vue (moins rigoureux celui-là) des domaines de référence. Voici un bref aperçu des résultats:

534 mots sont communs aux quatre listes (avec bien sûr des fréquences inégales), dont 1/4 de mots grammaticaux, proportion assez élevée si on se réfère à la structure moyenne des listes de référence (environ 17% de mots grammaticaux). Il reste 75% de mots lexicaux, constitués d'1/3 de noms, 1/4 de verbes, les 20% restant se répartissant également entre adjectifs et adverbes. La structure des domaines de référence des noms est complexe, du fait de leur forte polysémie; pourtant, en deçà des réseaux qu'ils constituent, on peut arriver à dégager quelques domaines référentiels prépondérants:

Quatre groupes représentent chacun environ 20% de ces 534 mots:

1) celui des mots abstraits qui renvoient à des notions de classement, d'organisation et d'évaluation et

2) celui des termes de temps (dates et durées) et d'espace (localisation et dénominations de lieux).

3) celui de mots désignant la personne humaine (homme, femme, monsieur, etc.), la parenté, les parties du corps, des sensations et sentiments.

4) celui de mots qui renvoient à l'environnement quotidien (objets, animaux, lieux).

L'examen des mots non communs aux quatre listes fait apparaître des divergences thématiques sans doute dues essentiellement aux différences de sources et d'époque des corpus:

Sur les 159 mots du TLF qu'on ne retrouve dans aucune autre liste, il y a plus de 75% de noms, les catégories des adjectifs et des verbes étant inférieures à 10%. Les domaines de référence sont très spécifiques, du fait de l'origine littéraire et étendue dans le temps du corpus. On y trouve: des désignations d'institutions et de groupes humains (gouvernement, nation, société, race, peuple, humanité), des titres de noblesse; du vocabulaire relatif à la culture, notamment à la littérature (philosophie, poésie, livre, ouvrage, chapitre). Les autres mots abstraits appartiennent en grande partie à des groupes spécifiques peu représentés dans les autres listes, concernant les sentiments ou les valeurs morales. Les noms concrets renvoient eux aussi à des univers de fiction littéraire: aile, ange, arme, branche, château, palais, désert, flot, rayon, source.

Sur les 351 mots du Français Fondamental absents du TLF, à peine la moitié sont des noms et près d'1/4 sont des verbes; adjectifs et adverbes sont à égalité (10%): les différentes catégories sont donc représentées à peu près comme dans la liste des 907 mots du TLF, sauf la catégorie des mots grammaticaux, qui ne sont ici que 10% contre 15% dans l'ensemble des 907 mots.

Les domaines de référence des noms sont caractéristiques d'un corpus oral: la proportion de termes concrets dénotant des objets du quotidien est la plus importante (1/3); les mots relatifs à l'homme et à la profession constituent aussi un groupe notable (près de 20%). Les mots abstraits, peu nombreux, renvoient pour la plupart au domaine des sensations.

Sur les 199 mots de Juilland absents du TLF, à peine la moitié sont des noms, comme dans la liste du Français Fondamental, mais la proportion de verbes est plus importante (30%). En revanche, contrairement au Français Fondamental, ce sont les noms abstraits qui sont les plus nombreux (1/4), avec une proportion de mots relatifs à l'économie supérieure à tous les autres corpus; mais le temps et l'espace sont peu représentés, et les mots concrets représentent moins de 15% du total.

Enfin, sur les 283 mots de Engwall absents du TLF, la proportion de noms (35%) et de verbes (50%) est inversée par rapport à Juilland. Les autres catégories sont sous-représentées.

La structure référentielle des noms se rapproche de celle du Français Fondamental: les mots concrets sont les plus nombreux et les mots abstraits, en petit nombre, renvoient surtout à des sensations.

Il semblerait, d'après nos premières conclusions, que les écarts dans les représentations des différentes catégories grammaticales soient plutôt liées à l'âge des corpus (par exemple, la proportion de verbes augmente au fur et à mesure que la source des corpus est plus récente); en revanche, les variations dans les domaines de référence s'éclairent par la composition des corpus (notamment, on remarquera que le corpus oral du FF et la liste Engwall, fondée sur l'exploitation de romans contemporains, présentent les mêmes caractéristiques de "concrétude").

O DICCIONARIO DA LINGUA GALEGA DA REAL ACADEMIA GALEGA

MANUEL GONZÁLEZ GONZÁLEZ
Real Academia Galega

1. Introducción

Un dos obxectivos que se propuxo a Real Academia Galega no momento da súa fundación no 1905 foi a redacción dun dicionario da lingua galega. Pero, por razóns que non é o caso aquí analizar, este dicionario que se empezou a publicar no ano 1913 en cadernos de 16 páxinas en folio a tres columnas, viuse interrompido no 1928 na letra C (exactamente na voz *cativo*). Houbo que esperar moito tempo ata o ano 1990, en que viu a luz o *Diccionario da lingua galega*, publicado conxuntamente pola Real Academia Galega e o Instituto da Lingua Galega. Pero trátase dun dicionario modesto en canto ó número de entradas, aínda que, desde o punto de vista do tratamento lexicográfico, supón un avance fundamental con respecto a todo o que se levaba feito dentro da lexicografía galega.

En realidade o *Diccionario* que hoxe presentamos é unha continuación deste publicado en 1990. Trátase dun dicionario de lingua, de tipo medio, polo número de entradas, que recolle cerca de trinta mil artigos. É ademais un dicionario *normativo*, no que non ten cabida o léxico marcadamente dialectal, e do que por suposto quedan excluídas tódalas formas espurias innecesarias. A forma das entradas segue a recollida no *Vocabulario ortográfico da lingua galega*, publicado en versión provisional polo Instituto da Lingua Galega e a Real Academia Galega no 1989. Para a fixación das formas normativas, fíxose un estudio histórico exhaustivo de cada unha das voces e unha análise da súa vitalidade actual.

Evidentemente tanto no *Vocabulario ortográfico da lingua galega* coma no *Diccionario* da RAG non se lles deu cabida ás voces non necesarias procedentes do castelán, do galego ou doutras linguas. Pero non podemos subtraernos a unha realidade presente na nosa

lingua: a presenza do castelán na sociedade galega ten como consecuencia que un gran número de voces procedentes desta lingua se vaian introducindo no léxico habitual de moitos galegofalantes, e que a voz xenuinamente galega vaia perdendo terreo progresivamente fronte á invasora castelá. Algunhas destas voces non galegas que gozan de certa difusión na fala de moitos habitantes de Galicia, pero que se consideran incorrectas nun galego culto, tiveron tamén cabida no noso dicionario, precedidas dun asterisco, non para darlles certificado de lexitimación, senón para indicar que se trata de voces que deben ser evitadas, e substituídas por outras autenticamente galegas, ás que desde elas se remite.

2. A selección de entradas

Dixemos que se trata dun dicionario de cerca 30.000 entradas, polo tanto un dicionario de dimensións relativamente reducidas, que implica un forte proceso de selección léxica, pero no que non falte a zona lexical común a tódolos locutores que se expresan habitualmente en galego.

Para a escolla das entradas o primeiro principio, aínda que non o único, foi o da súa frecuencia de uso. Se ben os escasos estudos de frecuencias cos que contamos non son estatisticamente fiables por riba das dúas mil palabras máis utilizadas, estes aseguráronnos polo menos de que ningunha das voces de maior presenza na fala cotiá quedase fóra do dicionario.

Procurouse non dar cabida a palabras dunha familia con significado que sexa totalmente deducible a partir doutros membros da familia máis frecuentes que aparecen recollidos como entradas. Por exemplo, non teñen cabida nesta obra os adverbios en *-mente*, porque o seu significado é claramente deducible a partir do do adxectivo sobre o que se forman. Só figuran aqueles casos en que o adverbio adquiriu un significado especial distinto do da base. Isto explica que, por ex., non apareza *cortesmente* que sempre significa 'de maneira cortés', pero que en cambio haxa unha entrada *seguramente*, porque o seu significado máis habitual non é 'de maneira segura', senón 'con moita probabilidade'. Un problema semellante preséntano os adxectivos en *-ble*, que tampouco aparecen recollidos, a non ser cando teñen un significado especial non deducible da base ou cando xa non existe na lingua actual a base sobre a que se formaron.

Introduciuse tamén un certo número de tecnicismos e de léxico especializado dos distintos ramos da ciencia, procurando que non

faltasen aqueles que teñen unha maior presenza no mundo escolar non universitario.

Non podemos esquecer que o galego é unha lingua en proceso de normalización, e que como tal aínda se atopa con certos problemas de estandarización. E este feito condicionou a presenza dalgúns termos neste dicionario. Por exemplo, a voz *fresa* 'instrumento para ampliar un burato' probablemente non debería incluírse nesta obra; pero déuselle entrada para indicar que en galego só existe esta *fresa* e non o seu homónimo que designa un froito que en galego é coñecido co nome de 'amorodo' ou 'amorote'.

3. Corpus co que se traballou e material auxiliar.

3.1. Unha boa parte do material de base co que se redactou esta obra constitúe o *Tesouro lexicográfico da lingua galega* do Instituto da Lingua Galega, que actualmente conta cunhas oitocentas mil fichas de léxico do galego vivo¹.

3.2. Outra fonte que se tivo en conta para a redacción deste dicionario foi o léxico utilizado polos escritores en lingua galega na época moderna, non só o daqueles que fixeron uso da nosa lingua con pretensións literarias, senón tamén o dos que a empregaron en obras e escritos de etnografía, ensaio, xornalismo, etc.

3.3. Aínda que este dicionario non ten carácter acumulativo, de recopilación do léxico xa recollido nos dicionarios e obras lexicográficas anteriores, revisáronse tódolos traballos lexicográficos coñecidos sobre o galego (dicionarios monolingües e bilingües; monografas de carácter dialectal; glosarios de obras literarias; léxicos especializados, etc.)

3.4. Tivéronse presentes en todo momento os bos dicionarios doutras linguas románicas e non románicas do ámbito europeo occidental, co que temos unha maior proximidade sociocultural.

¹ Unha descrición do contido e organización deste ficheiro pode verse en Manuel González González, "Os traballos lexicográficos no Instituto da Lingua Galega", *Actas do XIX Congreso Internacional de Lingüística e Filoloxía Románicas*, Fundación "Pedro Barrié de la Maza, Conde de Fenosa", A Coruña, vol. VIII, 1996, pp. 771-786.

3.5. Outra fonte importante para o tratamento dunha parcela do léxico está constituída polos dicionarios científicos e técnicos, que se mostraron útiles sobre todo para a definición dun bo número de cultismos e préstamos doutras linguas propias do mundo da ciencia e da técnica.

4. Microestructura

Cada entrada dun dicionario, por riba da aparente diversidade, encerra unha estrutura considerablemente rixida. Os diversos tipos de información seguen unha ordenación preestablecida e están marcados formalmente por unha serie de símbolos e normas tipográficas. Vexamos cá é a organización da información no noso dicionario.

4.1. A entrada vai en letra grosa. Cando aparece precedida dun asterisco quere dicir que se trata dunha forma non recomendada, por ser unha voz non galega, ou ben porque, aínda sendo galega, ten un carácter marcadamente dialectal ou vulgar, ou porque é recomendable outra ortografía. Por ex.:

- *mimo² *s.m.* meco
- *gavián *s.m.* gabián
- *postre *s.m.* sobremesa

As entradas homógrafas van diferenciadas por un número ó final en superíndice:

á¹ Contracción da prep. *a* e do art. f. *a*. *Virou á esquerda.*

á² *s.f.* 1. Membro do corpo das aves, dos insectos e dalgúns outros animais que lles serven para voar. *O morcego é un mamífero con ás.* 2. Cada unha das partes laterais do avión que presentan ó aire unha superficie plana e permiten mante-lo aparato en voo...

4.2. Indicación sobre a pronuncia. Naqueles casos en que pode existir algunha dúbida sobre a maneira correcta de pronunciar algunha palabra indícase a pronunciación entre corchetes.

É de sobra coñecido que en galego, coma en moitas outras linguas, hai grafemas que poden representar varios fonemas. En moitos destes casos é necesario aclarar a cal fonema ou grupo fonemático representa un grafema na palabra en cuestión. Por ex.:

fecha [e] s.f.
póla [ɔ]
paradoxo [ks]

4.3. A continuación, os substantivos e adxectivos que teñen moción de xénero, levan indicada a forma do feminino:

cliente -a s
rapaz -aza s
bailarín -ina s
avó -oa s
catalán -ana *adx.*
irmán s.m. (f. irmá)
ladrón s.m. (f. ladroa, ladrona, ladra)
herexe s.

Nalgúns casos o feminino exprésase por medio dun morfema derivativo. Cando isto ocorre dásele entrada á forma de cada un dos xéneros. Por ex., *sacerdote* e *sacerdotisa* aparecen como entradas distintas, e o mesmo ocorre con *actor*, *actriz*, aínda que o tratamento que se lle dea ó artigo vaia ser diferente nos dous casos: no primeiro tanto *sacerdote* como *sacerdotisa* levan unha definición, porque teñen significados distintos (é dicir, a diferenza entre *sacerdote* e *sacerdotisa* non se limita á simple oposición masculino/feminino); en cambio no caso de *actor*, *actriz*, unicamente aparece definida a entrada *actor*, e en *actriz* remítese para *actor*.

Cando a diferenza de xénero se expresa por medio de heterónimos, é dicir por un lexema distinto (*home/ muller; carneiro/ovella...*) ámbalas palabras aparecen como entradas.

4.4. A flexión de número só se indica naqueles casos de formacións irregulares. Xeralmente isto ocorre con voces non patrimoniais:

shérif s.m. (pl. shérifs).

4.5. Tódalas entradas levan indicada por medio dunha abreviatura a súa categoría gramatical, a non ser no caso de palabras que só aparecen en certas locucións, por ex.:

esguello Palabra empregada na *loc. adv. de esguello...*

Ademais da categoría, sinálase o xénero no caso dos substantivos e adxectivos, e naqueles casos en que é pertinente tamén o número. Isto ocorre cando unha palabra posúe un significado léxico distin-

to en singular e plural, ou ben cando só pode ser utilizada en sg. ou en pl.

No caso dos verbos indícase tamén o réxime *transitivo*, *intransitivo* ou *pronominal* que teñen en cada unha das súas acepcións.

Cando, dentro dunha mesma entrada, existe un cambio de categoría, de xénero, de réxime verbal ou ben un uso específico en plural, este cambio márcase por medio de dobre barra.

4.6. Nos casos en que é pertinente, unha entrada ou no seu caso unha acepción concreta desta, leva indicación sobre o nivel de lingua de que é propia, e sobre o campo científico ou técnico a que pertence. Por ex.:

tálamo *s.m.* 1. *lit.* Leito conxugal. *Na escuridade da noite abandonou o tálamo nupcial.* 2. *Bot.* Extremo do pedúnculo onde se asentan as partes da flor. *Os estames e o pistilo están insertos no tálamo.*
SIN. *receptáculo.*

4.7. Ó final de cada acepción dunha entrada indícanse os sinónimos, antónimos e termos moi próximos semanticamente, introducidos respectivamente polas abreviaturas SIN., ANT. e CF. Aínda que estas indicacións van sempre referidas a cada unha das acepcións, sabemos que a sinonimia e antonimia totais se dan en moi poucos casos, e que raramente dous termos son substituíbles sempre, polo que sentímo-la necesidade de especifica-los contextos en que se dá a sinonimia, e para lograr este fin optamos en moitos casos por referi-la información sobre os sinónimos e antónimos a cada exemplo particular, baseándonos nos propios contextos de fala.

Cando un sinónimo, antónimo ou palabra afín valen para todas as acepcións dunha entrada, en vez de repetilos tras cada unha delas, sitúanse ó final precedidos dun rombo.

Unha cuestión que afecta tamén á sinonimia é a do rexistro ou nivel de lingua; por iso, en moitos casos os sinónimos levan indicacións do tipo "culto", "máis culto", etc.

4.8. Nun dicionario debe de existir un certo equilibrio entre o número de entradas e o número de acepcións que se recolle para cada entrada. A dimensión do número de entradas do noso dicionario aconsella prescindir de acepcións arcaizantes, así como daquelas que pertencen a un ámbito demasiado específico ou teñen un uso moi restrinxido.

Evidentemente, así como se prescindiu para a selección de voces daquelas que, aínda tendo certa vixencia na fala, non son propiamente galegas, dado o carácter normativo deste dicionario, tamén na selección das acepcións se prescindiu daquelas que son espurias, debidas principalmente á interferencia co castelán. Así, por ex., na entrada **cubo** non se recolle a acepción de 'caldeiro ou balde', nin na entrada **fallo** se inclúe a acepción de 'sentencia ou veredicto'.

En canto á ordenación das acepcións, a ninguén se lle oculta a dificultade que encerra seguir un criterio dunha maneira sistemática, calquera que este sexa: lóxico, histórico, uso, etc. Xeralmente, nós situamos en primeiro lugar a acepción primaria da palabra, seguida das derivadas. Cando non está claro cuál é a acepción primaria, dásele prioridade á máis usual. Esta regra, non obstante, está na estruturación deste dicionario subordinada á clasificación por categorías gramaticais. Por ex., se a acepción primaria dun verbo corresponde a un uso transitivo, colócanse tras esta tódalas acepcións transitivas, mesmo as figuradas e as menos usuais. Sabemos que é un criterio discutible, coma case todos, pero estamos convencidos de que contribúe a reforza-lo esqueleto formal da obra, e a darlle maior sistematización interna.

Esta é grosso modo a descrición externa da obra que presentamos, pero non quixera deixa-la palabra sen referirme aínda que sexa brevemente ó que constitúe o miolo e a parte máis importante dunha obra lexicográfica, a **definición**.

4.9. Deixando á parte o caso das interxeccións e das palabras gramaticais (preposicións, conxuncións, etc.) procurouse que a definición estivese constituída por unha perífrase que sexa sinónima da palabra definida.

Como puxo de relevo hai xa moitos anos Josette Rey-Debove, resulta evidente que polo menos as palabras gramaticais, non son susceptibles dunha definición en sentido estricto, unha definición formulada en *metalingua de contido*, que responda á pregunta de *qué significa* tal palabra. Por iso nestes casos, é necesario substituí-la definición por unha explicación, que nos diga *qué é esa palabra, cómo e para qué se emprega*.

Fóra destes casos, na definición dos verbos, substantivos, adxectivos e moitos adverbios intentamos aternos sempre na medida do posible ós principios que rexen a lexicografía actual de corte clásico, e fundamentalmente ó principio de substituíbilidade. Unha defi-

nición deberá poder ser empregada no canto da entrada en calquera contexto de fala, sen que varíe o seu significado.

Isto implica, por un lado, que a perífrase da definición deba ter unha forma tal que poida desempeña-la función sintáctica propia da palabra definida; e, por outro lado, a separación na definición, do núcleo e do contorno, tal como hoxe se acepta na maior parte das obras lexicográficas. No noso dicionario o contorno aparece entre corchetes.

No núcleo da definición tratamos de que os semas que inclúe a perífrase sexan suficientes para distingui-la palabra de tódalas demais e evitar ambigüidades, e seguimos unha ordenación gradual, desde o máis xenérico ó máis específico. Comezamos polos trazos comúns a outras unidades léxicas do mesmo campo e seguimos con aqueles trazos necesarios para delimita-la especificidade da unidade definida, tratando de dar cumprimento neste sentido ó principio aristotélico da definición de carácter lóxico baseada na estrutura *xénero próximo + diferencia específica*.

Pero non debemos ocultar que ás veces nalgún caso de palabra con contido léxico, ó descartármola definición mediante un sinónimo, non fomos capaces de subtraernos ó elemento explicativo. Tal ocorre, por exemplo, na acepción 4 da palabra **simple**, que redactamos da seguinte maneira:

4. Precede a certos substantivos para indicar que unha cousa ou persoa é unicamente o que o substantivo expresa, e nada máis ca iso. *É un simple trámite. Crese moita cousa, pero é un simple subordinado.*

En tódalas definicións procuramos facer uso dun nivel de lingua e dun léxico que resulte comprensible para o lector, pois non podemos nunca perder de vista que, á parte doutras utilizacións que pode ter un dicionario de lingua, todas elas moi importantes, hai unha fundamental, que é a de encontra-lo significado dunha voz que o usuario descoñece ou non coñece con precisión. O noso dicionario, seguindo unha tendencia cada vez máis xeneralizada, unicamente utiliza nas definicións voces que figuran como entradas do dicionario; isto, aínda que ás veces dificulta o labor do lexicógrafo, facilita sen dúbida a comprensión do usuario.

Ademais, cada acepción vai ilustrada cun exemplo que facilite a súa comprensión e que oriente sobre o contexto correcto de utilización da palabra.

No último apartado do artigo lexicográfico incluímo-las locucións de todo tipo, as frases feitas e os modismos, por estricta orde alfabética dos seus elementos. Ás veces non é fácil decidir en qué artigo do dicionario hai que incluír estes elementos do mundo da fraseoloxía. Xeralmente optamos por incluílos na entrada da palabra que ó noso xuízo achega unha maior carga semántica ó significado global da expresión. E, ademais, sempre nos fixémo-la pregunta: *se nós fosemo-los usuarios do dicionario ¿baixo qué entrada buscaríamos esta locución?*. Nalgúns casos problemáticos, optamos por face-la inclusión en dúas entradas.

Ben, non quero estenderme máis. Trátase dunha presentación global, e xa que logo superficial, do *Diccionario da lingua galega*, elaborado pola RAG. Teño que dicir que o labor foi arduo, pero tamén que persoalmente tiven a sorte de coordinar un grupo humano extraordinario, ilusionado co traballo que realiza, integrado por Bieito Arias Freixedo, Carme López Taboada, Carme Cambre García, Xosé Luís Martínez Allegue, M^a Dolores Sánchez Palomino, M^a Dolores Sánchez Vales e María Rivas Barrós, que dedicou a última parte da súa vida a esta empresa, e que hoxe por desgracia xa non está fisicamente connosco; pero estou seguro que desde o outro lado da vida tamén se atopa hoxe aquí. A todos eles, gracias.

DICCIONARIO ITALIANO-GALEGO

ISABEL GONZÁLEZ

Universidade de Santiago de Compostela

1. Introducción

A área de Filoloxía Italiana da Universidade de Santiago está a desenvolver, desde hai algún tempo, o proxecto dun dicionario bilingüe italiano-galego.

Dado que, ata o momento, non hai ningunha obra lexicográfica que poña en relación as linguas italiana e galega e os falantes dunha e outra lingua vense obrigados a recorrer a unha terceira, coma o castelán, para se achegar á realidade lingüística italiana ou galega, pensamos que podería ser útil a elaboración dun dicionario bilingüe italiano-galego, que, en principio, estivese dirixido a un único grupo de usuarios.

Debido á complexidade que suporía o dobre enfoque do dicionario (que, no noso caso, estaría orientado a galegos e italianos ó mesmo tempo) consideramos que é máis conveniente dirixir este primeiro dicionario cara ó público galego.

Deste xeito, por unha banda, centrámonos no grupo de usuarios que maioritariamente utilizará esta obra (editarase e distribuirase en Galicia) e, por outra banda, evitamos posibles incoherencias dentro de cada artigo (porque atenderemos a unha única finalidade). Non descartamos que poida ser consultado por italianos, pero, pensando particularmente na estrutura da obra, o enfoque será unívoco.

Con respecto á súa finalidade específica, pretendemos elaborar un dicionario bilingüe destinado á comprensión e tradución de textos italianos contemporáneos.

Este primeiro dicionario está baseado na lingua estándar contemporánea italiana e galega, recollendo ademais, de modo xeral, as diferentes variedades da linguaxe. Terá ó redor de 10.000 entradas, que pretenden recolle-lo léxico básico.

2. Escolla das entradas

Fixémo-la nosa escolla partindo dos dicionarios de frecuencia de uso do italiano, basicamente o realizado por Tullio de Mauro¹, e tamén daquelas obras monolingües e bilingües de extensión semellante á que xa estamos elaborando.

Realizámo-la comprobación da escolla inicial partindo da elaboración de campos semánticos e mesmo incluímos algunhas verbas que se facían necesarias ó observámo-la diferenza de profundidade entre os diferentes campos, coa finalidade de recoller unha variedade de termos o máis ampla posible, segundo unha distribución homoxénea dos temas.

Con respecto ó problema da homonimia, considerando a reducida extensión do noso dicionario, tentamos evitar, na medida do posible, dúas entradas para a mesma palabra, tendo en conta a raíz etimolóxica do termo e a conciencia lingüística dos falantes.

Por outra parte, non consideramos no noso listado de entradas os nomes propios, os termos específicos non moi comúns, as voces correspondentes á linguaxe de tipo administrativo ou literario que sexan pouco frecuentes nos textos actuais, os adverbios rematados en *-mente*, os préstamos lingüísticos de uso restrinxido, prefixos, sufixos e termos relacionados con correntes históricas ou artísticas.

Tendo en conta o criterio da frecuencia, incluímo-los adxectivos rematados en *-abile/-ibile* (*amabile, possibile...*), as palabras compostas (*apricatole* 'abrelatas', *macinacaffè* 'muiño do café'...), a terminoloxía específica de campos semánticos concretos (como a medicina, a música, o dereito, a economía...), as interxeccións (*ahi, oh...*) ou os substantivos en *-mento* (*aggiornamento* 'actualización', *giuramento* 'xuramento', *mantenimento* 'mantemento'...)

Con respecto ós xentilicios, incluímos aqueles que aluden ós continentes, ós países máis coñecidos polo falante medio e mesmo os referidos a rexións e cidades italianas particularmente relevantes (*milanese, piemontese, romano...*).

Incluímo-los participios de uso xeral que teñen un significado adicional ó do verbo, como *abbagliante*, 'luz de estrada', *educato*, 'educado' ou *elevato*, 'elevado'. Noutros casos, non recollémo-lo infinitivo (por ser pouco usado ou arcaico) pero si o participio, como *sperduto* 'perdido, remoto'.

¹ T. de Mauro et alii, *Lessico di frequenza dell'italiano parlato*, Etaslibri, Milano, 1993.

Con respecto ás familias de palabras, recollémo-los termos máis frecuentes.

Atendendo ás necesidades do usuario galego, incluímos como entradas que remiten a outras as variantes ortográficas ou rexionais máis importantes ou ben aquelas voces que poidan ser difíciles de recoñecer por parte do usuario galego. Este é o caso de termos como *anguria* (variante rexional, que remite a *cocomero*, 'sandía'), *comperare* (variante ortográfica, que remite a *comprare*, 'comprar') ou *tuoi* (plural, que remite a *tuo*, 'teu').

3. As fontes

Desde o comezo do traballo, baseámonos fundamentalmente nos dicionarios monolingües Zingarelli², Garzanti³, Devoto-Oli⁴ ... e outros, monolingües e bilingües, segundo unha orde xerárquica de consulta. A elección foi condicionada pola sona e importancia das obras, a fiabilidade e a dispoñibilidade material.

O punto de chegada, é dicir, a nosa base para a escolla das equivalencias é o *Diccionario da Academia Galega*⁵, tendo en conta ademais que, en breve, dispoñemos dunha nova edición mellorada e ampliada. Contamos tamén co *Vocabulario ortográfico da lingua galega*⁶, elaborado polo Instituto da Lingua Galega e a Real Academia Galega, como obra básica de consulta.

Naturalmente, utilizamos outras fontes, non só italianas ou galegas, que nos parecen axeitadas para o desenvolvemento do noso traballo.

4. A macroestructura

A estrutura xeral do dicionario que estamos elaborando consta de tres partes fundamentais:

² N. Zingarelli, *Lo Zingarelli 1995, Vocabolario della lingua italiana*, (eds.) M. Dogliotti e L. Rosiello, Zanichelli, Bologna, (12ª ed.) 1994.

³ AA.VV., *Il grande dizionario Garzanti della lingua italiana*, Garzanti, Milano, 1987 (reimpr. 1988).

⁴ G. Devoto e G. C. Oli, *Il dizionario della lingua italiana*, Felice Le Monnier, Firenze, 1995.

⁵ Real Academia Galega e Instituto da Lingua Galega, *Diccionario da lingua galega*, R.A.G., A Coruña, 1990.

⁶ Instituto da Lingua Galega-Real Academia Galega, *Vocabulario ortográfico da lingua galega*, edic. non venal, Santiago, 1990.

-Introducción, formada polas páxinas titulares, o prólogo, a guía de uso, o listado de abreviaturas e unha pequena síntese gramatical na que recollerémo-las nocións básicas da gramática italiana.

-Corpo do dicionario, ordenado alfabeticamente.

-Os apéndices, que tratarán basicamente os topónimos, os antropónimos e os proverbios italianos máis usuais.

5. A microestructura

En canto á información contida en cada artigo, tivemos en conta a finalidade específica do noso dicionario, para proporcionar-lles ós usuarios galegos os datos estritamente necesarios para comprender ou traducir textos italianos. Ademais, toda a información complementaria está redactada en lingua galega, como veremos máis adiante.

A orde do artigo é a seguinte:

entrada

corpus

caracterización gramatical

acepción

indicacións sobre nivel de lingua e tecnolexos

equivalencia

discriminación de significado

exemplos en italiano e galego

locucións

observacións

5.1. A entrada

Presentámo-lo lema ou entrada sen flexión. É interesante destacar que os substantivos e os adxectivos con variación de xénero van acompañados da desinencia do feminino.

Os plurais irregulares indícanse ó final do artigo, no apartado de observacións, coa finalidade de non saturar-lo usuario con demasiada información, probablemente innecesaria para a comprensión e tradución da entrada, que é a finalidade do dicionario.

lirico -a adx. Lírico -a. Questa edizione raccoglie tutti i componimenti lirici di Leopardi. *Esta edición recolle tódalas composicións líricas de Leopardi.* II s.m. *Lírico.* L'influsso dei lirici classici è presente nella letteratura europea. *O influxo dos líri-*

cos clásicos está presente na literatura europea. OBS.: *m.pl.* lírici, *f.pl.* líriche.

5.2. O corpus

Prescindimos da etimoloxía e da información fonética, pero orientacións relativas a esta última irán recollidas na síntese gramatical.

5.2.1. Caracterización gramatical

Indicámola categoría gramatical por medio de abreviaturas, respectando o sistema gramatical italiano.

En canto ó verbo, sinalamos entre corchetes os auxiliares que corresponden a cada tipo e incluímos, por medio de abreviaturas, a información necesaria para os verbos que sexan irregulares. Esta remite a un apéndice no que se recollen os paradigmas dos verbos regulares e irregulares (facendo referencia a un número determinado).

anteporre *v.tr.* [*aux.* avere] (*fig.*) *Antepoñer, antepor.* Devi anteporre lo studio ai tuoi divertimenti. *Tes que antepoñe-lo estudio ás túas diversións.* OBS.: *v. irreg. V. táboa VII.*

Cando un termo ten dúas ou máis categorías gramaticais, estas pódense ordenar mediante un criterio de frecuencia de uso e un criterio lóxico. A categoría gramatical refírese á palabra da entrada, polo que as variacións de categoría ou de xénero no equivalente galego (estas últimas bastante comúns) non se especifican. O exemplo ten que reflectir con claridade tal variación.

lavagna *s.f.* *Encerado.* Sono allergica al gesso e perciò non posso scrivere alla lavagna. *Son alérxica ó xiz e por iso non podo escribir no encerado.*

5.2.2. As acepcións

A orde das acepcións dentro de cada categoría gramatical baséase nos criterios lóxico e de frecuencia de uso e a súa numeración vén indicada por números arábigos.

Recollémo-las acepcións básicas (omitindo tecnicismos non comúns, usos dialectais, arcaísmos, acepcións literarias...) para ser coherentes co obxectivo da obra.

A acepción consta de catro partes fundamentais, que, segundo as necesidades de cada entrada, poden non aparecer en tódolos artigos. Son as seguintes: indicacións sobre o nivel de lingua e

tecnolectos, a equivalencia, a discriminación de significado ou a extensión de uso e os exemplos.

5.2.2.1. Indicacións sobre o nivel de lingua e tecnolectos

Indicamos esta información entre parénteses e abreviada. A súa utilidade é a de proporcionarlle ó usuario unha maior información sobre cada acepción e, ó mesmo tempo, diferenciala doutras, incluíndoa nun campo semántico ou indicando a súa aplicación mediante as etiquetas de uso.

alibi s.m.inv. 1 (Der.) Coartada. Non hanno potuto accusarlo del reato perché aveva un alibi. *Non puideron acusalo do delicto porque tiña unha coartada. 2 (fig.) Pretexto, excusa.* I tuoi alibi non convincono nessuno. *Os teus pretextos non convencen a ninguén.*

5.2.2.2. A equivalencia

Practicamente, non existen equivalentes absolutos, debemos falar, máis ben, de equivalentes parciais e, con frecuencia, non podemos recorrer nin sequera a estes, no caso das chamadas *palabras culturais* (que non atopan referentes na lingua de chegada).

Para dar solución ás palabras culturais, consideramos que convén proporcionarlle ó usuario unha explicación ou unha equivalencia próxima ó termo da lingua de partida. Se é posible as dúas cousas. Cando se perden parcelas de significado na equivalencia facilitada logo da explicación, antepoñemos entre parénteses a abreviatura *aprox.*, para advertirlle ó lector que o termo é parcialmente equivalente respecto da entrada.

ricotta s.f. Tipo de queixo italiano feito co soro cocido do leite de ovella.

omertà s.f. Acordo tácito entre os membros do crime organizado para encubrir un delicto. (aprox.) Encubrimiento do delicto. L'omertà ha impedito che fossero concluse le indagini. *O encubrimiento do delicto impediú que rematasen as investigacións.*

Na medida do posible, procuramos dar máis dunha equivalencia para cada acepción, de xeito que evite ó usuario o ter que acudir frecuentemente a un diccionario de sinónimos.

As observacións gramaticais referidas ó equivalente galego non nos parecen necesarias, dado que partimos da base de que os usuarios galegos teñen suficientes coñecementos sobre a súa propia lingua.

Por exemplo, aínda que o feminino *egua*, baixo a entrada *cavalo*, apareza indicado como equivalencia, non é preciso especificar que se trata dun feminino irregular. Por motivos de coherencia coa entrada, recollémo-las formas masculina e feminina, mostrando, ó mesmo tempo, que en italiano non existe un feminino irregular similar a *egua*.

cavallo -a s. 1 *Cabalo egua*. Il cavallo nitrisce. O *cabalo rincha*. La cavalla ha partorito ieri sera. *A egua pariu onte á noite*. 2 *Cabalo (do xogo)*. Si è mangiato il cavallo con l'alfiere. *Comeu o cabalo co alfil*. Ho perso il fante con il cavallo. *Perdín a sota co cabalo*. 3 *(Mec.) Cabalo*. Quanti cavalli ha la tua macchina. *¿Cantos cabalos ten o teu coche?* 4 *(Dep.) Cabalo*. A lezione di ginnastica dobbiamo saltare il cavallo. *Na clase de ximnasia temos que salta-lo cabalo*. & **a ferro di cavallo**: *con forma de ferradura*; **cavallo di battaglia**: *especialidade*; **essere a cavallo**: *estar salvado*; **febbre da cavallo**: *febre de cabalo*.

5.2.2.3. A discriminación de significado

Trátase dunha información bastante útil para o usuario, xa que axuda a delimita-lo significado da entrada en cada unha das súas acepcións, respecto das demais e mesmo con respecto á extensión semántica da equivalencia galega.

Esta parte da acepción non se limita unicamente a proporcionar unha información case definitoria, senón que pode indicar tamén o contexto da equivalencia, complementándoa.

germe 1 s.m. *Xerme (microorganismo)*. Devi disinfettare bene la ferita per eliminare tutti i germi patogeni. *Tes que desinfectar ben a ferida para eliminar tódolos xermes patóxenos*. 2 *(fig.) Xerme, xermolo*. Quella forte discussione è stata il germe della discordia fra di loro. *Aquela forte discusión foi o xerme da discordia entre eles*.

lisca s.f. *Espiña (do peixe)*. Il gatto si è mangiato le lische che erano sul piattino. *O gato comeu as espiñas que estaban no pratoño*. OBS.: *pl. lische*.

5.2.2.4. Os exemplos

Tentamos axustarnos ó que considerámo-lo exemplo ideal: breve, ilustrativo e diferenciador respecto doutras acepcións. En todo caso, non amplía o significado do lema italiano.

A inclusión dos exemplos tanto en italiano coma en galego párecenos moi útil. Somos conscientes de que aumenta considerablemente a dimensión do corpo do dicionario, pero párecenos aconsellable para facilita-la asimilación da lingua italiana, a través da comparación das estruturas sintácticas de ámbalas dúas linguas.

Evitamos dar exemplos das palabras culturais que carezan de equivalencia (como viamos no artigo de *ricotta*, xa citado). Cando, pola contra, nos aventuramos a dar un equivalente aproximado, presentámo-lo exemplo, coa finalidade de mostra-la aplicación do termo nun contexto (tal era o caso da entrada *omertà*, vista anteriormente).

Tódolos exemplos proceden da lingua viva e procuramos ofrecerlle ó usuario as diferentes particularidades do italiano, contrastado ó mesmo tempo co galego.

5.2.3. Locucións

Baixo este título recollemos toda indicación ás locucións, as frases feitas, os modismos... Todas estas foron seleccionadas atendendo unicamente ó criterio da frecuencia de uso e están ordenadas no artigo por orde alfabética. Este apartado vai situado xusto despois das acepcións.

Respecto da equivalencia galega, intentamos proporcionar unha locución, frase feita ou un modismo similares. Cando non é posible, presentamos un termo que recolla o significado da expresión ou, en último caso, ofrecemos unha explicación.

non fare niente: *non dar pancada*

fare il muso: *torce-lo fociño*

fare presto: *apurar*

far vedere bianco per nero: *enganar*

scendere a patti: *verse obrigado a chegar a un acordo*

scendere in piazza: *participar nunha manifestación*

5.2.4. Observacións

O apartado de observacións recolle informacións de distinto tipo e sitúase tanto ó final de cada acepción (no caso de que se refira só a esta), coma ó final do artigo, tralas locucións.

Como dicíamos, o contido das observacións é variado. Pode recolle-los plurais irregulares, as variacións de acento, a primeira persoa do presente do verbo (para mostra-la caída do acento ou ben unha pequena peculiaridade da súa conxugación), a referencia a unha

táboa de verbos, unha información aclaratoria (no caso de que haxa falsos amigos), a delimitación semántica dalgunha acepción ou mesmo da entrada (cando non fora posible conter toda a información necesaria en equivalencias, explicacións ou exemplos...). A continuación mostramos algúns breves exemplos do uso deste apartado.

analfabeta OBS.: *o xénero deste termo é invariable só no singular; m.pl. analfabeti, f.pl. analfabete.*

giallo -a II s.m. 2 *Novela policial.* Agatha Christie scrisse tanti gialli. *Agatha Christie escribiu moitas novelas policiais.* OBS.: *tamén é aplicable ós filmes.*

ginocchio OBS.: *pl. ginocchi e tamén ginocchia, referíndose este último ós dous xeonllos dunha persoa.*

Se o noso dicionario tivese como finalidade a produción de textos, toda esta información debería ser máis ampla. Xa que o propósito da obra é axudar a comprender ou traducir textos italianos, consideramos que o apartado de observacións pode completa-la información básica que precisa o usuario. Deste xeito, o lector pode prescindir da consulta deste apartado, se o desexa, ou amplia-los seus coñecementos sobre a entrada.

A redacción dos artigos estanos resultando difícil -a ningún se lle escapa a cantidade de problemas que se presentan en calquera traballo de lexicografía- pero coa dedicación e ilusión que dende o primeiro momento (levamos algo máis de dous anos traballando neste proxecto) puxeron tódolos autores, para min, está resultando un pracer e tamén unha honra, dirixir a este grupo de traballo formado por Carmen Blanco Valdés, Benedict Buono, Asunción Caneda Cabrera, Consuelo de Frutos Martínez, María Díaz Bernárdez, Ana Domínguez Ferro, Javier Gutierrez Carou, Vincenzo Jacomuzzi, Cristina Marchisio, María Montes López, Alessandra Pavanello, Carmen Prados Anaya e María Teresa Sanmarco Bande, todos eles, Licenciados, Doutores e Profesores de Filoloxía Italiana e Galega. Moitas grazas.

O DICCIONARIO CASTELÁN-GALEGO DA REAL ACADEMIA GALEGA

ISABEL GONZÁLEZ
Universidade de Santiago de Compostela

1. Introdución

O dicionario castelán-galego que nós familiarmente chamamos "pequeno" (arredor das 12.000 entradas) está case no prelo. Para ser exactos, estámolle dando as últimas correccións, tratando de mellora-la coherencia interna da obra. Estase facendo a revisión final do "pequeno", pero xa hai algún tempo que se comezou a tarefa de selección de entradas dun segundo -agora falamos do "grande"- dicionario castelán-galego, que incluírá ó redor de 30.000 entradas.

A continuación exporemos brevemente as características do xa feito.

Un dicionario bilingüe sempre está enfocado cara a unha das linguas tratadas (ou, no caso dos bilingües dobres, cada metade cara a unha lingua). Nós partimos dun suposto usuario coñecedor do castelán, que busca no dicionario a correspondencia dun termo en galego. Así, tódalas informacións estarán redactadas en castelán e trataremos de estar atentos ós problemas e ás dúbidas que pode atopar en galego, incidindo especialmente, claro está, nos aspectos diferenciais.

2. Selección de entradas

A pesar do limitado número de entradas, tratamos de que non quedaran fóra aquelas máis correntes e tentamos seleccionalas tendo en conta a súa presenza na fala. Botamos man de estudos de frecuencias feitos sobre o castelán, cotexándoos con dicionarios bilingües do castelán con outras linguas, dicionarios de castelán de tipo básico e medio, etc.

Pero nun dicionario bilingüe non se pode atender unicamente á lingua de partida, senón que debemos ter presente sempre tamén a de

chegada, e así démoslles cabida a formas pouco frecuentes en castelán pero moi frecuentes en galego, xa que poden ser útiles para atopar palabras moi usuais en galego, como por exemplo a voz *grelo*.

3. Microestructura

Os dicionarios bilingües actuais non lle dan ó lector información léxica unicamente, senón que lle proporcionan igualmente unha serie de datos encamiñados á utilización correcta da lingua que o usuario non domina ou non domina perfectamente. Nós quixemos seguir tamén esta tendencia, e para iso incluimos información sobre a categoría gramatical, a sintaxe, etc., así como exemplos.

En canto ó uso da tipografía, utilizamos distintos tipos de letra para cada idioma: para o castelán letra redonda e para o galego letra cursiva.

3.1. A entrada aparece en letra grosa redonda, de maior tamaño có resto do artigo. Para a distinción dos homónimos seguímo-lo criterio utilizado polo DRAE, e aparecen no seu caso sinalados con superíndices.

3.2. A organización do artigo vén determinada polas posibilidades de traducción, e se unha mesma traducción é común para varios usos da palabra da entrada, agrúpanse. Así, se a(s) posibilidade(s) de traducción dunha palabra son sempre a(s) mesma(s), aparecerán reunidas, aínda que se correspondan con distintas categorías que poida te-la entrada.

agente *adj., s.m. y s. axente. Adj.:* Escribe una oración pasiva con complemento agente. *Escribe unha oración pasiva con complemento axente. /S.m.:* La lluvia es un agente atmosférico. *A chuvia é un axente atmosférico. / S.:* María trabajaba como agente en una compañía de seguros. *María traballaba como axente nunha compañía de seguros.* Los paró un agente y les pidió la documentación. *Parounos un axente e pediulle-la documentación.*

Se existen distintas posibilidades de traducción presentaranse organizadas do seguinte xeito:

1. De agrúpárense varias traducións comúns, estas aparecen en primeiro lugar, aínda que se correspondan con distintas categorías. No caso de adxectivos ou substantivos que teñan forma masculina e femi-

nina, só aparecerá o masculino, e no caso de substantivos que se utilizan en singular e plural só o singular, xa que as demais formas serían facilmente deducibles a partir destas e xa se recollen posteriormente nos exemplos.

2. As posibilidades de traducción que son distintas presentaranse agrupadas por categorías, en parágrafos á parte, sinalados coas letras (a), (b), (c), ... , en caracteres grosos e entre parénteses. Dentro de cada parágrafo, ou categoría, sinálanse con números: 1., 2., 3., etc.

contar (a) *v.t. e i. (común) contar. V.t.:* Cuenta las personas que hay presentes. *Conta as persoas que hai presentes.* Cuéntame que tal te va. *Cóntame que tal che vai.* Entre los gastos del viaje cuentan las comidas. *Entre os gastos da viaxe contan as comidas.* / *V.i.:* Cuenta hasta cinco y échate a correr. *Conta ata cinco e bótate a correr.*

(b) *v.t. 1. (un suceso o historia) contar, relatar, narrar.* Nos contó el caso con todos los detalles. *Contounos ~ relatounos ~ narróuno-lo caso con tódolos detalles.* 2. (una edad, alguien o algo) *contar, ter.* Cuando emigró contaba veinticinco años. *Cando emigrou contaba ~ tiña vintecinco anos.*

(c) *v.t. y p. (alguien o algo en un grupo o categoría) contar(se), incluir (t.). V.t.:* Te cuento entre mis mejores amigos. *Cóntote ~ inclúote entre os meus mellores amigos.* / *V.p.:* Este vino se cuenta entre los más selectos. *Este viño cóntase entre os máis selectos.*

(d) *v.i. (algo, por su importancia) contar, importar.* Para él lo que cuenta son los sentimientos. *Para el o que conta ~ importa son os sentimentos.*

A traducción que aparece en primeiro lugar é a que agrupa varios usos da entrada, ou, se non hai unha traducción común, a que se corresponde coa acepción máis usual de palabra castelá. A continuación, dentro do mesmo parágrafo, seguen as que pertencen á mesma categoría, aínda que outras traduccions incluídas en seguintes parágrafos se poidan considerar máis usuais. Nisto seguímo-lo mesmo criterio có *Diccionario da Lingua Galega* que acabamos de ver.

coger (a) *v.t. 1. (común) coller. Coge* el vaso con las dos manos. *Colle o vaso coas dúas mans.* Cógelo de la mano para cruzar. *Cólleo da man para cruzar.* Si llegas pronto cógeme sitio. *Se chegas pronto cólleme sitio.* Sin antena no cogerás la emisora. *Sen antena non collera-la emisora.* 2. (lo que se cosecha) *coller, apañar, recoller.* En otoño cogemos las castañas. *No outono collemos ~ apañamos ~ recollémo-las castañas.* 3. (a alguien, de cierta forma) *co-*

ller, pillar. Lo cogimos de buen humor. *Collémolo ~ pillámolo de bo humor*. 4. (a alguien que escapa o está desprevenido) *coller, pillar, cachar*. Lo cogieron copiando. *Collérono ~ pillárono ~ cachárono copiando*. 5. (una enfermedad o un vicio) *coller, pillar, agarrar*. Has cogido una buena gripe. *Colliches ~ pillaches ~ agarraches unha boa gripe*. 6. (a alguien, un vehículo) *coller, pillar, atopelar*. No cruces sin mirar que te puede coger un coche. *Non cruces sen mirar que te pode coller ~ pillar ~ atopelar un coche*. 7. (el significado de algo) *coller, cachar*. No cogí la indirecta. *Non collín ~ cachei a indirecta*.

(b) *v.i.* (fam., una persona o una cosa en un lugar) *coller, caber, entrar*. No cogemos tantos en un solo coche. *Non collemos ~ cabemos ~ entramos tantos nun só coche*.

3. Cada división dentro do artigo levará sempre unha explicación que estableza as diferencias entre traducións que non se poden utilizar indistintamente. Isto parécenos fundamental, xa que en moitos bilingües ofréceselle ó usuario un abano de posibilidades en absoluto diferenciadas ou insuficientemente diferenciadas que poden inducir ó lector a incorreccións, ó usalas indistintamente cando isto realmente non é posible.

Así, despois de cada número, ou da indicación da categoría, se dentro dela non hai divisións, aparece entre parénteses unha indicación relativa ós usos ou contextos nos que se poden utiliza-las traducións dadas. Estas indicacións serán distintas segundo o tipo de palabra de que se trate; seguimos en principio as propostas por Colin Smith¹: para o substantivo, sinónimos aproximativos ou definicións parciais; para o adxectivo, substantivos cos que poida concordar; para o verbo transitivo, substantivos que poidan se-los seus complementos directos e para o verbo intransitivo e pronominal, substantivos que poidan se-los seus suxeitos. O mesmo autor recoñece que só se deben seguir cando resulten útiles, recorrendo a outras en caso contrario, e foi tamén o que fixemos nós cando nos pareceu máis axeitado para da-la equivalencia máis exacta posible.

Como xa expuxemos anteriormente, o dicionario está enfocado cara ós falantes de castelán, por iso estas indicacións están redactadas en castelán.

adormecer (a) *v.t.* 1. (a alguien, produciéndole sueño) *adormentar, adurmiñar*. El ruido del mar lo adormeció. *O ruído do mar*

¹. Colin Smith, *Diccionario Español-Inglés English-Spanish*, Grijalbo, Barcelona, Introducción, p. xviii.

adormentouno ~ adurmiñouno. 2. (algo, haciéndolo menos intenso) *adormentar, clamar* ≡ *acalmar, mitigar*. Con estas pastillas consiguió adormecer el dolor. *Con estas pastillas conseguíu adormentar ~ calmar* ≡ *acalmar ~ mitiga-la dor*.

(b) *v.p.* 1. (alguien, comenzando a dormir) *adormecer (i)*. No me adormecí hasta muy tarde. *Non adormecín ata moi tarde*. 2. (fig., una parte del cuerpo) *adormecer (i), entumecerse*. Se le adormeció una pierna. *Adormeceulle unha perna. Entume-céuselle unha perna*.

brujo -a (a) s. (persona con poderes) *bruxo, meigo*. Una bruja le echó las cartas. *Unha bruxa ~ meiga botoulle as cartas*.

(b) *adj.* (que atrae fuertemente) *meigo, feiticeiro*. No se podía resistir a su encanto brujo. *Non se podía resistir ó seu encanto bruxo ~ feiticeiro*.

4. Depois da categoría ou de cada explicación entre parénteses, cando as hai, figura(n) a(s) traducción(s), en caracteres grosos e cursiva. Cando hai máis dunha palabra aparecen separadas por comas, comezando pola máis próxima á entrada, sempre que sexa tamén usual en galego. De ser unha forma pouco frecuente aparecería en primeiro lugar a palabra considerada máis habitual. A continuación as restantes posibilidades, tratando de ordenalas segundo a súa presenza na fala, e evitando aquelas que, aínda sendo perfectamente válidas, só se utilizan nunha área moi restrinxida do territorio. Se entre as posibles traduccions existen dúas ou máis que son variantes da mesma forma, aparecen separadas entre si polo símbolo ≡.

habituat *v.t. y p.* *habituat(se), afacer(se) ≡ facer(se), acostumar(se)*. *V.t.*: Hay que habituat al niño a comer de todo. *Hai que habituatar ~ afacer ≡ facer ~ acostuma-lo neno a comer de todo.* / *V.p.*: Se habituó a la vida fácil. *Habituouse ~ afixose ≡ fixose ~ acostumouse á vida fácil*.

5. Se, con respecto ó indicado para a entrada, a palabra dada como traducción presenta algunha variación de xénero, número ou réxime verbal, indícase entre parénteses e con letra cursiva despois de cada traducción.

miel *s.f.* *mel (m.)*

olvidar *v.t. y p.* *esquecer (t. e i.)*

6. Se á voz que se dá como equivalente en galego lle corresponde un feminino ou un plural irregulares indícase a continuación e entre parénteses en letra grosa cursiva, precedidos da correspondente abreviatura en cursiva.

aquel, aquela *pron. dem. aquel (pl. aqueles)*
ladrón -a *adj. y s. ladrón (f. ladrona, ladroa o ladra)*

7. A continuación sitúanse os exemplos. En primeiro lugar o exemplo en castelán en letra redonda e logo o exemplo en galego en letra cursiva. Se se dera máis de unha posibilidade de tradución, estas aparecen no exemplo separadas polo símbolo ~, ou, no caso de seren variantes, polo símbolo ≡. Se unha forma verbal vai seguida pola segunda forma do artigo, só se engadirá ó verbo situado en último lugar. Se algunha destas palabras variase en xénero ou en número con respecto á(s) outra(s), repetírase tamén aquela parte do exemplo que non concorde. No caso de verbos que tivesen uso transitivo, intransitivo ou pronominal diferente, optouse na maioría dos casos por repeti-lo exemplo enteiro para que a construción resultase máis clara.

enredo *s.m.* 1. (de hilos o cosas semejantes) **enredo, maraña** (*f.*). No pudo deshacer el enredo y cortó el hilo. *Non puido desface-lo enredo ~ a maraña e cortou o fio.* 2. (asunto complicado) **enredo, maraña** (*f.*), **zarangallada** (*f.*) ≡ **zarapallada** (*f.*). Se metió en un enredo que lo arruinó. *Meteuse nun enredo ~ nunha maraña ~ zarangallada ≡ zarapallada que o arruinou.*

apartar *v.t. y p.* 3. (alguien o algo de donde estaba) **apartar(se)**, (*t. i. y p.*), **quitar(se)** (*t. i. y p.*). *V.t.*: Aparta las cajas de la entrada. *Aparta ~ quita as caixas da entrada.* / *V.p.*: Apártate de la ventana que te puedes caer. *Apártate ~ quítate da ventá, que podes caer.* *Aparta ~ quita da ventá, que podes caer.*

Ás veces inclúense varios exemplos, ben porque se agrupen distintos usos da entrada ou ben porque se agrupen distintas categorías. Neste último caso, antes de cada exemplo ou exemplos sinálase coa abreviatura que corresponda. A distinción entre os exemplos de categorías distintas márcase cunha barra (/).

lila *adj., s.m. y s.f.* **lila**. *Adj.*: Se puso de moda el color lila. *Púxose de moda a cor lila.* / *S.m.*: El lila y el blanco combinan bien. *O lila e o branco combinan ben.* / *S.f.*: Le regaló un ramo de lilas. *Regaloulle un ramo de lilas.*

8. Como complemento á información dada no artigo incluíranse nas observacións aspectos que non se traten de forma sistemática dentro da microestrutura, pero que se consideren útiles para un caso concreto. Normalmente refírense a cuestións gramaticais, aínda que tamén poden ser de tipo semántico.

Situaranse, precedidas pola abreviatura OBS. en letras versais, en parágrafo á parte se se refiren á entrada en xeral ou, no caso contrario, detrás da división correspondente. Da mesma maneira que as explicacións entre parénteses, tamén estarán redactadas en castelán.

alguno -a *adj. y pron. indef. [...]*

OBS. En gallego, en frases negativas, sólo se pueden utilizar las formas *ningún, ningunha*: No presentó oposición alguna. *Non presentou oposición ningunha*.

cesta *s.m. cesta, cesto (m.) [...]* OBS. Según las zonas, en gallego *cesto* o *cesta* pueden designar el de mayor o menor tamaño.

9. Detrás dunha dobre barra (//) inclúense locucións ou expresións fixas, e se a palabra ten varias categorías aparecerán as locucións en parágrafo á parte. Atopa-la equivalencia para as frases feitas non é labor doado e tampouco se fixo nesta obra unha recompilación exhaustiva das expresións fixadas en castelán.

Decidir baixo qué entrada deben figurar é sempre problemático e decidimos segui-los criterios expostos polo DRAE. A palabra que se corresponde coa entrada aparece indicada só coa primeira letra sempre que coincidan exactamente ou sexa unha forma flexiva facilmente deducible pola concordancia.

A frase en castelán figura en letra grosa redonda e, separada(s) por dous puntos a(s) súa(s) traducción(s) en galego, en letra grosa cursiva. A continuación inclúense os exemplos correspondentes en castelán e galego, ó igual ca no resto dos casos. De existiren varias posibilidades parcialmente coincidentes, o fragmento común non se repetirá no exemplo.

caballo *s.m. caballo [...]* // A c.: *a caballo, dacabalo*. Iba a caballo de una yegua blanca. *Ía a caballo ~ dacabalo dunha egua branca*.

10. Á parte, co símbolo ♦ introducíranse os casos nos que o verbo rexa preposición. Se a forma verbal coincide totalmente coa entrada escribírase de maneira abreviada. A construción en castelán apa-

rece en letra grosa redonda e, separada(s) por dous puntos, a(s) equivalencia(s) en galego. Como nos casos anteriores, ilustraranse con exemplos e non se repetirá a preposición se é a mesma para máis dunha posibilidade de traducción.

hacer [...]

- ♦ h. de: *facer de*. Hizo de protagonista. *Fixo de protagonista*.
- ♦ h. por: *facer por, tratar de, intentar, tentar*. Haz por llegar a tiempo. *Fai por ~ trata de ~ intenta ~ tenta chegar a tempo*.
- ♦ h. con: *facerse con*. Se hizo con el poder. *Fíxose co poder*.

11. Consideramos de grande utilidade incluír datos relativos á gramática, aínda que só se tratará de cuestións básicas. Algunhas destas informacións xa van incluídas na estrutura do artigo, como vimos nos parágrafos 5. e 6.; outras poden aparecer nas observacións (contraccións, emprego da segunda forma do artigo, etc.); pero, para evitar unha complicación excesiva, outros aspectos gramaticais recóllense nun apéndice apartado do corpo do dicionario, enfocado de xeito comparativo entre as dúas linguas tratadas.

A nosa aspiración con este dicionario bilingüe castelán-galego é proporcionarlle ó lector a correspondencia máis exacta posible entre a entrada e a traducción e, dentro disto, o maior número de posibilidades existentes, e con este obxectivo fómoslles dando solucións a tantas dificultades que atopamos. Así, incluímos unha explicación que oriente ó usuario cando as posibilidades de traducción non se poden utilizar indistintamente, os cambios de xénero, número ou réxime verbal na lingua de chegada con relación á entrada, observacións, etc., e, como xa vimos, exemplos que poidan servir de ilustración e de modelo. E ó mesmo tempo non podemos esquecer a importancia que ten a facilidade no manexo do dicionario.

Moitos foron os obxectivos e moita a nosa vontade de conseguilos, pero somos conscientes da dificultade da tarefa e, se cadra, non alcanzamos máis ca unha mínima parte. Isto xa se saberá cando o público o xulgue. Das súas críticas aprenderemos para acometer novos e máis ambiciosos proxectos.

Non quero rematar sen salientar dun xeito especial o labor desenvolvido polo grupo de traballo que tiveron a sorte e a honra de coordinar desde hai algo máis de dous anos. O grupo, formado por Xesús Domínguez Dono, Ana Fachal Fraguela, M^a Xesús Gómez Sanjurjo, Pablo Lado López, M^a do Carme Pazos Balado, Manuela Pérez Rodríguez, M^a do Carme Prieto Guibelalde e Isolina Salgado Basoa, traba-

llou arreo, cunha grande ilusión e vontade a pesar de ser unha tarefa moi difícil e ardua. Moitas gracias a todos eles.

LEXICOGRAFÍA VASCA, EL DICCIONARIO GENERAL VASCO

ENCUESTA
Vascongada

El tiempo que necesitó para recopilar algunas respuestas referentes a la lengua vasca y su historia, debía ser de que le correspondiera postular no trabajo que sería completamente independiente por pertenecer a otros países.

La lengua vasca se encuentra fragmentada en una serie de dialectos, algunos de ellos de marcada personalidad. Hay en ella el castellano que los dialectos vascos son desde el vascón, el guipuzcoano, el alto-navarro y el vascallés se extiende por el pirineo y el alto-navarro, el basco-navarro y el euzkoano en el denominado País Vasco-Francés. Como es sabido, una de las peculiaridades de la lengua vasca es su carácter regional. En su pueblo, algo más reciente la etimología que se maneja sobre la evolución de la lengua vasca, de parte de dialectos se haya impreso en el uso común. Por otra parte, el país vascongado se divide en varias zonas, en el centro el vascón y el guipuzcoano se reparten, con algunas variantes, cerca del 50% de los habitantes vascos, y en el norte navarro, el vascallés es el de los que son como dialecto vascos.

Volviendo al vocablo vasco, como son los dialectos que han merecido la clasificación de la lengua vasca, primer lugar el vascón, el dialecto de la zona centro vascongada, es decir del País de Lehera, dialecto en el que se escribió nuestra literatura clásica, desde un momento en el siglo XVII, cuando se originó el guipuzcoano, cubriendo desde un tiempo del siglo XVIII y el dialecto que abarca el rango de dialecto navarro y vascón del XIX. Mención que merece el dialecto, dialecto de la zona más oriental del País Vasco continental, que puede ser considerado dialecto vascongado desde tal momento por su uso.

Una cuestión que merece llamar la atención es relativa a una producción del total de libros publicados en el período 1950-

attractiva que se mostra a través de alguns dos seus usos, pois
levamos ao grupo. Como resultado, a análise mostra algumas diferenças
para a sua utilização e importância de los contextos para cada classe po-
sibilitada de transição.

Para o (a)

- * A. de forma de. Usos de progressiva. A. de transição
- * B. para descrever, contar, explicar, avaliar. Usos de forma +
tempo. Ex: *Estava a estudar* - *estava a estudar* - *estava a estudar* + tempo.
- * A. para fazer um plano ou a parte. Ex: *Estava a fazer*.

* B. Consideramos de grande utilidade todos estes verbos e a
possibilidade de se usarem de maneiras distintas, algumas delas
relacionadas com a utilização de estruturas de frases, como o uso de
partículas A e B. Outros pontos que merecem ser observados concernem
além, sobretudo, ao modo de usar de frases, ou seja, para criar
uma complexidade gramatical, ou para expressar particularidades que
aparecem apenas de certos contextos, sobretudo de certo
contexto onde se dá lugar a estas estruturas.

A parte seguinte dos nossos documentos bilíngues contém o glossário
proprietário de termos e expressões, com uma lista de palavras, uma
lista de estruturas e, dentro disso, a análise de cada uma das estruturas
relacionadas com os usos observados. Também há uma lista de
palavras que aparecem. Há, inclusive, uma explicação que, além de
mostrar como se utilizam as estruturas, há de se ter em conta as
particularidades, os contextos de uso, a relação de cada verbo no lin-
guagem de origem com relação à estrutura, observando, por exemplo, os
verbos, exemplos que podem servir de ilustração e de apoio. Há
também alguns verbos que podem ser usados de maneiras diferentes, de
modo a ser possível separar a importância que tem a utilização
de um verbo, no contexto.

Muito foram os observados e analisados, com o intuito de conside-
rar, para cada um dos verbos, as particularidades de cada um, de modo a
mostrar como se utilizam para isso, e se há ou não a possibilidade
de serem usados de maneiras diferentes para os contextos, e não
simplesmente para o.

Nota: Este trabalho foi elaborado com o auxílio de alguns de-
senvolvidores, para quem se agradece que teve a oportunidade de con-
tato com os alunos, com o apoio de alguns grupos formados por Nilda
Domingos, Rosa, Ana Paula, Priscila, M.ª Xosé, Glória, Rosângela,
Tânia, Lúcia, Lígia, M.ª da Glória, Paula, Raquel, Estéfania, Nilda, Ros-
ângela, M.ª da Glória, Priscila, Glória e todos os alunos da turma, entre

LEXICOGRAFÍA VASCA: EL DICCIONARIO GENERAL VASCO

IBON SARASOLA
Euskaltzaindia

1. Tengo que comenzar por explicar algunos aspectos referentes a la lengua vasca y su historia, única forma de que la exposición posterior no tenga que verse constantemente interrumpida por paréntesis aclaratorios.

La lengua vasca se encuentra fragmentada en una serie de dialectos, algunos de ellos de marcada personalidad. Hoy en día se considera que los dialectos vascos son siete: el vizcaíno, el guipuzcoano, el alto-navarro y el roncalés en territorio peninsular, y el labortano, el bajo-navarro y el suletino en el denominado País Vasco-Francés. Como es sabido, esta diferenciación dialectal más bien fuerte no es un hecho excepcional. Sí es, quizás, algo más llamativa la circunstancia que en cinco siglos de cultivo literario ninguno de estos dialectos se haya impuesto en el uso escrito. Por otra parte, el peso demográfico de estos dialectos no es el mismo: el vizcaíno y el guipuzcoano se reparten, casi a partes iguales, cerca del 80% de los hablantes actuales, y en el otro extremo, el roncalés acaba de desaparecer como dialecto hablado.

Volviendo al cultivo escrito, cuatro son los dialectos que han merecido la consideración de literarios: en primer lugar el labortano, el dialecto de la zona costera vasco-francesa, es decir del País de Labort, dialecto en el que se escribió nuestra literatura clásica, fundamentalmente en el siglo XVII; vienen a continuación el guipuzcoano, cultivado desde mediados del siglo XVIII y el vizcaíno que adquiere rango de dialecto literario a comienzos del XIX. Mención aparte merece el suletino, dialecto de la zona más oriental del País Vasco continental, que puede ser considerado dialecto marginal desde casi todos los puntos de vista.

Unas cuantas cifras pueden ilustrar el peso absoluto y relativo de esta producción. El total de libros publicados en el período 1545-

1880 asciende a 600, que se reduce a 200 si no tenemos en cuenta las reediciones, a 100 si descontamos también las traducciones, y a una docena si sólo tenemos en cuenta las obras no religiosas. Esta producción se fue incrementando paulatinamente hasta los años de la segunda República, pero sin llegar nunca a cotas espectaculares. En la actualidad la producción parece que se ha estancado alrededor de los mil títulos anuales. Por lo que se refiere a la distribución por dialectos, la correlación entre número de hablantes y peso literario es muy baja: las hablas alto-navarras, de gran peso demográfico en siglos anteriores, no han sido prácticamente cultivadas literariamente y el mismo vizcaíno, que como hemos señalado supone hoy en día el 40% de los hablantes, no llega al 15% de la producción editorial del período antiguo antes citado. Históricamente son el guipuzcoano, dialecto central del País Vasco peninsular, y el labortano, de gran prestigio literario y que ha servido como norma escrita a los autores de habla bajo-navarra, los dos dialectos literarios principales.

2. Se pueden distinguir tres períodos en la lexicografía vasca: el anterior a la publicación del Diccionario de Larramendi (1745), el que media entre éste y el diccionario de Azkue (1905-6), y el del presente siglo. La época anterior al diccionario de Larramendi se caracteriza por la práctica ausencia de diccionarios publicados. Así, las tres principales obras de la época, el vocabulario español-vasco recopilado por el italiano Landuchio (1562), el gran diccionario vasco-francés (con referencias en otros idiomas) de Silvain Pouvreau y el diccionario latín-vasco, sólo redactado hasta la letra C, de Pierre d'Urte, han permanecido inéditos hasta fechas recientes, excepción hecha del de Pouvreau que aún hoy día sólo puede consultarse en microfilm.

La primera obra lexicográfica vasca de envergadura que ve la imprenta es el *Diccionario Trilingüe del Castellano, Bascuence y Latín* del jesuita guipuzcoano Manuel de Larramendi, que se publicó, como ya hemos señalado, en 1745. Esta obra de Larramendi fue publicada tras una apología de la lengua (*De la antigüedad y universalidad del Bascuence en España: de sus perfecciones y ventajas sobre otra muchas lenguas*, 1728), y una gramática, que lleva el título revelador de *El imposible vencido* (1729). No hace falta decir más para comprender que la intención del autor al publicar estas obras era establecer la infraestructura para convertir el euskera en idioma de cultura. Dentro de este contexto Larramendi tuvo que inventar gran cantidad de neologismos para paliar la falta de vocabulario técnico -lo que él denominaba *voces facultativas*-, labor que combinó con un notable trabajo de campo y un pormenorizado estudio de la tradición literaria. No

consiguió su propósito por la situación sociolingüística del euskara, idioma relegado ya para aquella época a lengua marginal destinada a la confección de catecismos y devocionarios para el pueblo llano: él mismo tuvo que publicar estas obras en castellano, porque en caso contrario, como había ocurrido con Joanes Etcheberri y Domingo Bidegaray en intentos parecidos, la resonancia de su obra habría sido literalmente nula. A pesar de todo hay que señalar que tanto su gramática como su diccionario tuvieron un notable éxito, y que son la base de las normas literarias guipuzcoana y vizcaína a las que nos referíamos más arriba.

Por lo que atañe a su diccionario, por una parte, prácticamente toda la producción lexicográfica vasca del siglo XIX -las obras de Añibarro, Hiribarren y el gran diccionario de Harriet por ejemplo- es deudora en gran medida de la obra de Larramendi, y por otra parte, raro es el escritor de aquella época en la que no se detecten palabras inventadas o puestas en circulación por el *Diccionario Trilingüe*. Hay además un hecho significativo: cuanto más técnico es un texto, más palpable es la influencia de Larramendi en él. Así, el movimiento de los Juegos Florales del último tercio del siglo pasado, que supuso una renovación de una ambición cultural desconocida hasta entonces en lengua vasca, utiliza un vocabulario absolutamente larramendiano. No es de extrañar que sea en esa época cuando se publica el diccionario de Aizkibel (1883-85), que no es prácticamente más que una adaptación del *Diccionario Trilingüe* al formato vasco-castellano.

La tercera época de la lexicografía vasca viene marcada por la aparición en los años 1905-1906 del *Diccionario Vasco-Español-Francés* de R. M^a de Azkue, "obra verdaderamente colosal" en palabras de L. Villasante. Se trata de un diccionario que pretendía recoger todo el caudal léxico de todos los dialectos del idioma. Este proyecto quedó en realidad muy mutilado debido a los prejuicios con los que fue llevado a cabo. En primer lugar, Azkue era purista y populista. Esto le llevó a descartar prácticamente todo el patrimonio de procedencia románica. Este tipo de préstamos se salvó en tanto en cuanto ese origen no fue detectado, o bien por curiosas circunstancias, como en el caso de *puta*, recogida porque "Astarloa sostiene que es palabra originaria del vascuence". Por otra parte, su populismo, y en esto fue también hijo de su tiempo, le hacía desconfiar de la tradición literaria, producida por autores que al recibir toda su formación en lengua estraña estaban irremisiblemente "contaminados". Sólo se libraron de la quema los que según su intuición -necesariamente arbitraria- eran genuinos representantes del habla popular. Su populismo se deja

igualmente traslucir en ciertos comentarios, como el que acompaña a *zeru goien* 'cielo empíreo': "voz muy usada entre los predicadores. No sé si el pueblo se ha valido nunca de ella". Unida a esta visión del idioma literario, o de cualquier registro alto, está la idea implícita que todo escritor no puede sino escribir en el habla de su pueblo natal: el concepto, no ya de norma supradialectal, sino simplemente de norma "supralocal" no tiene sentido para él. En consecuencia, toda palabra que encuentre, por ejemplo en Lardizabal, escritor de Zaldibia del s. XIX, será recogida como "guipuzcoano de Zaldibia". Además, para Azkue, el tiempo no pasa para los idiomas. Así, aunque alguna vez usa la abreviatura *arc.*, no tiene ningún problema en marcar como "alto-navarro de Elcano", sin más, un vocablo documentado en un texto de 1780.

Hay otro aspecto del populismo de Azkue que no podemos pasar por alto: es su actitud ante la actividad neológica de Sabino Arana, fundador del Partido Nacionalista Vasco. Este último combatió los neologismos larramendianos por razones aparentemente lingüísticas que en realidad escondían motivos políticos: Larramendi, para quien el euskera era el antiguo idioma de toda la península, era "españolista". Pero Arana Goiri, al contrario de su antecesor, no sólo inventó términos nuevos para conceptos sin tradición en el idioma, sino también para las palabras de origen extraño, fueran empleadas por el pueblo o no. En consecuencia fabricó centenares de neologismos "según el genio del idioma", la mayoría de los cuales tuvo gran aceptación, sobre todo en los escritores nacionalistas. Azkue, como buen populista, cierra el paso en su diccionario a los neologismos larramendianos (lo que no impide que bastantes de ellos se encuentren, a pesar suyo, en el diccionario), pero tampoco acepta los de Arana por estar "mal formados". Su purismo le lleva a fabricar sus propios neologismos, que como no podían por razones obvias aparecer en el *Vasco-Español-Francés*, encontraron acogida en su *Diccionario de bolsillo* (1918). Hay que señalar que ninguna de las propuestas neológicas de Azkue tuvo el más mínimo éxito.

Todo ello no obsta para que el Diccionario Vasco-Español-Francés deba de ser considerado como la obra lexicográfica más importante de la historia de los estudios vascos. Toda la actividad lexicográfica de este siglo se ha limitado a tratar de completar la obra de Azkue, nunca a sustituirla. El mismo *Diccionario Castellano-Euskera. Euzkel-Erdel Iztegia* (1916), la obra lexicográfica de más éxito de toda la primera parte del este siglo, no es más que un plagio de la obra de Azkue a la que se añadieron todos los neologismos de la escuela arañista.

Como hemos señalado, la lexicografía de investigación de este siglo se ha limitado a tratar de completar la obra de Azkue. Desgraciadamente, se ha incidido sobre todo en los aspectos más negativos de la obra de éste: cuanto más extraña, más marginal, más "curiosa" era una voz, más probabilidades tenía de aparecer en la típica "contribución al diccionario de Azkue", máxima expresión de esta actividad. Como resultado de todo ello, la lexicografía vasca presentaba en los años 80 una distorsión tal respecto al patrimonio léxico del idioma que se hacía absolutamente necesaria una obra que abordara el problema desde planteamientos radicalmente diferentes.

3. En 1956 la Academia de la Lengua Vasca encargó a Luis Michelena una nueva edición del *Diccionario* de Azkue, una vez agotada la primera edición. Se trataba en realidad de proyectar una obra de nueva factura, basada en la obra de Azkue, pero con el formato lexicográfico modificado según nuevas exigencias y planteamientos, algunos de los cuales han sido señalados en los párrafos anteriores. Luis Michelena comenzó a trabajar inmediatamente en el proyecto, que en una primera etapa consistió en la recogida de materiales a la que siguió el inicio de un primer borrador provisional de las primeras letras, hasta la G.

Diversas circunstancias fueron retrasando la continuación de la redacción, aunque la recogida de materiales no se detuvo en ningún momento. De hecho, estos alcanzaron hacia finales de la década de los setenta un volumen que obligaba a pensar en la confección de un diccionario absolutamente diferenciado del *Diccionario* de Azkue. Además, tras el Congreso de Aránzazu de 1968, que puso las bases de la ortografía y la morfología del estándar literario, el idioma se encontró con la urgente necesidad de contar con una obra lexicográfica que permitiera abordar el problema del establecimiento de las formas y contenidos del léxico unificado. Todo ello llevó a Luis Michelena a comenzar a idear una obra de otro formato y otro orden de prelación al cambiar los intereses a que intentaba servir.

La imposibilidad de que él solo pudiera llevar a cabo la obra, le llevaron a ponerse en contacto con el autor de estas líneas. Ambos iniciamos el curso 1979-80 en Salamanca un primer borrador de la nueva obra. Pero pronto vimos que la empresa necesitaba de la colaboración de un equipo que se dedicara exclusivamente a la redacción del diccionario. La ayuda necesaria llegó en 1984 de la mano de las instituciones vascas. Ese mismo año comenzó el despojo sistemático de los textos de la tradición literaria, complemento absolutamente indispensable por aquellas fechas a los centenares de miles de fichas

que el profesor Michelena había recopilado a lo largo de casi treinta años para confeccionar su *Diccionario General Vasco*.

Y es que con la ayuda de las instituciones y en los años ochenta se podía abordar unos de los sueños de Michelena, que era a la vez una tarea imperiosa para la nueva etapa de los estudios vascos en general y del estándar literario en particular: el despojo de todos los textos anteriores a 1745, de los principales libros hasta 1880 -con discriminación positiva a favor de los textos no religiosos y de dialectos marginales-, y de una muestra amplia y representativa de la producción anterior a 1970, trabajo perfectamente realizable en una lengua de literatura reducida como la nuestra.

4. Para entender lo que sigue hay que tener en cuenta que en la tarea de crear esa nueva infraestructura el tiempo era quizás el factor primordial. Había que trabajar en condiciones de emergencia, y preparar en menos de dieciocho meses un estructura de consulta que, aunque estaba pensada en primera instancia como soporte del *Diccionario*, fuera lo suficientemente versátil como para poder ser explotada también por la comisión de la Academia encargada de ir esbozando una gramática del estándar literario.

El reto obligó a tener en cuenta factores que siempre se soslayan en trabajos filológicos. Para citar sólo un botón como muestra, la elección de la fotocopidora podía convertirse en una circunstancia crucial. Supongamos que se trata de hacer 36.000 fotocopias al mes (lor primeros meses superamos ampliamente en más de una ocasión dicha cifra). El utilizar una fotocopidora que realizara una copia por segundo (que es la que elegimos) o una que necesitara 4 segundos por copia suponía una diferencia de treinta horas al mes, es decir, casi una semana de trabajo de una persona al mes, una cantidad desde luego a tener en cuenta en "filología industrial". Igualmente, años más tarde tuvimos que renunciar a un local por lo demás magnífico para la redacción del diccionario, porque su distribución de los espacios nos hubiera obligado a perder miles de segundos al mes en la consulta de originales.

Esta premura en el tiempo condicionó todo tipo de decisiones, y, lo que es más interesante, nos obligó a aplicar cálculos de costes y rendimientos propios de la ingeniería de organización industrial. Por ejemplo, era para nosotros absolutamente necesario, para conocer la influencia real del diccionario de Larramendi en los textos del s. XIX, disponer de esta obra ordenada en sentido vasco-castellano. Una primera evaluación de costos, sobre todo de tiempo, nos hizo ver la inviabilidad del proyecto, pero un estudio más profundo de lo que

era en realidad esencial para nosotros en dicha operación nos indicó que invirtiendo la cuarta parte del tiempo podíamos disponer de un sistema que nos daba cuenta de todos los items de parte vasca con indicación de la entrada castellana en la que se encontraba cada uno de ellos, cosa que suponía prácticamente todo lo que nos interesaba del proyecto inicial.

La mecanización de despojo de textos de la tradición escrita, que en adelante denominaremos el *corpus*, se realizó de la siguiente manera. Por exigencias de versatilidad se eligió un sistema de procesamiento de textos y no uno de base de datos. Se trataba por otra parte de llegar a una solución de compromiso entre dos factores antagónicos: por una parte había que lograr -salvo en los libros más complicados desde un punto de vista filológico- una media de 12.000 pulsaciones por hora en la introducción de textos, lo que obligaba a manipular éstos lo menos posible, pero por otra parte, los redactores deberían disponer de las palabras-texto de corpus correspondientes a un mismo lema agrupadas a poder ser en un único emplazamiento según el orden alfabético. Por ejemplo, *biçiac, biciac, biziac, biziak, viçiac, viciac* etc., deberían aparecer juntas, si no queríamos complicar en extremo -y eternizar- el trabajo de redacción. En consecuencia, no tuvimos la más mínima duda en modificar la ortografía original, aún a sabiendas que no es ésta la práctica habitual en los despojos mecánicos: todas las variantes anteriores se introdujeron en la forma *biziak*, es decir con la ortografía del estándar. De todas maneras conservamos las aspiradas, palatales y todo aquello cuyo interés trascendiera el meramente ortográfico. Igualmente, la palabras que en el original aparecían "marcadas", para señalar su carácter foráneo etc., fueron afectadas por diversas marcas, %, *w*, etc., que dado el interés que tenía para nosotros el respetar el orden alfabético, se añadieron al final de la palabra. En el mismo sentido, la sexta vocal del dialecto suletino, que se representa *ü*, fue transcrita *v*, para que las ocurrencias del común (*h*)*artu* y el equivalente en ese dialecto *hartü* aparecieran lo más próximas posible -hay que tener en cuenta que nuestro procesador ordenaba *ü* inmediatamente delante de *a*-.

Con estas pautas se procedió a procesar las trescientas obras más importantes de la literatura vasca. No se mecanizaron revistas, porque para ello hubiéramos tenido que buscar un sistema de codificación mucho más complicado, sin que las ventajas compensaran los inconvenientes. De hecho, un sistema de 5 cifras fue suficiente para una codificación que nos permitió clasificar cada obra según época (por períodos de 50 años), según dialecto (tanto los literarios como los no literarios) y según tipo de obra con los tres primeros dígitos,

reservando el cuarto y quinto para autor y obra. Así, por ejemplo, 83641 representa una obra del período 1900-1950 (8), en dialecto labortano (3), de prosa literaria (6), de Jean Etchepare (4), en concreto *Buruxkak* (1). Este código se introducía al empezar a mecanizar el libro; el cambio de página se hacía manualmente en el primer punto o punto y coma tras el cambio de página físico, para que el contexto de las últimas palabras no quedara mutilado.

Todo este dispositivo se estableció trabajando conjuntamente con los miembros del equipo informático, en función de los problemas que ellos podían subsanar con costes razonables en el programa informático. Tras el trabajo de los programadores, pudimos disponer de los aproximadamente cinco millones y medio de palabras-texto del corpus en orden alfabético en la estructura que puede verse en la página contigua.

Por razones de todo tipo, en nuestra situación y en 1984, era absolutamente impensable cualquier tipo de lematización automática. De todas formas, gracias al trabajo de los programadores pudimos disponer de listados en los que, como puede apreciarse, la *h* no cuenta para el orden alfabético, y las palatales cuentan como las no palatales, en cualquier posición en la palabra: es decir *habailla*, *abailla*, *abaila*, *habaila* son indistintos a efectos de orden en el listado. Las ocurrencias o palabras-texto formalmente iguales se ordenan por el orden alfabético de la palabra siguiente, lo que facilita enormemente el trabajo en la redacción de locuciones o verbos compuestos, tan frecuentes en euskara como *(h)itz egin* "hablar": si las más de 700 ocurrencias de *(h)itz egin* estuvieran desperdigadas entre las aproximadamente 5.200 ocurrencias de *(h)itz*, el redactor de esta subentrada perdería previamente centenares de segundos punteando aquellas entre éstas. Además, las ocurrencias que coinciden también en la palabra siguiente están ordenadas alfabéticamente, es decir, las más de 700 ocurrencias de *(h)itz egin* se presentan al redactor en orden cronológico de antiguas a modernas. Esta disposición del corpus permite tener el material de más del 90% de las entradas léxicas en un solo bloque, con un contexto suficientemente amplio y perfectamente identificadas, es decir con todas las ventajas de la lematización concordada.

Quedaba otra tarea pendiente antes de abordar la redacción: la mecanización de los centenares de miles de fichas lexicográficas recogidas por Luis Michelena. En este caso se escogió una especie de rudimentaria base de datos. Michelena tenía previsto un uso múltiple para algunas de su fichas. Por ejemplo la ficha ABANDONATU Iz Als 0000 de la muestra de la página contigua, se utilizaría, después de

en la entrada *abandonatu*, en *dezente*, *irristaka*, *euri*, y *gurdi mutur*, sacando la ficha de su localización inicial e insertándola, una vez finalizada la redacción de la entrada léxica *abandonatu*, en las fichas correspondientes a *dezente*, para pasar posteriormente a *euri* etc. El modelo elegido trataba de imitar este proceso. La ficha mecanizada, aparte de las casillas de código, disponía sólo de dos campos; uno para el texto de la ficha y otro para el encabezamiento, en el que una coma separaba los ítems. Así, por ejemplo, con la secuencia *gurdi mutur* la ficha aparecería en la *g*-pero no en la *m*-, mientras que con *gurdi, mutur* aparecería en las posiciones *gurdi* y *mutur*. Un dispositivo tan simple permitió una mecanización notablemente rápida y eficaz de las fichas del autor del diccionario. Este sistema se utilizó también para mecanizar el despojo -no sistemático desde luego- de las revistas más relevantes para nosotros de los últimos cien años.

5. Como indica el autor en el prólogo de su obra "el Diccionario General Vasco trata de recoger y sistematizar el patrimonio de todas las épocas y dialectos. Se trata, por lo tanto, de una obra general y descriptiva, no de un diccionario normativo. Se basa en un material clasificado mecánicamente que comprende: a) los testimonios clave de las distintas épocas en que la lengua vasca aparece atestiguada de alguna manera, desde la documentación antigua hasta una fecha reciente que se puede fijar, provisionalmente, en 1970 [...]; b) los textos propiamente dichos desde la documentación medieval hasta la fecha citada [...]; c) aunque se trata de un elemento adicional en el texto destinado a la ubicación, se ha tenido siempre muy presente la onomástica personal y local, fundamental para bastantes siglos".

Es decir, se trataba de confeccionar una obra que recogiera todo aquello que tuviera interés para la lexicografía vasca y, en general, para los estudios vascos. Esta variedad de objetivos nos llevó a un formato lo más versátil posible, con una disposición de los artículos que permitiera dar en éstos el mayor tipo de informaciones posible sobre el mayor número de palabras posible: en principio, sobre todas las palabras y expresiones empleadas realmente por los vascos cuando han hablado o escrito en su idioma.

Las fuentes del *Diccionario General Vasco* son orales y escritas, en partes muy desiguales, claro está. En cuanto a las orales, no se ha hecho, por motivos obvios, prácticamente trabajo de campo, pero se ha recogido sistemáticamente todo lo publicado de este tipo. Por lo que respecta a las fuentes escritas, se ha recogido sistemáticamente toda la información procedente de otros léxicos, pero señalando nuestra prevención hacia las "palabras de diccionario" de las que no

haya testimonio independiente, y desde luego se ha primado claramente, por los motivos apuntados más arriba, la información recogida del despojo de la tradición literaria, es decir de los textos. En consecuencia, se puede afirmar que el *Diccionario General Vasco* es sobre todo un diccionario histórico, que ha de ser necesariamente crítico. Esta labor crítica, que es una exigencia general al ser la lexicografía un terreno en el que el acarreo de errores, datos mal interpretados etc., es la norma común, tiene que ser ejercida en la lexicografía vasca con una especial intensidad, pues como no se cansó de repetir Luis Michelena, la historia de nuestra lexicografía, en manos en muchos casos de personas demasiado inexpertas, es particularmente turbia o turbulenta.

Acabaremos con una descripción resumida de los aspectos más relevantes de la estructura del diccionario, del que se ofrece una muestra en la página contigua.

En el encabezamiento de los artículos se agrupan todas la variantes atestiguadas bajo un único lema, ya que la dispersión de datos, que es una característica común de casi toda la actividad lexicográfica vasca, aparte de desdibujar la unidad del idioma, tiene desde un punto de vista práctico el nocivo efecto de desorientar al lector. Junto a cada variante se indica entre paréntesis, el área dialectal en la que se emplea o se ha empleado, y tras ella y dentro del mismo paréntesis, los lexicógrafos que la recogen. Hay que señalar que en este diccionario las ausencias son significativas, y que la falta de referencia a S. Pouvreau, Urte, Larramendi, Harriet etc, indica que, salvo error, dicha variante no se documenta en el lexicógrafo correspondiente.

Las acepciones en que se divide cada artículo se ordenan en principio cronológicamente. En cada una de ellas se da el equivalente castellano y, siempre que sea pertinente, sinónimos y antónimos vascos. A continuación se recogen las informaciones semánticas de interés aportadas por otros lexicógrafos.

Tras la abreviatura *Tr.* se da la información relativa a la historia de la palabra o acepción correspondiente, a la que siguen los ejemplos de utilización que ilustran lo esbozado en la información anterior. Pretendemos con los ejemplos ofrecer una imagen fiel del amplio corpus descrito más arriba, dando al menos un ejemplo de todos los autores importantes. También aquí la falta de ejemplos de alguno de ellos indica que, salvo error, no lo hemos encontrado en nuestro corpus ni en los textos que hemos despojado sistemáticamente. Hay que señalar que dado el excesivo volumen que estaba tomando la obra, a partir del quinto volumen se ha restringido drásticamente tanto el número como la longitud de los ejemplos, de manera

que ahora sólo se dan los ejemplos interesantes en sí. De los demás se da simplemente la referencia, bien en *Tr.*, bien en el nuevo apartado *v. tbn.* Dado el carácter de la obra los textos se dan en lo que en la jerga de la filología vasca se denomina "ortografía dialectal modernizada", es decir la del estándar, pero respetando la notación de palatales, aspiradas, grafías etimológicas (*hitztegi*), etc.

Finalmente, tras las acepciones de la entrada léxica se han especificado las formas compuestas y locuciones de interés. Estas sub-entradas, de importancia crucial en muchos casos en el establecimiento del estándar literario, han sido confeccionadas con la misma estructura y la misma exhaustividad que las referencias y ejemplos de las entradas generales.

En algunos casos, al final de los artículos se añaden la informaciones disponibles sobre el origen de dicha entrada léxica, cuando, como señala el autor en el prólogo de la obra "se trata de cuestiones muy disputadas o procedencias cuyos eslabones se acierta a fundamentar". El estado actual de los estudios vascos no permite ir más lejos.

Nada más, salvo añadir que en estos momentos está punto de ver la luz el octavo volumen de la obra que abarca las entradas *fe-gub.* Según nuestros cálculos, con este volumen se ha llegado a la publicación de casi la mitad del diccionario, que esperamos acabar en ocho volúmenes más, el año 2003 o 2004.

abaillan mintza oi zen, otoiizean diardunaren antzo. Etxde Itxas 22. ♦ 3. «Abailla, con intensidad, con fuerza: curia egin du abailla: Hantxe ari duk abailla lanean. Abailla demonio etorri nintzen» Asp Gebi.

■ *Etim.* No parece que Schuchardt llegara a encontrar un origen rom. satisfactorio.

habailada (Lar, Añ, A; Lh, que cita a Hb), *habailkada*. ♦ Hondazo, golpe de honda. ★ *Goliat erraldoia, abailkada bat eman-ta Dabid erregeak Terebintoko ibarrean il zuana.* Berron *Kijote* 20.

habailaka (SP, A). «À coups de fronde» SP.

habailari (SP; Lar, A; Lh, que cita a Hb), *habalari, abailkari*. ♦ *Hondero.* ★ *Bainan mendietara burbildu zirenean habalariek atheratu ziren hekien kontra.* Dv *Iudith* 6, 8. *Eta bivia inguratua izan zen habalariez, eta zathi handi batean joa.* Dv 4 Reg 3, 25. *Bañan, mendira urbildu ziranean, abailkariak aien aurka irten ziran.* Ol *Iudith* 6, 8. *Antxiñako gerretan, abillari euskaldunak agertu izan ziran.* Etxba *Eib. Abailariak, arein aurka arriak jaurtirik, igotea galerazoten eutseen artean.* Ker *Iudith* 6, 12. *Diagü Utziagaiñalako / elkebigia etxekeitzen, / Baitere Akberzothalian / habalariek gordatzen.* Casve *SGrazi* 160.

habailarri. ♦ Piedra lanzada por una honda. ★ *Bai, amar... edo asko beintzat, ardi, beor, moxal ta otso danak nastean [...], abail-arria <aria> baño zunburruntuago, ego aizerik gogorrena baño zazpi milla bidar ariñago ta zakarrago.* Ag G (ed. 1935) 128 (1ª ed. 157 *aballako arria*; cf. *DRA* s.v. *aballari*).

habailatu. «ondrer, lancer avec la fronde; sin. de *habailkatzea, habailkatzea*, attaquer à coups de

fronde, de pierres lancées avec la fronde» H.

abaildu (Vc, G, AN-gip). Ref.: A (*abaildu, abaindu*); Iz *ArOñ*. ♦ 1. *Abatir(se), fatigar(se).* v. *hebaindu; akitu*. ■ *Difundida, al parecer, a través del dicc. de Azkue.* ★ *Zu, launa, arika ta abaildurik | iarri zinean neure bila. 'Quaerens me sedisti lassus' (Dies Irae). Azc PB 29 (la versión original, no retocada por su hijo dice (v. *CancB* II 27 y *Eusk* 1959, 60): *Zu, Jaun andija, jarri ziñian / nekatu ta neure billa*). *Baño gorputza lotua ta abaildua gelditu bazitzaion, ez orratik bere anima: anima beñere baño argiago, bi-ziago ta azkarragoa zeukan.* Ag G 204. *Jangar ta nakar, abaildua ta ezertarako eza.* Ib. 79. *Leon Olazarreta, oso abaildua dago, latzak eta beltzak ikusten bizi da errukarri au.* A *Ardi* 99s. *Errukarriago, atsekabetuago, abailduago ta gogoz ilago.* Ib. 100. *Biurkortasun ori ezthanean, geiegizko irutzipenaz gogo abaildu ta argaltzen da edo edozein adigakuntzaz kezutuzen da.* Zink *Crit Ees* 1918, 172. *Euskera ludiko ele kemendungazteenetakorik bat dala diozute: aurgintzaren nekeak ezpaitu abaildu, ez-eta igatu erabilli luzeak.* Ldi *IL* 86. *Bera izan dan gizon sendoz-tal... Era orretan abailduta, dangilizka, intziriz ikusi ta entzunez, barru guztiak ere bira-egin dauts.* Erkiag *Arran* 74. *Gorputz eta arima, ondatuta dago. Gorputza, makal ta indarbage, geisorik. Gogo abailduta, adorea galduta.* Erkiag *BatB* 191. *Lanaren lanez abailduta dauntzala-ta.* Zait *Plat* 117. ♦ 2. *Abatirse los árboles con el peso de la fruta. «Asi xate abailketan erramaak, inclinarse las ramas con el peso de la fruta».* Iz *ArOñ*. ★ *Arto musker, / mendi, baserri zârrak; / ale gorriz / abailduta sagarrak.* Ldi. *BB* 64. *Arloak, un-**

dazken-aienez abailduta, zuri dagozkitzu uzta betean eta maats-billetak gaiñezka dagi kupeletan apar-iaro. Ibiñ Virgil 82. Edoi tzarren antzera pillotzen dira ta taldeka zugaitz gaiñean kokatzen adarrak abailduz. Ib. 118. Basoak ere edotariko igaliz abailtzen dira. Ib. 91. ♦ 3. Sg. A «(V-ple-oro-z-m) madurarse bien la fruta».

- ABAIL EGIN. ★ Oian utzi dot ondo-ezik, abail-abail egindda, eta nik ulertzen dodanez beruenak artuta. Otx 161.

■ Etim. Cf. acaso rom. *aballar*, v. DCECH, y REW *ballare*.

abaildura. ♦ Fatiga. ★ *Oarrikaspenak egiztutzen du soñaren edo giartzaren abaildura utsak ez duala beti adimen eta zillei beera-pena ekartzen. Zink Crit Ees 1924, 117.*

abailera. ♦ Disposición, marcha (?). ★ *Ettxeko aballera txarrak tabernarako gogoia ematen du: emakumeak jakin bear du etxea txukuntzen [...], etxea pozgarri egiten. Ayerb Ees 1916, 33.*

abain(a?) v. **amaraun**.

abakando (V-m), **abakondo** (V-ger, G-azp). Ref.: A (*abakando*, *abakondo*), Laff 132 (*abakondo*). ♦ Bogavante (*Homarus gammarus*). ★ *Nasa edo otar oetan, langosta ezik, misera (abakando) sangurru, olagarro edo amarratz (pulpo), ta noizipeiñ aingiraren bat edo beste ere artzen zituzten. Elizdo Ees 1926, 7.*

abakondo. «1° (V-ger) nudo de árbol. 2° (V) tronco seco» A.

abakondo v. **abakando**.

abalaio (V-m), **abalu** (G), **abario** (V-m). «Un pececillo de mar» A.

abaldu v. **afaldu**.

abalu v. **abalaio**.

abaña (Sc ap. A; H sg. Lh, pero no lo hemos encontrado). Cf. «*Abañ-oibal* (S), gaze, tul» Lh. ♦ *Telaraña* v. **amaraun**. ★ *Bere obretan haiñ aizaldir non hil denian*

abaiñak beno azkarrago ezpeitirade. Egat 257.

abandonamendu. ♦ Abandono. ★ *Onbesten ditut oro-zko abandona-mendua, ene bibotz kbrrestasunak, ene langidurak, ene penak, ene tentazione eta ene gaitz guzien embrazuak, zure iustiziari satisfatzeko. JesBih 455s.*

abandonatu (Urt I 17, *VocBN*, Gèze, Arch *VocGr*; no en SP, Lar, Asizk, H, A ni Lh). ♦ **Abandonar.** ■ Tr. Su uso parece restringido a los dialectos septentrionales, e incluso en éstos, los únicos que lo emplean con cierta profusión son autores de importancia menor o ligados a la tradición oral. El equivalente es por lo general *utzi*, incluso en *Leizarraga* (para Ax v. *Letamendia ASJU* 1981 s.v. *desero*), y parece que cuando hay que precisar más, autores como *Duvoisin*, *J. Etchepare* etc. se inclinan por fórmulas como *bazterrer(a) utzi*. En los dialectos meridionales sólo lo hemos encontrado en *Xenpelar* y *Udarregi*, y en el *baztanés Echenique*; los demás emplean *utzi* o fórmulas como *bertan bera utzi* etc. En *DFrec* hay 21 ejs., uno de ellos septentrional. ★ *Ene Iainkoa, ene Iainkoa, zeren abandonatu nauk?* Lç Mt 27, 46, (TB *zertako utzi nauzu?*, HeH, Dv *zergatik utzi nauzu?*, Ur *zergatik utzi nazu?*, Hual *zeren desanparatu nauzu?*, Ol *zergatik zapuztu nauzu?*); v. tbn. Mc 15, 34. *Hai-zeari unzia abandonaturik eramaiten gener.* Lç Act 27, 15 (TB *eramamak izan ginen, utzirik untzia haizeari*, Dv *haizeen gogara utzirik, barekin eramamak ginen*, Ol *ontzia narrastuz; ta ezin ari burutuz bere oldarrari utziak genbiltzan*). Ez *nazazula othoi abandona neure azkeneko partidan*. Arg *DevB* 154. *Bañan, Jauna, ez nazazula abandona, eta ez zaitzela urrum eneganik.* Gç

46. *Eta ez dezala abandona haren etsaiaren furia*. Ib. 54. *Abandonatu zen osoki Jainkoaren borondata*. Ch I 25, 2 (SP jarri, Mst jarri). *Ez dakigun sekulan abantz, manamendu xoil batean huts egitea, lege guziaren haustea eta abandonatzea dela*. CatLav 102 (V 57). *Paganotasuneko superstizio-neak eta banitate guziak, zeinak izan baitziren utziak eta abandonatuak Herri hautako jendez San Leonek predikatu ziotenean*. Ib. 420 (V 200). *Ez nazazuela, ni zuen zerbitzaria, abandona edo ene mena utz*. He Gudu 93. *Orduan abandonatzen dio bere burua xargiñari, eta gaitzen ditu esparantza guziak*. Ib. 52. *Merezi zinduen beraz Jainkoaz abandonatua izatea*. Brte 249. *Haizearen pare den obo-rezko pontu bano baten gatik [...]* *ahantzi, abandonatu, meziprezatu zaitudala*. Ib. 200. *Zeren baitakit ez nauzula abandonatuko*. Mih 105. *Erakustera eman zien Jinkuak abandonatü ziela*. CatLan 44. *Eta proposatzen nüzü azkarki zure eta zure seme Dibinuaren zerbütxiaren ez jagoiti abandonatzer*. Mercy 22. *Oh ehiza nabi huillantu / ene Jainko handia? / Nabi duka abandonatü / Nabarruko jentia?* Xarlem 581. *Jinkuaren legiak oro / abandonatürik dütüzü*. Ib. 1273. *Gisa hontan hiltzeko / gütüka abandonatzen?* Ib. 1149. *O Baboumet gure jinkua / Ezkütüko ikusten (sic)? / Abandenatü [sic] gütiála / Orai diat sinhesten*. Ib. 999. *Ez nazazula, othoi, abandona*. Dh 44. *Abandona gaiten probidentzia maitagarri haren eskuetara*. Jaur 389. *Eta deklaritzen derauku ez duela abandonatuko bere eskuko obra, begia haren gainean izanen duela*. Ib. 106. *Abandonatzen ziuzten bere onthasunak, edo po-brei ematen ziuzten*. Ib. 202. *Ez othoi abandona zure semiak!* Etch 636. *Malurra malur dela / nun da*

ohoria? / abandonatu dute / beren faboria. Bordel f165. *Ez gaitzan sekulakotz, / otoi, abandona, / tormentu aundietan / sar ez gaiten barna*. Ib. 198. *Ez da herririk hain abandonaturik zeinetan ez baitire aurkhitzen zenbeit arima leial*. Marll 8. *Jainkoaren Ama ez baita kasik ezagutua ere hainitez, eta gehienez abandonatua baita!* Ib. 7s. *Ene jinkua zeren abandonatü naizü*. UskLiB 42. *Hitzaz abandonaturik nizala / orai diat ikusten*. Xarlem 1336. *Eta gero gitian retira / Baionarat hebeti / abandona Biarno / gure etsaiari*. Ib. 438. *Baiñan orai zabartua / naduka abandonatua*. Gy 261. *Nere Yangoikoa, nere Yangoikoa, zertako abandonatu naizu?* Echn Mt 27, 46. *Denbora da aitzina juaiten, / Sekül'ez argitzen, / Üdüritzen zait naizüla abiñdenatzen*. ChantP 116. *Zure arartekotar-zünari hersatü dena izan dela abandonaturik*. Ip Hil 48. *Ez nezazula abandona, o ene Jainkoa!* Arb Igand 203. *Anima galdu kulparik gabe, / gorputzak abandonatzen... / despeida txarra guretzat dator / ezpagerade mudatzen*. Xe 343. *Andriak gaitu penatzen, / asi zait abandonatzen; / pazientziya eziñ arturik / bizi naiz ernegatutzen*. Ib. 273. *Emakume sobatuak, / frantsesak in oratuak, / itz batian esateko / abandonatuak*. Ib. 215. *Txala bere salletik / abandonatuba*. Ud 102. *Jainkoaz othoi, ez gaitzazula abandona!* Lf Murtuts 28. *Zeruetako Jaunari oiugiten zion saminki bhaktasun ikaragarri artan abandona etzezan*. Etxde JJ 248. *Gogorra zait joatea: nere sorterriari utzi, nere gizaroa galdu, nere erriari aldegin...* *Miren bera gezurtara ta abandonatu, dena sentitzen dut errai-muñean*. Txill Let 106. *Andra txarra zuan baña olan ezingo juat abandonau*. SM Zirik 120. *Ez naiz itzuliko egundo sekuletan /*

abandonatu *dudan haritzi beltzera*. Arti MaldanB 195. Bere seme alabez abandonatua. Xa Odol 207.

habanera. ♦ Habanera. ★ Batek esaten du «Vals». Jotzen dute. Gero «Habanera», ta au bukatzian pasakallia juaz dijuaz. Sor Gabon 50. Abarkadun gutxi ikusiko dezu erderaz mintzatzen edo abanera, xotis ta beste arrotz dantza nastuetan. Ag G 96. Maitearen illobi gañean. (Habanera-aidea). Netx LBB 380.

habano, habana. ♦ (Puro) habano.

★ Txomiñek kafea, koñaka eta habano bat ekarri zuan. Osk Kurl 217. Karamelu luze bat, habano baten itzurazkoa. Ib. 49. Geinertikoeek habana luzeak erretzen zituzten. Mde Pr 134.

abaño. v. 2 amaño.

abant (Vc, G, L ap. A; Aq 352),

abanti (Aq 352). ♦ Acción de remar. «Ea, andar, entre los que reman ea, abanti, abant ababor o estribord» Aq 352. ★ Arraiñaren jarrai urrunago ta urrunago, beti abant eta abant, tosta ganean jezarrita, indarren indarrez untzia abiau-araziaz. Erkiag Arran 79.

- ABANT EGIN (A s.v. abastean).

Remar. ★ Alu orrek, abant, eizu abant. A s.v. alu.

- ABANTAN (A). Remando. ★ Zelan siniztu dozu, / Apolon kuttuna, / izan dala abantean / goitua eskalduna? Azc PB 225. Ikusten Bizkaitarrak / santzoka abantean: / eta egaz txalupea / euren ez aurrean. Ib. 225. Aunek emakumeok be, eztira makalak abantean [...]. Donostiako estropadetan eskasagorik izango da txanpan egiteko. Erkiag Arran 121.

DICCIONARIOS Y CORPUS TEXTUALES. PERSPECTIVAS PARA EL CATALÁN

JOAQUIM RAFEL I FONTANALS

Universitat de Barcelona
Institut d'Estudis Catalans

La lexicografía actual y los corpus textuales

No creo que sea necesario explicar ante un auditorio como éste que el concepto de corpus textual está íntimamente relacionado con el método lexicográfico moderno, puesto que sobre todo en los últimos años se ha ido haciendo habitual relacionar los dos conceptos que figuran en el título de mi conferencia: diccionarios y corpus textuales.

Es de todos sabido que la lexicografía experimenta, fundamentalmente a partir de los años sesenta, una renovación metodológica que se va afianzando con lentitud en los años subsiguientes¹. Los primeros artículos críticos sobre aspectos concretos del método lexicográfico y las primeras obras de conjunto, como las de Jossette Rey-Debove, de Jean y Claude Dubois o de Ladislav Zgusta², por ejem-

¹ Es digno de mención el precedente de Julio Casares, que en su obra *Introducción a la lexicografía moderna* (Madrid, 1950) lleva a cabo una de las primeras reflexiones que se conocen sobre la actividad lexicográfica. En fechas anteriores y en los mismos años cincuenta, en que es general la falta de interés por los aspectos relacionados con el método de trabajo en lexicografía, son contados los trabajos que se ocupan de este tema. También merece una referencia explícita la reunión de lingüistas y lexicógrafos que tuvo lugar en noviembre del año 1960 en Bloomington (actas publicadas en Fred W. Householder & Sol Saporta; *Problems in Lexicography*, Indiana University, Bloomington, 1962) y que representa un hito en la bibliografía sobre el tema; los especialistas reunidos en aquella ocasión tomaban conciencia de que una nueva época comenzaba para la lexicografía, como declara C. F. Voegelin, editor de la serie en que se publica la obra, en las breves palabras que escribe al inicio del libro: «The planners and the authors of the conference deserve congratulations on summing up the end of an epoch in lexicography, now that we are on the eve of another epoch which will give more emphasis to typology and include the use of computers in lexical work.»

² J. Rey-Debove; *Étude linguistique et sémiotique des dictionnaires français contemporains*, Mouton, La Haya-París, 1971.

plo, han sido seguidos por una gran cantidad de títulos, tanto sobre cuestiones particulares como de carácter general, que se han ido multiplicando con profusión; la bibliografía sobre lexicografía publicada por el mismo Zgusta³ nos da una buena información sobre los trabajos aparecidos hasta finales de los años ochenta, y la reciente y extensísima obra enciclopédica en tres volúmenes, de carácter exhaustivo, editada a cargo de Franz Josef Haussmann⁴, fija el estado de la cuestión en los más diversos aspectos del quehacer lexicográfico. Además, en los últimos años han surgido publicaciones periódicas especializadas en léxico y en lexicografía, aparte de algunas recopilaciones de artículos sobre este tema⁵.

Este proceso renovador surge como reacción ante el carácter asistemático de la lexicografía que denominamos tradicional, caracterizada fundamentalmente por la acumulación de datos y por la ausencia de unas bases teóricas, teniendo como fuente principal de inspiración los otros diccionarios publicados previamente. En las bases de este movimiento de renovación encontramos, en cambio, la necesidad de incorporar a la actividad lexicográfica, por una parte, los avances que se han producido en el campo de la lingüística en los últimos decenios, y, por otra parte, las innovaciones tecnológicas que han tenido lugar en el ámbito del tratamiento masivo de la información con la aparición y el avance vertiginoso de las técnicas informáticas.

No es ésta la ocasión adecuada para analizar con detalle cada uno de los aspectos en que incide la renovación metodológica a que me refiero; sólo voy a mencionar aquí que uno de los logros fundamentales de la nueva lexicografía es la objetivación del método: la subjetividad y la intuición dejan de ocupar un lugar preeminente en

J. & C. Dubois; *Introduction à la lexicographie: le dictionnaire*. Larousse, París, 1971.

L. Zgusta; *Manual of Lexicography*, Academia-Mouton, Praga-París, 1971.

³ L. Zgusta; *Lexicography Today. An Annotated Bibliography of the Theory of Lexicography*. Max Niemeyer, Tübingen, 1988.

⁴ F. J. Haussmann (ed.); *Wörterbücher. Ein internationales Handbuch zur Lexikographie*, de Gruyter, Berlín-Nueva York, I, 1989; II, 1990; III, 1991.

⁵ Aparte de la ya histórica *Cahiers de Lexicologie* (París, desde 1959), recordemos las más recientes *Dictionaries*, órgano de la Dictionary Society of North America (DSNA) (desde 1979), *Lexique* (Lille, desde 1982), *Lexicographica. International Annual for Lexicography*, publicada por la misma DSNA y por la European Association for Lexicography (EURALEX) (Tübingen, desde 1985) y *International Journal of Lexicography* (Oxford, desde 1988). Entre los libros que recogen un conjunto de artículos sobre lexicografía, cabe citar *Lexicography: Principles and Practice*, editado por R. R. K. Hartmann (Academic Press Inc., Londres, 1983) y *A Spectrum of Lexicography*, editado por Robert F. Illson (John Benjamins Publishing Company, Amsterdam, 1987).

la actividad lexicográfica para dejar paso a los datos objetivos, y en este proceso de objetivación del método lexicográfico tienen un papel muy destacado los corpus textuales. La utilización de un corpus textual adecuadamente representativo como fuente principal en el proceso de elaboración de un diccionario, substituyendo en este cometido al conjunto de diccionarios preexistentes, se convierte así en una de las características principales de la lexicografía actual, hasta el punto que se ha establecido una frontera entre los diccionarios basados en corpus y los diccionarios no basados en corpus.

En la última parte de mi intervención voy a comentar más detalladamente algunos de los aspectos de la configuración del producto final -el diccionario- en que este cambio de orientación metodológica influye de una manera decisiva. Voy a puntualizar sin embargo ya desde ahora que existe una gran diferencia entre la utilización de un corpus textual en la forma que acabo de indicar y el empleo de la información que nos proporciona un corpus como mera ayuda o complemento a una labor lexicográfica efectuada a partir de otros supuestos. Dicho de otro modo, cuando hablamos de los corpus textuales relacionados con la elaboración de diccionarios no nos referimos a un uso accesorio, como cuando se remoja un edificio viejo y se le da una apariencia de nuevo, sino que aludimos a una utilización de los datos proporcionados por el corpus desde el momento de la concepción del diccionario, de tal modo que esos datos condicionan la estructura y la configuración tanto general como particular de la obra.

Tengo que decir, sin embargo, antes de dar fin a estas palabras introductorias, que, para que un corpus pueda desempeñar cumplidamente en una empresa lexicográfica la función que aquí le atribuímos, debe reunir ciertas características indispensables, que ahora, para no alargarme excesivamente en cuestiones preliminares, resumiré en una sola palabra: representatividad. En efecto, para que un corpus pueda ser utilizado con garantía como base fundamental para la elaboración de un diccionario de la lengua, debe tratarse de un corpus realmente representativo de la realidad lingüística que se pretende describir en aquella obra, tanto desde la perspectiva cronológica (ámbito temporal que abarca el diccionario), como desde la perspectiva tipológica (clases de manifestaciones lingüísticas que pretende tener en cuenta). Dejo de lado, pues, otro tipo de consideraciones más particulares en las que ahora no puedo extenderme.

El catalán y la lexicografía basada en corpus

Es precisamente dentro del contexto metodológico que acabo

de presentar de una manera muy sumaria, que hace ya unos años la Secció Filològica del Institut d'Estudis Catalans⁶ concibió el proyecto de elaborar un diccionario descriptivo de la lengua de aproximadamente los últimos 150 años; en el marco de este proyecto, que se ha denominado Diccionari del català contemporani (DCC), y como primera fase, o fase previa, de él se concibió la elaboración de un corpus, denominado Corpus textual informatitzat de la llengua catalana (CTILC), que en el momento actual se halla próximo a su culminación⁷.

⁶ El Institut d'Estudis Catalans es una corporación académica, científica y cultural que tiene por objeto la alta investigación científica, principalmente la de todos los elementos de la cultura catalana. Fue creado el año 1907 por un acuerdo de la Diputación Provincial de Barcelona, a iniciativa de su presidente Enric Prat de la Riba, que más tarde sería presidente de la Mancomunidad de Cataluña; después de un primer período floreciente y de una existencia agitada, siguiendo los avatares por los que ha pasado el país, actualmente su situación legal se basa en un Real Decreto del año 1976. De acuerdo con sus estatutos actuales (renovados en 1989), sus finalidades son: a) Ocuparse del estudio de la lengua, establecer su normativa y velar para que su proceso de normalización sea coherente en todo su ámbito lingüístico. b) Contribuir a la planificación, la coordinación y la realización de la investigación en las distintas áreas de la ciencia y la tecnología. c) Impulsar con las actividades que le son propias el progreso y el desarrollo generales de la sociedad y realizar, cuando corresponda, trabajos de asesoramiento de los poderes públicos. Aunque son competencia del conjunto de la institución, los trabajos relacionados con la lengua conciernen especialmente a la Sección Filológica. Actualmente las otras secciones del Institut son la Histórico-arqueológica, la de Ciencias biológicas, la de Ciencias y tecnología y la de Filosofía y ciencias sociales.

⁷ La decisión de llevar a cabo este proyecto fue tomada en 1984. Durante los primeros cuatro años, con recursos escasos, se establecieron los detalles de las bases metodológicas de los trabajos, se creó la infraestructura material que lo había de soportar y se elaboraron los programas básicos para llevarlo a cabo, se formó el equipo de especialistas necesario y se iniciaron los trabajos propiamente dichos a título experimental. A partir de 1989, con todos los requisitos previos y el proyecto claramente definido para poder trabajar de una manera productiva, un incremento notable de los recursos disponibles (gracias a un convenio específico con la Secretaría de Estado de Universidades e Investigación del Ministerio de Investigación y Ciencia) permitió acelerar la realización del proyecto hasta 1992, en que un replanteamiento de su financiación produjo un lapso en los trabajos hasta 1994; a partir de esta fecha, un nuevo convenio con la Secretaría de Estado de Universidades e Investigación y la Comisión Interdepartamental de Investigación Científica y Técnica de la Generalidad de Cataluña ha permitido reemprender el ritmo de trabajo y asegurar la terminación, a final de 1997, del corpus previsto y de las labores previas a la elaboración del diccionario descriptivo.

Una información más detallada sobre la justificación, el inicio y la ejecución de este proyecto puede encontrarse en varios trabajos que he publicado a lo largo de su desarrollo; a través de ellos puede seguirse su trayectoria desde el origen: *Cap a un diccionari del català contemporani*. «Segon Congrés Internacional de la Llengua Catalana, IV, Àrea 3: Lingüística Social» (celebrado en Palma de Mallorca del 30 de abril al 4 de mayo de 1986), Palma, 1992, págs. 589-595. *El «Corpus textual informatitzat de la llengua catalana»* (en colaboración con J. M. Solanellas). «Actas de las II Jornadas Españolas de Documentación Automatizada. 20-22 de Noviembre de 1986. Ponencias y

Características generales del CTILC

Las características fundamentales del CTILC, que nos permiten hacernos cargo de su naturaleza, son las siguientes: a) Desde el punto de vista temporal abarca desde 1833, que es la fecha simbólica de la recuperación de la lengua catalana escrita en la época contemporánea, hasta 1988, que marca el final del último grupo cronológico en el momento en que se llevó a cabo la selección de los textos; cubre, por lo tanto, un período de 155 años. b) Desde el punto de vista tipológico, el CTILC es un corpus de la lengua escrita; incluye textos de carácter literario y textos correspondientes a la lengua no literaria (es decir, textos que se ocupan de la más diversa variedad de temas, con los más distintos objetivos, sin una intención artística o estética); cada uno de estos dos tipos de texto ha sido dividido en otros, que han permitido una selección equilibrada de la cantidad de texto a tener en cuenta, no sólo para cada tipo de lengua (literaria o no literaria), sino para cada subtipo (los cuatro géneros tradicionales: ensayo, narrativa, poesía y teatro, para la lengua literaria, y diez grupos temáticos o funcionales distintos para la lengua no literaria: 0. Correspondencia, 1. Filosofía, 2. Religión y teología, 3. Ciencias sociales, 4. Prensa, 5. Ciencias puras y naturales, 6. Ciencias aplicadas, 7. Bellas artes, ocio y deportes, 8. Lengua y literatura, 9. Historia y geografía). c) Por lo que respecta a la extensión del corpus, supera los 52 millones de ocurrencias o palabras de texto (el recuento actual es exactamente de 52.374.665)⁸.

Otras características del corpus, que sólo mencionaré, son las

comunicaciones» (Torremolinos-Málaga, 1986), págs. 147-161. *El «Diccionari del català contemporani»*. «Serra d'Or», 1987, págs. 428-431. *El «Corpus textual informatitzat de la llengua catalana» i el «Diccionari del català contemporani». Un projecte del Institut d'Estudis Catalans*. Anthropos, 1988, Documentación cultural e información bibliográfica, págs. V-VII. *El «Diccionari del català contemporani»: treballs realitzats i previsiones de futur*. «Llengua & Literatura», Revista anual de la Societat Catalana de Llengua i Literatura, 5, 1992-93 [1994], págs. 733-737. *Le «Diccionari del català contemporani» et le «Corpus textual informatitzat de la llengua catalana». Brève description du projet et état des travaux*. «Actes du XX Congrès International de Linguistique et Philologie Romanes, Université de Zurich (6-11 avril 1992)», IV, Tübingen-Basel, 1993, págs. 816-821. *Un corpus general de referència de la llengua catalana*. «Caplletra» 17, Valencia, 1994, págs. 219-250. *El Corpus textual informatitzat de la llengua catalana*. (Diccionari del català contemporani). «Quaderni Ibero-americani», 77, Roma, 1995, págs. 69-83. *The CTILC-DCC. A reference corpus for modern Catalan*. «Computers and the Humanities» [en col. con J. Soler] (en prensa). Por otra parte, en las diferentes memorias de actividades del Institut d'Estudis Catalans se puede hallar información sobre el estudio previo, la aprobación, el inicio y el desarrollo del proyecto desde 1983 hasta ahora.

⁸ Datos actualizados en diciembre de 1995.

siguientes: los textos tomados en consideración son siempre textos publicados, cualquiera que sea su forma (libro, folleto, pliego u hoja suelta, periódico, revista); las únicas excepciones que se han hecho a la condición de publicado son las cartas personales y las escrituras notariales, por razones obvias; los textos seleccionados se incorporan íntegramente al sistema informático (sólo en casos realmente excepcionales, en obras de una extensión excesiva para el equilibrio del corpus, se ha seleccionado una parte claramente delimitada); el texto fuente es siempre la primera edición de la obra, la cual se utilizará como referencia definitiva; el contenido de las obras se incorpora sin ninguna corrección ni modificación, conservando incluso el formato material del original, con lo cual se mantiene en el corpus la diversidad de formas, e incluso de grafías, que se encuentran especialmente en la época anterior a la regularización ortográfica del catalán (1913).

Selección de los textos

En el momento de llevar a cabo la selección de los textos, es decir, a la hora de decidir qué textos concretos iban a constituir el corpus, se tuvo como máximo objetivo lograr el mayor grado de representatividad posible, de acuerdo con los principios generales a que he aludido; la representatividad se ha entendido en sentido cronológico y en sentido tipológico a la vez; estas dos dimensiones han orientado en todo momento la selección, y, para ello, paralelamente a los distintos subtipos de lengua considerados, se establecieron unos grupos cronológicos (de diez años en la parte más antigua, hasta 1913, y de cinco años cada uno en el resto, de 1914 a 1988). Durante el proceso de selección de los textos se ha intentado que dentro de cada uno de los veintitrés grupos cronológicos que resultan de esta división se halle representado cada uno de los subtipos de lengua establecidos, de una manera aproximadamente relativa al volumen total de texto publicado en dicho período. Hay que decir todavía que, por lo que se refiere a la lengua no literaria, la mayor parte de las áreas temáticas han sido divididas a su vez en un número variable de subáreas, hasta un máximo de diez, y se han seleccionado textos de cada una de ellas; en la selección de los textos literarios, sin necesidad de subdividir los géneros de una manera sistemática, se han tenido en cuenta diferencias de carácter temático o de otra naturaleza (por ejemplo, narrativa imaginativa, memorias o libros de viajes; teatro costumbrista o clásico; poesía religiosa o profana, etc.). Con estas medidas se ha intentado lograr el mayor grado posible de equilibrio entre las más variadas manifestaciones de la lengua escrita. La aplica-

ción sistemática de estos criterios ha dado lugar a la repartición del texto que puede observarse en el cuadro de la figura 1. En las figuras 2 y 3 pueden verse las representaciones gráficas y los porcentajes correspondientes, respectivamente, a la distribución de los géneros, por lo que se refiere a la lengua literaria, y a la distribución de las áreas temáticas, por lo que respecta a los textos no literarios.

El número de obras seleccionadas es de 3.302, que se reparten en 2.296 correspondientes a la lengua no literaria y 1.006 a la lengua literaria.

La lematización

Una característica relevante del CTILC es que se trata de un corpus lematizado. Como es sabido, en una buena parte de los corpus más o menos extensos que existen para las distintas lenguas la unidad de trabajo es la palabra entendida como cadena de caracteres gráficos, con lo cual esas unidades tienen un elevado grado de ambigüedad, porque una misma grafía puede corresponder a diferentes unidades gramaticales o léxicas: la forma gráfica *cap*, en catalán, puede corresponder al sustantivo masculino CAP ('cabeza', 'cabo'), al verbo CABRE ('caber') (3ª persona del singular del presente de indicativo), a la preposición CAP ('hacia') o al adjetivo indefinido CAP ('ningún', 'ninguna').

La información que contiene un corpus sin lematizar sólo es accesible a través de las formas gráficas, con todos los inconvenientes que ello comporta para la mayoría de las consultas que se pueden querer realizar. En el caso del ejemplo citado, a través de la grafía *cap* accederíamos a la vez y de una manera indiscriminada a la información sobre las distintas unidades lingüísticas a que corresponde: obtendríamos mezclados los casos de sustantivo, los de verbo, los de preposición y los de adjetivo; además, por lo que respecta al sustantivo, sólo tendríamos información sobre la forma de singular, y, por lo que se refiere al verbo, sólo sobre una de las formas de su conjugación; para obtener información sobre el sustantivo o sobre el verbo en su conjunto habría que preguntar por cada una de sus formas, con el agravante de que las variantes morfológicas o gráficas tienen un cierto grado de imprevisibilidad.

El CTILC, en cambio, por el hecho de estar lematizado, tiene un nivel de funcionalidad mucho más alto. En él se ha llevado a cabo una operación de análisis lingüístico en virtud de la cual se ha categorizado gramaticalmente cada una de las ocurrencias de cada forma gráfica y se ha asociado a una unidad léxica de referencia. Con ello se

cubren dos objetivos: por una parte se desambiguan los casos como el del ejemplo citado u otros parecidos, que son muy numerosos; por otra parte, como consecuencia de la misma operación, se relacionan entre ellas las distintas formas de una serie flexiva, las cuales quedan asociadas a esa unidad de referencia, llamada *lema*. La principal consecuencia práctica de esta operación es que se puede realizar la consulta del corpus a partir de los lemas y obtener así información sobre todas las formas asociadas a cada uno de ellos a partir de una sola consulta. En el caso del ejemplo que nos ha ocupado, podemos interrogar el corpus de una manera selectiva sobre el sustantivo CAP, sobre el verbo CABRE, sobre la preposición CAP, y sobre el adjetivo CAP, cosa que no podríamos hacer en un corpus no lematizado. En el caso de nuestro corpus, a causa del diseño específico y característico que posee, en cuyos detalles no podemos entrar aquí, además de las distintas formas de flexión, propiamente dichas, que corresponden a un mismo lema, recuperamos las distintas variantes formales o gráficas, a las que he hecho referencia, y también las eventuales formas derivadas con elementos apreciativos (diminutivos, aumentativos, despectivos) o intensivos, para las cuales, siguiendo criterios de carácter lingüístico, no se ha creado un lema propio. En este conjunto de información, a causa del marcaje específico a que son sometidas durante el proceso de lematización, aparecen la ocurrencias que corresponden a usos citativos, diferenciadas con un código especial (el código MET = usado metalingüísticamente) que impide que se confundan con los usos normales de la misma forma.

Interrogación del corpus

Toda la información generada durante el proceso, tanto los datos sobre las obras incorporadas (autor, título, año, tipología), como el texto propiamente dicho y la información gramatical originada durante el proceso de lematización, son debidamente almacenados en forma de base de datos (Base de dades textual de la llengua catalana - BDTLC). La BDTLC contiene, pues, toda la información necesaria para la explotación del corpus; para posibilitar esta explotación han sido elaborados una serie de programas que permiten, por una parte, la consulta interactiva del corpus por lo que se refiere a lemas, formas, frecuencias, localizaciones y contextos, y, por otra, en consulta diferida, la obtención de listados de diversa naturaleza.

A través del sistema de consulta interactiva, si preguntamos

por un lema⁹, obtenemos un primer bloque de datos: la frecuencia total del lema en el corpus¹⁰, las formas que tiene asociadas (grafía y código morfológico) y la frecuencia de cada una de ellas. Veamos a título de ejemplo la información que obtenemos si preguntamos por el adjetivo invariable GRAN; en el cuadro de la figura 4 vemos la relación completa de formas que en este momento nos ofrece el corpus a partir de este lema, con la frecuencia de cada una de ellas: las abreviaturas G., g., la forma de singular *gran* con diversas variantes gráficas (*grán*, *grân*, *grant*), la forma de plural *grans* con sus variantes gráficas (*gran*, *grands*, *grâns*, *grants*), una larga lista de derivados apreciativos o intensivos (*regran*, *grandassot*, *grandàs*, *grandet*, *grandíssim*, *grandot*, *grandassàs*, *grandolàs*, *grandissó*) con sus variantes morfológicas (femenino, plural) y una serie de variantes gráficas no normalizadas que aparecen en los textos; al final de la lista pueden observarse los usos metalingüísticos a que me he referido más arriba. El sistema de consulta nos permite elegir una de esas formas y observar las distintas ocurrencias que presenta en el corpus a través de sus referencias; la figura 5 nos muestra, como ejemplo, la imagen que aparece en la pantalla después de haber preguntado por la forma *gran*, donde vemos el inicio de la lista de las 17.134 ocurrencias de esta forma. A partir de aquí podemos optar por solicitar el contexto de todas las ocurrencias de la forma en cuestión, de una parte, o de una sola de ellas. La figura 6 nos muestra la pantalla donde solicitamos el contexto correspondiente a la ocurrencia de la obra referenciada con el número 3, página 123, línea 5, palabra número 7; y en la figura 7 aparece dicho contexto con una información complementaria en la cabecera (autor, título y año de la obra a que pertenece, tipo de texto, forma, lema, frecuencia de la forma, frecuencia del lema, etc.). Si el contexto ofrecido en primera instancia (tres líneas del original) no es suficiente para la

⁹ Además de la consulta a partir de un lema, también se puede interrogar el corpus a partir de una forma determinada o a través de una localización (por ejemplo, la palabra número 3 de la línea 7 de la página 125 de la obra 2136); pero estas modalidades de consulta no son habituales para un usuario externo al propio proyecto.

¹⁰ La consulta puede realizarse sobre la base de todo el corpus, pero también existe la posibilidad de que el usuario defina un subcorpus a partir de parámetros cronológicos y tipológicos; es decir, puede restringir la consulta a un período de tiempo determinado y a un sector tipológico más o menos amplio (lengua literaria, teatro, ciencias puras y aplicadas, etc.), o a una sola obra o a un conjunto de obras determinado. Además, ante una forma con una frecuencia excesivamente elevada para poder consultar interactivamente todas sus ocurrencias, existe la posibilidad de consultar un número más reducido de ellas (un número determinado o un tanto por ciento del total, establecidos por el usuario); en este caso, las ocurrencias concretas son determinadas por un programa a base de una selección aleatoria.

finalidad que nos proponemos, podemos solicitar el párrafo completo en que se encuentra la ocurrencia que nos ocupa (figura 8).

Mediante el sistema de consulta diferida podemos obtener distintos tipos de listados: de lemas, de lemas y formas, de todo el corpus o de una parte del corpus (de una obra o de un conjunto de obras), con información sobre la frecuencia de aparición. La naturaleza y el formato de estos listados puede ser muy diversa; adjuntamos unas muestras de ellos a fin de que el lector pueda tener una idea de alguna de las posibilidades que existen (figuras 9, 10 y 11).

En el conjunto de los programas preparados hasta ahora para la consulta y explotación del corpus, quizás sea interesante destacar que se ha incorporado un procedimiento para obtener datos sobre la repartición de las distintas ocurrencias de un lema entre los distintos tipos de texto, es decir, entre diversas partes del corpus. Sabido es que, aunque la frecuencia absoluta (es decir, el número de veces que aparece una palabra en un corpus) es el indicio más característico del peso que tiene ésta en la lengua, es decir, de su valor desde el punto de vista estadístico, dicho valor puede ser matizado a partir del conocimiento de la distribución de las distintas ocurrencias en los distintos grupos en que puede ser dividido el corpus; a igual frecuencia, se considera más básico o fundamental para la lengua general la palabra que está más uniformemente repartida; la que presenta una repartición irregular, en cambio, tiene una consideración secundaria, que es tanto mayor cuanto más irregular es la repartición. El programa aplicado calcula un coeficiente de dispersión con valores extremos entre 0 y 1, y multiplica este coeficiente por la frecuencia para obtener una frecuencia modificada en función de la repartición, que se denomina *uso*¹¹. En la figura 12 vemos una muestra de esta información tal como nos la presenta la pantalla del ordenador tras la consulta realizada a partir de un lema. En las figuras 13, 14 y 15 podemos ver una determinada representación gráfica de la distribución para varios lemas, con el valor de la frecuencia calculado relativamente al contingente de cada grupo, de tal modo que para una palabra reparti-

¹¹ Este procedimiento se ha aplicado por ahora solamente a los subgrupos correspondientes a la lengua no literaria, y no a la parte correspondiente a la lengua literaria, que, como he dicho, no está totalmente lematizada; por otra parte, el sistema está inspirado en el que utiliza Juilland en su serie de diccionarios de frecuencias sobre el español (1964), el rumano (1965), el francés (1970) y el italiano (1973), adaptado a un corpus como el nuestro en que cada uno de los grupos tipológicos tiene un contingente de ocurrencias distinto como consecuencia del criterio de representatividad que hemos aplicado a la hora de seleccionar las obras y establecer la cantidad de texto correspondiente a cada uno de estos grupos.

da de un modo absolutamente uniforme, la gráfica sería horizontal con el valor 1. Por último, las figuras 16, 17 y 18 nos muestran fragmentos de listados de lemas del subcorpus no literario con los valores de la frecuencia, de la dispersión y del uso; clasificados por frecuencia, clasificados por uso y clasificados por dispersión, respectivamente.

Estado actual del corpus y futuro inmediato

Como ya he dado a entender, la constitución del CTILC se halla en este momento en un estado muy avanzado; la parte del corpus que se refiere a la lengua no literaria está totalmente terminada y se ha incorporado íntegramente a la base de datos consultable. Por lo que respecta a la parte correspondiente a la lengua literaria, los textos seleccionados están todos introducidos en el sistema informático y verificados; una parte importante de ellos (10.878.813 ocurrencias) está ya lematizada y podemos prever que a final de 1996 se habrá culminado la lematización del resto (12.129.499 ocurrencias) y el corpus estará totalmente terminado y consultable.

Ello no es, sin embargo, sino el final de una etapa. No debemos olvidar que la constitución de este corpus tiene como objetivo fundamental la redacción de un diccionario descriptivo de la lengua catalana escrita durante los últimos 150 años¹²; en consecuencia, durante los próximos dos años se llevarán a cabo los trabajos preparatorios para la redacción de este diccionario. La labor prevista para este período puede resumirse en tres objetivos específicos¹³.

¹² Al margen de este objetivo fundamental, que es lo que motivó el proyecto, el CTILC puede ser utilizado como fuente no sólo para la realización de otros diccionarios, sino de los más diversos trabajos de investigación sobre la lengua. En el momento de proyectarlo y durante su realización se ha tenido muy en cuenta no cerrar horizontes ni limitar posibilidades, lo cual ha implicado un trabajo mucho más concienzudo, y costoso al fin y al cabo, que si se hubiera orientado única y exclusivamente a la elaboración de la obra prevista; la magnitud de la empresa y el volumen de los recursos necesarios para llevarla a buen puerto aconsejaban la aplicación de unos criterios que permitieran su utilización para las finalidades más diversas dentro de la investigación lingüística. Hasta este momento el corpus, incluso antes de su completa terminación, ha sido utilizado en la elaboración de varias tesis doctorales, tanto en el ámbito estricto de la lingüística, como en el de la estadística.

¹³ Paralelamente a la terminación del corpus textual y a estos trabajos previos a la redacción del diccionario, se constituirá durante los dos años próximos lo que hemos llamado un corpus lexicográfico: una base de datos destinada a contener los principales diccionarios catalanes aparecidos en los siglos XIX y XX. Aunque la fuente principal del nuevo diccionario descriptivo sean los datos obtenidos del CTILC, en el momento de describir la lengua de este período no puede olvidarse dar cuenta de aque-

El primero de ellos se refiere a la elaboración y puesta a punto de un procedimiento informático para la categorización semántica del corpus, puesto que el marcaje que se lleva a cabo durante el proceso de lematización es meramente morfosintáctico y conviene que sea así por razones de carácter metodológico; pero a la hora de redactar el diccionario hay que analizar los distintos contextos en que aparece una palabra y clasificar cada una de sus ocurrencias de acuerdo con el significado que le corresponde. Así, pues, siguiendo con el ejemplo de que me he valido más arriba, si durante el proceso de lematización hemos distinguido entre aquellas ocurrencias de CAP que corresponden a un sustantivo, a un verbo, a una preposición y a un adjetivo, en esta segunda fase del análisis nos será dado establecer los distintos significados del sustantivo (los previsibles, como 'cabeza', 'cabo', 'jefe', y aquellos que el análisis detenido de los contextos nos depare), y así del verbo, del adjetivo y de la preposición, a la vez que podemos conocer las distintas estructuras sintácticas de que forman parte. Dentro del mismo procedimiento se llevará a cabo también la detección de las coocurrencias, es decir, de los elementos léxicos con los que se combina más frecuentemente una palabra determinada.

El segundo de estos objetivos consiste en el estudio y en la fijación de los criterios lexicográficos que se aplicarán durante la elaboración de la obra, los cuales deberán reflejarse ya en la redacción del proyecto de diccionario propiamente dicho.

El tercer objetivo hace referencia a la creación de la estructura material y lógica que, en forma de estación de trabajo lexicográfica, permita la elaboración de los artículos del diccionario sobre soporte informático, utilizando directamente los materiales de las bases de datos del corpus textual y del corpus lexicográfico. El resultado de este trabajo, aparte de dar lugar a un diccionario publicable en forma de libro, ha de originar una nueva base de datos que contendrá el diccionario en forma consultable por pantalla.

Una vez terminado todo este proceso, previsiblemente a principio de 1998, estaremos en condiciones de iniciar la redacción del diccionario descriptivo previsto en el inicio de este proyecto.

llo que dicen de ella los diccionarios coetáneos; los datos lexicográficos que contienen deben tener el lugar adecuado en un diccionario descriptivo, aunque sea como información complementaria. La creación y la utilización de esta base de datos de diccionarios debe entenderse no como una mera ayuda a la redacción de un nuevo diccionario, como sucede a menudo, sino como un modo de dar cuenta documentadamente en una obra lexicográfica descriptiva del tratamiento otorgado a la descripción de los significados de las palabras y a otros fenómenos léxicos en los diccionarios de la época.

El diccionario basado en un corpus

Consideraciones generales

Como he avanzado ya al principio de mi intervención, existe una diferencia fundamental entre la utilización de datos textuales como complemento de una labor lexicográfica a partir de la información proporcionada por los diccionarios preexistentes, y la utilización de este tipo de datos como fuente principal para la elaboración de un diccionario. En el primer caso el lexicógrafo organiza el diccionario a partir de la información que encuentra en los mejores diccionarios que le han precedido (nomenclatura, organización de los artículos, etc.) y recurre a los textos fundamentalmente para documentar, o simplemente para ejemplificar, los significados o los usos que ha establecido a partir de aquellos criterios; de este modo, el recurso al corpus textual es un mero complemento de la actividad lexicográfica, que, en el mejor de los casos, permite no llevar la dependencia hasta el extremo de reproducir los ejemplos de los otros diccionarios, y que en muchas ocasiones puede dar lugar a una mejora indudable con la aportación de usos nuevos, pero no deja por ello de tener un valor secundario, en todo caso añadido¹⁴.

Si se plantea, en cambio, el empleo de un corpus textual como fuente principal, es decir, como base fundamental o punto de partida para la elaboración del diccionario, este elemento pasa a tener una función preponderante. Entendida así esta cuestión, la naturaleza de los datos textuales puede llegar a determinar la estructura de la nueva obra, tanto en sus aspectos generales o globales, como en una serie de cuestiones más particulares, que afectan la organización interna de los artículos. Un corpus que cumple las condiciones de representatividad a que me he referido más arriba, nos depara un conjunto de consta-

¹⁴ Incluso en un diccionario como el *Trésor de la Langue Française*, primer diccionario que se basa en un corpus textual en soporte magnético, puede pensarse que la utilización de los datos ha sido banal: «Il y a naturellement une relation privilégiée entre une base de données textuelles, littéraires pour la plupart, et un dictionnaire général de langue. Une preuve en est que FRANTEXT, première base de ce type, a été réalisée par et pour les rédacteurs du *Trésor de la Langue Française (TLF)*. Ceux-ci l'ont utilisée, comme on sait, pour illustrer autrement les sens consacrés par la tradition lexicographique, mais aussi, et peut-être surtout, pour faire émerger de nouvelles acceptions, attestées dans le discours, mais non encore enregistrées. On objectera que c'est là l'exploitation banale de n'importe quel fichier d'auteur de dictionnaire.» (Éveline Martin; *Les mille et une ressources d'une base de données textuelles pour le lexicographe*, «Lexique», 12-13, 1995, pág. 451.)

taciones sobre la realidad empírica de la lengua que difícilmente podemos adquirir de otro modo. La manera de utilizar esta información de forma operativa es una cuestión en buena parte nueva todavía en la técnica lexicográfica, pero lo que parece evidente es que hay que tenerla en cuenta a la hora de definir la estructura interna del nuevo diccionario. Si se adopta este modo de proceder, se puede elaborar un producto que difiera en muchos aspectos de todos los diccionarios existentes de la misma lengua.

Determinación de la macroestructura

Una de las cuestiones más relevantes al analizar la naturaleza de un corpus es su contenido desde el punto de vista del léxico. Es evidente que la constitución de un corpus pone de relieve la utilización de las unidades léxicas en los actos comunicativos; el corpus nos muestra no solamente los distintos usos y valores que la realidad de la comunicación lingüística hace de cada elemento léxico, sino que también nos revela qué palabras son realmente usadas y cuáles no se utilizan o se utilizan menos¹⁵. A partir del análisis de un corpus nos percatamos de que ciertas palabras que se encuentran en la mayoría de los diccionarios de la lengua no se hallan ordinariamente en los textos, y, por otra parte, que en estos textos aparecen numerosas palabras que no han sido recogidas en ningún diccionario. Podríamos decir, por tanto, que la lengua escrita habitualmente se aparta en mayor o menor medida de aquello que los diccionarios contienen, o, dicho de otro modo, que los diccionarios no son un reflejo de la naturaleza de la lengua utilizada en la realidad. Nos encontramos, pues, con la aparente paradoja de que el léxico contenido en un corpus es a la vez más rico y más pobre que el que contiene un diccionario; lo mismo podemos decir del léxico del diccionario con respecto al léxico del corpus. Los dos conjuntos de unidades léxicas, el del corpus y el del diccionario, se hallan en relación de intersección.

No podemos entrar aquí en el análisis detallado de la naturaleza interna de esta desviación, pero lo que he dicho hasta ahora creo que es suficiente para plantearse la pregunta de hasta qué punto el léxico que aparece en un corpus tiene que determinar la naturaleza de

¹⁵ Para citar sólo un caso concreto, se da la paradoja, por ejemplo, que una palabra que no consta en los diccionarios catalanes más representativos, como *irrepetible*, tiene en el corpus una frecuencia igual a 36, mientras que la palabra de significado opuesto, que sí está en los diccionarios (*repetible*) tiene en el corpus una frecuencia de sólo 12. Casos como éste nos llevarían a plantear si en el establecimiento de la nomenclatura de un diccionario debe prevalecer el sistema o bien el uso.

la nomenclatura de un diccionario basado en este corpus, tanto más si tenemos en cuenta que desde el punto de vista cuantitativo, por lo menos a partir de los datos que tenemos para el catalán, las desviaciones tanto en un sentido como en otro no son nada despreciables.

En efecto, una simple comparación de los lemas que contiene el corpus y los lemas del diccionario de base¹⁶ nos permite saber que aquél contiene 129.007 lemas y éste contiene 85.337; ahora bien, los lemas comunes al corpus y al diccionario de base son sólo 54.970, lo cual quiere decir que el diccionario contiene 30.403 lemas que no aparecen en el corpus, y que, por otro lado, el corpus contiene 74.037 lemas que no aparecen en el diccionario; es decir, sólo el 64.39% de los lemas del diccionario aparecen también en el corpus, y sólo el 42.62% de los lemas del corpus aparecen en el diccionario¹⁷. Como vemos, estas cifras no son banales.

A partir de estos datos podríamos caer en la tentación de hacernos las preguntas siguientes: ¿Hemos de eliminar del diccionario todos aquellos lemas que no aparecen en el corpus? ¿Hemos de añadir al diccionario los lemas del corpus que no se encuentran en él? Sería, sin embargo, temerario intentar contestar a estas preguntas con una simple afirmación o negación. Este fenómeno de intersección entre el conjunto de lemas del diccionario y el conjunto de lemas del corpus, que deja sin correspondencia en éste una buena parte de los lemas del diccionario, y sin correspondencia en el diccionario más de la mitad de los lemas del corpus, es susceptible de ser considerado en su conjunto, como acabamos de hacer, pero de esta consideración general no podemos extraer conclusiones que nos permitan adoptar decisiones concretas. En cambio, el estudio más pormenorizado, por grupos

¹⁶ La naturaleza del procedimiento de lematización utilizado, de carácter semiautomatizado, implica la existencia de un diccionario electrónico o diccionario de máquina denominado Diccionari Bàsic Informatitzat (DBI), que se constituyó a partir de las entradas de los dos diccionarios más representativos y de uso general (el *Diccionari general de la llengua catalana* de Pompeu Fabra y el *Diccionari de la llengua catalana* de Enciclopèdia Catalana). Como resultado de la interacción entre el corpus y el DBI a lo largo del proceso de lematización, éste contiene en la actualidad, debidamente codificados, por una parte, los lemas presentes en los diccionarios citados, y, por otra parte, los lemas nuevos que han ido apareciendo en el corpus durante todo este proceso; respecto de los primeros queda constancia de aquellos que han aparecido en el corpus en algún momento y de aquellos que no han surgido hasta ahora en ningún texto. Esta circunstancia nos permite comparar el conjunto de los lemas de los diccionarios con el conjunto de los lemas del corpus y sacar conclusiones tanto de carácter cuantitativo como de carácter cualitativo.

¹⁷ Aunque estos datos pueden ser matizados antes de la terminación completa del corpus, su estado actual, que presenta más de 40 millones de ocurrencias lematizadas, nos permite darlos como representativos.

homogéneos de palabras, parece que puede arrojar una cierta luz sobre esta cuestión, como se desprende de un análisis tentativo que he realizado recientemente a propósito de los derivados en *-ble*, de los adverbios en *-ment* y de una serie de prefijos o formas prefijadas cultas¹⁸; desde el punto de vista cuantitativo, este tipo de palabras presentan una situación comparable a la que nos ofrecen los datos globales que acabamos de comentar y que se resume en una relación de intersección entre el conjunto de los lemas que aparecen en el corpus y el de los lemas del diccionario de base. El cuadro de la figura 19 nos ofrece de una manera comparativa los datos que se refieren al conjunto del corpus y al conjunto del diccionario, los que se refieren a los adjetivos formados con el sufijo *-ble*, y los correspondientes a los adverbios en *-ment*; en la figura 20 encontramos los datos relativos a las palabras formadas con veinte prefijos o formas prefijadas cultas¹⁹. La naturaleza de las palabras afectadas y los fenómenos que representan reclaman, sin embargo, un análisis detenido de esta cuestión en toda su complejidad, que no podemos abordar aquí. Lo que sí creo que se puede avanzar es que el resultado de este análisis ha de contribuir indudablemente a mejorar los diccionarios desde el punto de vista de la nomenclatura, es decir, de las palabras que contienen y de las que no contienen, eliminando quizás aquellas que no tienen ningún reflejo en la realidad de los textos -que son el resultado de la reproducción indiscriminada de diccionario en diccionario desde tiempo inmemorable-

¹⁸ J. Rafel; *Le réel et le virtuel dans le lexique. Le rôle des corpus textuels*. (en prensa en las actas del «XXI Congrès International de Linguistique et Philologie Romane», celebrado en Palermo del 18 al 24 de setiembre de 1995).

¹⁹ No podemos entrar aquí en el análisis de la naturaleza cualitativa de estos fenómenos. Voy a citar solamente unos ejemplos relativos a uno de los prefijos cultos que aparecen en el cuadro de la figura 19: el prefijo *bio-*, que desde el punto de vista numérico presenta una situación equilibrada entre los dos extremos, el de mayor y el de menor correspondencia entre el diccionario y el corpus. Entre las palabras exclusivas del corpus (ausentes en el diccionario de base), que son un total de 91, encontramos *bioclimàtic* [f. 3], *biocombustible* [f. 1], *biodinàmic* [f. 7], *bioelèctric* [f. 4], *bioestadística* [f. 3], *bioindicador* [f. 1], *bioinsecticida* [f. 4], *biomèdic* [f. 6], *biomatemàtic* [f. 4], *biomolecular* [f. 1], *biopolímer* [f. 4], *biopsíquic* [f. 5], *biosensor* [f. 5], *biotecnologia* [f. 196], *biotecnològic* [f. 52], *biotò* [f. 2]; entre las palabras del diccionario que no se encuentran en el corpus, con un total de 57, encontramos casos como *bioclima*, *bioconversió*, *biodeterioració*, *biodinamisme*, *bioflavonoide*, *biofotogènesi*, *bioherm*, *bioma*, *biòmetra*, *biometria*, *bioreacció*, *bioritme*; finalmente, entre las palabras comunes al corpus y al diccionario, con un total de 42, encontramos las que son quizás más corrientes: *biodegradable* [f. 13], *biofísica* [f. 8], *biogènesi* [f. 9], *biòleg* [f. 159], *biologia* [f. 480], *biològic* [f. 1.280], *biomassa* [f. 106], *biòpsia* [f. 21], *bioquímic* [f. 70], *bioquímica* [f. 53], *biosfera* [f. 104], *biosíntesi* [f. 14], *biòtop* [f. 65].

rial, o de la adición acrítica- y añadiendo por lo menos una parte de las que vienen atestiguadas en la lengua escrita²⁰.

Aparte de aquellos aspectos que se refieren estrictamente al establecimiento de la nomenclatura, el resultado del análisis detallado de toda esta problemática puede afectar otras cuestiones de la macroestructura diccionarial, como la disposición de los artículos, la posibilidad de establecer más de un nivel en la nomenclatura, o una subnomenclatura en el interior de los artículos. Todos estos elementos contribuyen, por supuesto, a configurar la estructura del diccionario de acuerdo con los datos que el corpus nos ofrece.

Naturaleza de la microestructura

El influjo del corpus en la estructura del diccionario no termina, sin embargo, con las cuestiones relativas a la macroestructura -éstas no constituyen sino una pequeña parte de su contribución-, sino que puede determinar muchos aspectos de la naturaleza interna de los artículos, tanto desde el punto de vista del contenido propiamente dicho, como de la organización de este contenido. Todo ello es consecuencia de que en el momento de redactar cualquier artículo del diccionario el lexicógrafo o el equipo de lexicógrafos tiene a su alcance una información objetiva y coherente sobre el uso de aquel elemento léxico; por el carácter representativo del corpus, esta información puede considerarse además exhaustiva en el sentido que contiene todas las opciones de las unidades léxicas que tienen un reflejo en la realidad del uso; es decir, en este caso el lexicógrafo posee unos datos más seguros y más objetivos que si se propusiera redactar un diccionario no basado en corpus.

Por lo que se refiere a la determinación de los significados de una unidad léxica, se puede partir así del análisis de los distintos contextos en que aparece y llegar a establecer su red de significaciones, en claro contraste con los diccionarios no basados en corpus, que deben recurrir necesariamente a la tradición lexicográfica precedente y, en

²⁰ Los autores del *Lessico di frequenza della lingua italiana contemporanea*, al ponderar la utilidad de los diccionarios de frecuencia para estudiar el léxico de una lengua, presentan también un panorama muy negativo por lo que respecta al contenido de los diccionarios habituales: «Con esso ci proponiamo di dare un utile strumento di lavoro per lo studio del lessico contemporaneo che, il più delle volte, nei dizionari normativi viene necessariamente quasi sommerso nel *mare magnum* di voci e forme tradizionali o già uscite dall'uso, o in via di sparizione, o limitate a lessici speciali o, anche se tuttora vitali, fornite di accezioni particolari.» (Bortolini, U., Tagliavini, C., Zampolli, A.; *Lessico di frequenza della lingua italiana contemporanea*, Garzanti, Milán, 1972, pág. 5)

todo caso, a la competencia que el mismo lexicógrafo tiene de la lengua en la que trabaja o al vaciado selectivo de textos. Este procedimiento puede llegar a dar resultados que transforman completamente lo que sería el artículo de un diccionario no basado en corpus, porque el análisis de los contextos aporta elementos que permiten establecer valores significativos que no se encuentran habitualmente en los diccionarios, desde los casos más obvios y claros hasta los matices más difíciles de establecer.

La información frecuencial inherente al concepto de corpus tiene también un valor operativo en la elaboración de un diccionario, aparte del valor meramente informativo que obviamente posee; en efecto, el diccionario puede dar en cualquier momento la frecuencia, sea de la unidad léxica globalmente considerada -el lema-, o de cualquiera de los elementos en que éste puede desglosarse con criterios morfológicos (por ejemplo, formas de singular y de plural, de masculino y de femenino, etc.) o bien por criterios semánticos (frecuencia de aparición de cada uno de los principales valores significativos establecidos) o con criterios combinados. Como sea que también tenemos a mano información sobre las estructuras sintácticas en que puede aparecer aquella unidad o sobre las combinaciones más habituales con otros elementos léxicos, también puede ofrecerse información sobre la frecuencia de este tipo de estructuras y de combinaciones. Pero el conocimiento de estos valores nos permite también utilizar criterios frecuenciales para la ordenación de los significados o de los artículos de entradas homónimas, si se opta por esta solución a la hora de determinar los criterios metodológicos que van a regir la elaboración del diccionario²¹.

Otra de las grandes aportaciones de la utilización de un corpus como fuente principal de un diccionario es justamente la información que podemos obtener sobre estructuras sintácticas y sobre coocurrencias léxicas, a que acabo de hacer referencia. En efecto, los diccionarios no basados en corpus no contienen una información sistemática sobre estas cuestiones, que el lector interesado debe rastrear en los ejemplos o en las construcciones que por su naturaleza se han

²¹ Diversos autores se han referido en más de una ocasión a la posibilidad o a la conveniencia de ordenar las acepciones de una entrada atendiendo a la frecuencia de cada una de ellas. Veámos sólo un testimonio de ello: «... comme les mots considérés dans leur fonctionnement actuel n'héritent que d'une partie de leurs valeurs d'emploi anciennes, c'est plutôt en fonction des scènes (sic) ou de la fréquence des emplois qu'il convient de classer les sens, l'échelle pouvant être ascendante (de rare fréquence à grande fréquence) ou descendante.» (R. L. Wagner; *Réflexions naïves à propos des dictionnaires*, «Cahiers de Lexicologie», 27, 2, 1975, pág. 87.

tratado como subentradas o como frases ejemplo, pero en cualquier caso la información que obtiene está muy lejos de ser un reflejo de la realidad del uso, sea porque es insuficiente, sea porque es inexacta o inadecuada, sea porque en ciertos casos no existe ninguna. No podemos entretenernos aquí en la ilustración de cada uno de los aspectos que vamos mencionando, puesto que no disponemos ni del tiempo ni del espacio suficiente. Sólo voy a poner un ejemplo muy simple de lo que quiero decir: el sustantivo femenino *calor*, en el sentido de «sensación producida por una elevada temperatura ambiental o corporal», se combina habitualmente en catalán -de una manera parecida a lo que sucede en castellano- con un reducido número de verbos, con ausencia o presencia controlada del artículo, dando lugar a expresiones que, aunque muy utilizadas, no son de interpretación obvia: *fer calor*, *tenir calor*, *passar calor*, *suportar la calor*, *entrar en calor*; por otra parte, la mayoría de los adjetivos que se combinan con esta unidad, o los que lo hacen con mayor frecuencia, son de carácter magnificador o indican cualidades no deseables: *calor forta*, *terrible*, *intensa*, *extraordinària*, *insuportable*; existe la expresión *fa una calor que empudega* (literalmente, «hace un calor que apesta») para referirse a un calor insoportable por lo intenso, y la frase de sentido figurado *no em (li, et, etc.) fa ni fred ni calor* con el significado de «me (le, te, etc.) da igual», «no me (le, te, etc.) importa», muy habitual en la lengua hablada. Pues bien, aunque pueda sorprender, ninguna de estas informaciones aparece en ninguno de los diccionarios catalanes a que puede recurrir un usuario; el *Diccionari general de la llengua catalana*, de Pompeu Fabra, para ese sentido trae sólo los ejemplos *Tenir, sentir calor. Les calors de l'estiu. Corrent he agafat calor*; el *Diccionari de la llengua catalana* publicado por Enciclopèdia Catalana, no trae ningún ejemplo para esta acepción y, en cambio, introduce *fer calor* como subentrada, entendiendo que no se refiere exactamente a la sensación, sino al fenómeno atmosférico. Esta es toda la información que encontramos en los dos diccionarios catalanes más representativos²².

²² Tampoco encontramos mayor información sobre estos aspectos de la palabra catalana *calor* en el *Diccionari Català Valencià Balear* de A. M. Alcover y F. de B. Moll (10 volúmenes publicados entre 1926 y 1962), que tiene una pretensión descriptiva, histórica y dialectal. Si echamos una ojeada a la situación del español para la palabra correspondiente, tenemos que recurrir al diccionario de María Moliner, que se define como «diccionario de uso» para encontrar un ejemplo de «hacer calor»: *Hoy hace más calor que ayer*, una referencia a *dar, sentir y tener*, y la subentrada *entrar en calor*. Nada de ello aparece, por ejemplo, en el diccionario de Casares. En el diccionario de la Real Academia Española no vemos ningún ejemplo que ilustre los usos habituales de esta palabra; encontramos, eso sí, la subentrada *entrar en calor* y otras expresiones

Un diccionario que pueda dar el tipo de información a que me refiero, porque parte de una fuente que la proporciona de una manera sistemática, se acerca más a lo que podríamos entender como descripción de la lengua que la mayoría de los diccionarios convencionales; éstos, en muchos casos, tienden más a dar información sobre los fenómenos de la realidad en general que a describir la naturaleza o la estructura de la lengua, es decir, la manera de configurar mentalmente esta realidad y de expresarla.

Una de las posibilidades que nos permiten los materiales del CTILC es la de utilizar información sobre los derivados apreciativos (diminutivos, aumentativos, despectivos) o intensivos, de los que el catalán -como el español o el italiano- hace un uso más o menos profuso²³. En efecto, la organización de la base de datos de nuestro corpus nos permite obtener fácilmente información sobre todos los derivados de esta naturaleza que un lema determinado presenta en el corpus, con la frecuencia de aparición de cada uno de ellos. Es importante poder dar cuenta de este tipo de información basada en la realidad del uso, puesto que el tratamiento de estas formaciones, que normalmente es confiado a las gramáticas, suele ser muy insuficiente; en efecto, muchas de ellas no hacen ninguna referencia a este tipo de derivados o tratan de ellos de una manera imprecisa o incompleta. Los diccionarios, por otra parte, no suelen acoger en sus páginas este tipo de palabras (salvo en los casos de lexicalización o de formaciones excepcionales), puesto que no constituyen unidades léxicas con un significado conceptual distinto del de la palabra primitiva corres-

como *al calor de*, *ahogarse uno de calor*, *asarse uno de calor*, *coger calor*, *dar calor*, *dejarse caer el calor*, etc.

²³ Julio Casares, en su obra *Introducción a la lexicografía moderna*, se ocupa de la necesidad de que el diccionario aporte algún tipo de información sobre estos elementos: «Por lo que se refiere a los aumentativos y diminutivos, es tal su variedad y abundancia en español, que si se registraran todos en el Diccionario ocuparían un espacio desmedido. Esta razón, muy atendible, es la que tuvo en cuenta la Academia para omitir las formaciones que ofrecen una cierta regularidad, poniendo, en cambio, al final del Diccionario y en la Gramática una serie de normas encaminadas a servir de guía en la elección de los respectivos sufijos; pero, aun sabidas a la perfección estas normas con sus correspondientes excepciones, queda todavía por averiguar si esta o aquella de las formaciones permitidas se ha usado efectivamente en el habla (...) Nuestra opinión es favorable a un temperamento intermedio, que consistirá en no hacer un artículo aparte para cada uno de los aumentativos y diminutivos llamados regulares, salvo cuando la significación tenga un matiz que no sea simplemente la modificación cuantitativa del positivo; pero mencionar, en el artículo dedicado a éste, todas las formaciones documentadas.» (págs. 320-321).

pondiente²⁴. A partir de la información que nos ofrece el corpus, conocemos, por ejemplo que adjetivos como *molt* «mucho», *bell* «bello», *antic* «antiguo», *cert* «cierto», *humà* «humano», *comú* «común», se combinan solamente con el sufijo intensivo *-íssim*, mientras que otros adjetivos, como *dolç* «dulce», *alt* «alto», *just* «justo», *estret* «estrecho», aparte de las formas intensificadas con *-íssim*, presentan testimonios con el diminutivo *-et*; y otros presentan un mayor número de posibilidades: *fort* «fuerte» presenta *fortíssim*, *fortet*, *fortot* y el intensivo prefijado *refort*; *trist* «triste» presenta *tristíssim*, *tristoi*, *tristet*, *tristot*, *tristó*; *poc* «poco» presenta *poquíssim*, *poquet*, *pocotet*, *repoc*; *gras* «gordo» presenta *grassíssim*, *grasset*, *grassó*, *grassonet*, *grassot*. Otros, todavía, nos ofrecen un abanico de posibilidades mucho mayor; veamos, la distinta distribución de este tipo de derivados para *gran* «grande» (*grandíssim*, *grandet*, *grandot*, *grandàs*, *grandassàs*, *grandassot*, *grandolàs*, *regran*) y para *petit* «pequeño» (*petitíssim*, *petitet*, *petitó*, *petitoi*, *petitonet*, *repetit*). Un panorama también muy variado nos ofrecen los sustantivos: *cap* «cabeza» (*caparró*, *capet*, *caparrot*, *caparronet*, *caparri*), *animal* «animal» (*animalet*, *animaló*, *animalic*, *animalàs*), *camí* «camino» (*caminet*, *caminoi*, *caminot*, *caminàs*), *veu* «voz» (*veuetà*, *veuarra*, *veuassa*, *veuota*, *veuerriña*), etc. A partir de estos datos, pues, el diccionario puede reflejar el uso que se ha hecho en el acto comunicativo de las posibilidades que la lengua presenta de una manera genérica.

Final

Después de las consideraciones que acabo de hacer, queda patente, creo, que la naturaleza, la estructura y la fisonomía de un diccionario diseñado y elaborado a partir de los datos proporcionados por un corpus textual representativo de la lengua que se quiere describir puede proporcionar una imagen de ésta muy distinta de la que nos ofrece un diccionario elaborado siguiendo los métodos convencionales, a condición de que los datos del corpus se utilicen realmente de una manera operativa, como los materiales de construcción de la estructura del edificio, si se me permite la imagen, y no simplemente

²⁴ Un caso particular es el *Diccionari Català-Valencià-Balear*, que en algunos artículos contiene un apartado titulado INTENS. (intensivos), donde recoge formas con este tipo de sufijos. Por ejemplo, en el artículo correspondiente a la entrada LLEIG (feo), encontramos *lletget*, *lletjot*, *lletgíssim*, *lletjó*; en el correspondiente a la entrada ESCALFOR (calor, del cuerpo, de una estufa), encontramos *escalforeta*, *escalforassa*, *escalforota*.

como una mera ayuda o complemento, para poner un adorno en algún lugar de la obra, como quien pone la guinda encima del helado.

Un diccionario elaborado a partir de estas premisas ofrece al usuario unas garantías que no puede ofrecer un diccionario tradicional, porque todas las palabras que contiene su nomenclatura, aparte del valor general que les confiere el hecho de pertenecer al sistema lingüístico, tienen el valor particular y concreto que les confiere el testimonio del uso, que puede ser ponderado cuando convenga con criterios cuantitativos. Las distintas significaciones que un diccionario así constituido nos aporta para cada lema tienen una base empírica ausente en los otros diccionarios, y pueden ser también valorados sobre la base de su frecuencia. Las frases que ilustran los significados en un diccionario convencional, los llamados ejemplos, no son propiamente ejemplos en un diccionario así concebido, sino testimonios de su uso real.

La garantía que ofrece un diccionario tradicional es una garantía indirecta -los otros diccionarios-, subjetiva y limitada -el conocimiento que el lexicógrafo tiene de la estructura y funcionamiento de la lengua-; mientras que la garantía de un diccionario basado en un corpus textual es una garantía directa, objetiva y general -el propio corpus.

Por lo que respecta a la lengua catalana, como acabamos de ver, poseemos un corpus suficientemente amplio y equilibrado, que va a permitirnos en breve llevar a cabo un trabajo de esta naturaleza.

TEXTOS LITERARIOS

	SECCIÓN I (1833-1873)	SECCIÓN II (1874-1913)	SECCIÓN III (1914-1988)	TOTAL
1. Ensayo	0	358.075	2.687.251	3.045.326
2. Narrativa	564.416	2.483.523	10.812.144	13.860.083
3. Poesía	263.433	468.032	1.769.984	2.501.449
4. Teatro	378.128	868.338	2.454.988	3.701.454
TOTAL	1.205.997	4.177.968	17.724.367	23.108.312

TEXTOS NO LITERARIOS

	SECCIÓN I (1833-1873)	SECCIÓN II (1874-1913)	SECCIÓN III (1914-1988)	TOTAL
0. Correspondencia	2.597	40.252	90.410	133.259
1. Filosofía	6.886	210.337	1.532.615	1.749.838
2. Religión y teología	235.218	533.475	2.203.350	2.972.043
3. Ciencias sociales	202.995	909.723	4.469.565	5.582.283
4. Prensa	109.686	462.242	3.023.971	3.595.899
5. Ciencias puras y naturales	23.715	372.237	1.828.993	2.224.945
6. Ciencias aplicadas	225.303	573.767	3.674.356	4.473.426
7. Bellas artes. Ocio. Deportes	5.027	416.406	2.386.413	2.807.846
8. Lengua y literatura	136.310	333.840	1.759.025	2.229.175
9. Historia y Geografía	107.855	426.895	2.962.889	3.497.639
TOTAL	1.055.592	4.279.174	23.931.587	29.266.353

Figura 1.

Repartición de los textos literarios

23.108.312 palabras

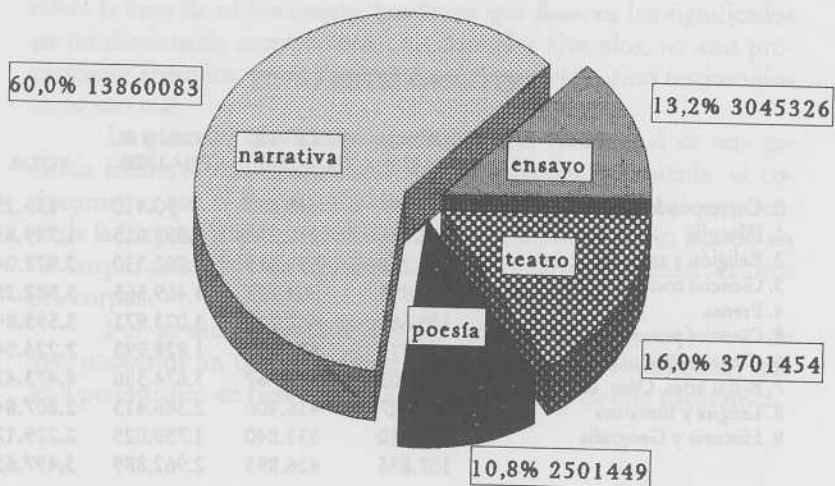


Figura 2.

Repartición de los textos no literarios
29.266.353 palabras

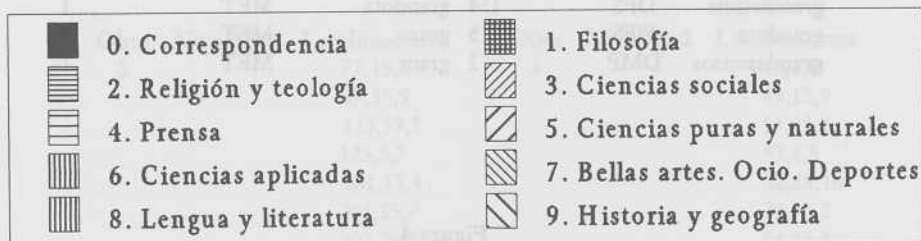
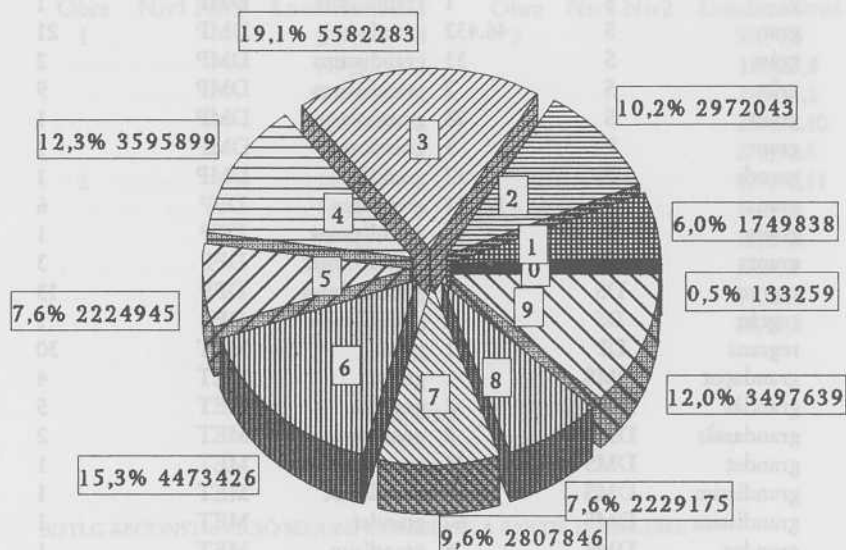


Figura 3.

Lema: GRAN

Cat. gram.: AI

Frecuencia del lema: 64.078

<i>forma</i>	<i>c.m.</i>	<i>frecuencia</i>	<i>forma</i>	<i>c.m.</i>	<i>frecuencia</i>
G.	S	2	grandassos	DMP	4
g.	S	1	grandassots	DMP	1
gran	S	46.432	grandets	DMP	21
grán	S	33	grandíssims	DMP	2
grân	S	2	grandíssims	DMP	9
grant	S	24	grandolassos	DMP	1
gran	P	3	grandots	DMP	1
grands	P	12	grandòts	DMP	1
grans	P	17.134	grandetes	DFP	6
gráns	P	1	grandísimas	DFP	1
grants	P	11	grandíssimas	DFP	3
regran	DS	3	grandíssimes	DFP	13
regrán	DS	2	grandissones	DFP	1
regrans	DP	1	gran	MET	30
grandaçot	DMS	1	grand	MET	4
grandàs	DMS	2	grandàs	MET	5
grandassàs	DMS	2	grandassa	MET	2
grandet	DMS	31	grandassàs	MET	1
grandíssim	DMS	8	grandassot	MET	1
grandíssim	DMS	60	grandet	MET	1
grandot	DMS	8	grandísim	MET	1
grandeta	DFS	17	grandíssim	MET	1
grandísima	DFS	2	grandíssim	MET	3
grandíssima	DFS	5	grandot	MET	2
grandíssima	DFS	154	grandota	MET	1
grandota	DFS	5	grans	MET	3
grandassassos	DMP	2	grant	MET	1

Figura 4.

BDTLC RECONSTRUCCIÓ SEGONS L'ORIGINAL A PARTIR D'UN LEMA

Selecció de les localitzacions de la forma

Forma: grans Codi morf.: P Freq. abs.: 17.134
 Lema: gran Cat. gram.: AI Codi lema: 42.814

Obra	Niv1	Niv2	Localitzacions	Obra	Niv1	Niv2	Localitzacions
1			17,7,5	2			59,34,5
			117,24,1				186,20,8
			127,11,1				265,34,1
			157,11,2				266,29,10
			173,5,9				276,30,5
2			21,21,7	3			10,25,6
			34,31,12				10,26,8
			41,32,3				10,31,8
			55,6,11				11,18,3
			58,11,1				28,18,8
		58,21,1					

Figura 5.

BDTLC RECONSTRUCCIÓ SEGONS L'ORIGINAL A PARTIR D'UN LEMA

Selecció de les localitzacions de la forma

Forma: grans Codi morf.: P Freq. abs.: 17.134
 Lema: gran Cat. gram.: AI Codi lema: 42.814

Obra	Niv1	Niv2	Localitzacions	Obra	Niv1	Niv2	Localitzacions
3			77,15,8	4			45,1,8
			97,35,9				49,15,9
			113,39,5				51,23,4
			123,5,7				53,4,5
			161,13,4				62,28,10
			201,25,7				79,31,2
			202,20,8				84,25,8
4			31,4,5			86,14,7	
			38,21,4			99,27,9	
			41,28,13			106,19,13	
			42,31,6			107,27,10	

Figura 6.

Autor: Joan Fuster

Títol: Diccionari per a ociosos

Obra: 3 N.1: N.2: Localització: 123,5,7

Any d'edició: 1964 Forma: grans

Tipus public.: Assaig C. morf.: P Plural

Total d'ocurrències: 53.136 Lema: gran

Total de lemes usats: 6.431 C. Gram.: AI Adj. de gènere invariable

Total de formes usades: 10.056 Codi lema: 42.814

CONTEXT

▣▣ novel·listes mediocres- en deformacions grosseres. I convé subratllar que si en les grans novel·les (excepte Dickens?) apareixen tan pocs infants, tan escassos ▣▣

Figura 7.

BDTLC RECONSTRUCCIÓ DEL CONTEXT SEGONS L'ORIGINAL

Autor: Joan Fuster

Títol: Diccionari per a ociosos

Obra: 3 N.1: N.2: Localizació: 123,5,7

CONTEXT

✠ ✠ Amb tot, el problema subsisteix, dempeus i invulnerable. I no sols respecte a Dostoievski. El defecte d'adultitzar els personatges sovinteja en els novel·listes de qualsevol època i de qualsevol lloc. Els adolescents, i sobretot els infants, que van i vénen per les obres d'imaginació, solen pecar d'adults, de massa adults en les seves paraules i en els seus gestos. El novel·lista els confereix pensaments i aventures que el lector, en la seva experiència personal -jutjant des de la seva experiència personal-, no gosa reconèixer com a normals en aquelles etapes de la vida. Una de les causes d'aquest desfocament, la més palesa, seria, en el fons, el mateix fet de

BDTLC RECONSTRUCCIÓ DEL CONTEXT SEGONS L'ORIGINAL

Autor: Joan Fuster

Títol: Diccionari per a ociosos

Obra: 3 N.1: N.2: Localització: 123,5,7

CONTEXT

ser l'escriptor un adult, propens, per tant, a crear els seus personatges com una segregació personal, com una rèplica alliberadora: a imatge i semblança seva. És lògic, doncs, que incorri en deformacions, i de vegades -diguem-ho tot: vegades de novel·listes mediocres- en deformacions grosseres. I convé subratllar que si en les grans novel·les (excepte Dickens?) apareixen tan pocs infants, tan escassos adolescents, això es deu, segurament, a una tendència clara a evitar-ne els perills: a evitar el perill de falsificar-los. Pensem de passada com alguns -Proust, en algunes pàgines- l'eviten: per l'evocació, per una evocació a través de l'adult. ✠ ✠

Figura 8.

<i>lema</i>	<i>c.g.</i>	<i>proc.</i>	<i>frequència</i>
a	PO	DFA	714.605
a	F	DFA	7
aalenia	A	DCC	3
aapa	AII	DCC	2
ababol	F	DEC	2
ababolera	F	DEC	1
ab absurdum	LA	DCC	1
àbac	M	DFA	59
abacà	M	DFA	5
abacial	AI	DFA	48
abacinat	A	DCC	2
abaciologi	M	DFA	2
abadal	AI	DFA	1
abadeig	M	DEC	9
abadessa	F	DFA	124
abadia	F	DFA	238
abadiat	M	DFA	14
ab aeterno	LA	DCC	3
abaixador	M	DFA	9
abaixament	M	DFA	45
abaixar	VTP	DFA	352
abaixat	A	DEC	34
abalançament	M	DFA	1
abalançar	VP	DFA	11
abaldufat	A	DEC	2
abalisar	VT	DFA	1
abalisat	A	DCC	1
abalori	M	DCC	1
abaltiment	M	DFA	12
abaltir	VP	DFA	7
abaltit	A	DCC	4
abalustrat	A	DCC	3
abancament	M	DCC	3
abancalar	VT	DFA	1
abancalat	A	DCC	1
abanderament	M	DFA	4
abanderar	VT	DFA	2
abanderat	A	DCC	14
abandó	M	DFA	359
abandonament	M	DFA	148
abandonar	VTP	DFA	2.754
abandonat	A	DFA	643
abandonisme	M	DFA	2

Figura 9. Principio de un índice de lemas clasificados alfabéticamente con indicación de la categoría gramatical, la procedencia y la frecuencia. Entendemos por procedencia si el lema en cuestión figura en alguno de los dos diccionarios que se han tenido en cuenta para constituir el diccionario de máquina básico (codificados DFA y DEC) o en ambos con alguna diferencia en la categorización gramatical (codificados DFE), o bien no figuran en estos diccionarios pero han aparecido en el corpus (codificados DCC).

freqüència		lema	c.g.	proc.
absoluta	relativa			
3.074.587	10,766737	el	AR	DFA
2.026.255	7,095637	de	PO	DFA
920.596	3,223787	i	C	DFA
714.532	2,502182	a	PO	DFA
698.838	2,447224	ell	P	DFA
618.622	2,166320	un	AR	DFA
579.610	2,029706	en	PO	DFA
566.862	1,985064	ésser	VI	DFA
503.704	1,763895	del	CT	DFA
487.325	1,706538	que	P	DFA
411.381	1,440593	per	PO	DFA
364.944	1,277978	que	C	DFA
319.093	1,117415	haver	VIA	DFA
283.261	0,991937	no	AV	DFA
238.958	0,836794	amb	PO	DFA
222.990	0,780877	al	CT	DFA
202.536	0,709250	aquest	A	DFA
163.955	0,574145	com	AV	DFA
156.210	0,547023	seu	A	DFA
149.960	0,525137	més	AV	DFA
148.551	0,520203	fer	VVP	DFA
147.837	0,517702	o	C	DFA
127.758	0,447389	tot	A	DFA
126.635	0,443456	poder	VT	DFA
119.642	0,418968	jo	P	DFA
117.057	0,409915	tenir	VTP	DFA
107.479	0,376375	hi	P	DFA
99.813	0,349530	altre	A	DFA
84.356	0,295401	dir	VVP	DFA
78.986	0,276597	si	C	DFA
76.417	0,267600	però	C	DFA
66.271	0,232070	anar	VA	DFA
61.579	0,215640	qual	AI	DFA
60.609	0,212243	nostre	A	DFA
58.616	0,205264	en	P	DFA
58.570	0,205103	mateix	A	DFA
52.299	0,183143	pel	CT	DFA
51.866	0,181626	any	M	DFA
51.549	0,180516	molt	AV	DFA
51.469	0,180236	donar	VTP	DFA
51.001	0,178597	estar	VVP	DFA
50.375	0,176405	gran	AI	DFA
50.217	0,175852	també	AV	DFA
48.722	0,170617	aquell	A	DFA
46.662	0,163403	dos	AN	DFA
46.219	0,161851	entre	PO	DFA
46.212	0,161827	veure	VTP	DFA

Figura 10. Inicio de un índice de lemas clasificados en orden decreciente de frecuencia.

lema	c.g.	proc.	freqüència	
			absoluta	relativa
solidaritzar	VT	DFA	66	0,000231
insolidaritzar	V	DCC	2	0,000007
dessolidaritzar	V	DCC	2	0,000007
estandaritzar	V	DCC	1	0,000003
nuclearitzar	V	DCC	1	0,000003
desnuclearitzar	V	DCC	3	0,000010
vulgaritzar	VT	DFA	47	0,000164
hongaritzar	V	DCC	1	0,000003
gargaritzar	VI	DFA	12	0,000042
familiaritzar	VTP	DFA	101	0,000353
velaritzar	VT	DFA	5	0,000017
escolaritzar	VT	DFA	10	0,000035
polaritzar	VTP	DFE	82	0,000287
bipolaritzar	V	DCC	1	0,000003
despolaritzar	VT	DFA	3	0,000010
exemplaritzar	V	DCC	5	0,000017
desexemplaritzar	V	DCC	1	0,000003
obstacularitzar	V	DCC	1	0,000003
secularitzar	VTP	DFE	17	0,000059
particularitzar	VTP	DFA	31	0,000108
vascularitzar	V	DCC	2	0,000007
regularitzar	VT	DFA	134	0,000469
irregularitzar	V	DCC	2	0,000007
singularitzar	VTP	DFE	76	0,000266
popularitzar	VTP	DFE	132	0,000462
proletaritzar	VT	DEC	6	0,000021
desproletaritzar	V	DCC	5	0,000017
militaritzar	VT	DFA	8	0,000028
remilitaritzar	V	DCC	1	0,000003
sedentaritzar	V	DCC	2	0,000007
tartaritzar	VT	DFA	2	0,000007
iberitzar	V	DCC	2	0,000007
suberitzar	VTP	DFE	1	0,000003
canceritzar	V	DCC	1	0,000003
atmosferitzar	V	DCC	1	0,000003
estrangeritzar	VTP	DFA	1	0,000003
encoleritzar	V	DCC	1	0,000003
polimeritzar	VT	DFA	1	0,000003
caracteritzar	VTP	DFA	2.263	0,007924
descaracteritzar	V	DCC	5	0,000017
eteritzar	VT	DFA	1	0,000003
cateteritzar	V	DCC	1	0,000003
cauteritzar	VT	DFA	11	0,000038
parqueritzar	V	DCC	1	0,000003
pulveritzar	V	DCC	49	0,000171

Figura 11. Fragmento de un índice de lemas clasificados en orden alfabético inverso, es decir, comenzando la ordenación por el final de la palabra en vez de por el principio, con la frecuencia absoluta y relativa.

BDTLC FREQUÈNCIA DE DISPERSIÓ I ÚS (NO LITERARI)

Lema: 53093 VTP donar

Grup Tipològic	Freq. prevista	Freq. real	Desviació
1. Filosofia.	3.120,105153	2.843	277,105153-
2. Religió i Teologia.	5.268,359445	6.197	928,640555
3. Ciències Socials.	9.897,123235	9.258	639,123235-
4. Premsa.	6.214,977508	6.368	153,022492
5. Ciències Pures i Naturals.	3.946,287534	3.389	557,287534-
6. Ciències Aplicades.	7.972,033227	8.535	562,966773
7. Belles Arts. Oci. Esports.	4.931,989321	4.700	231,989321-
8. Llengua i Literatura.	3.934,595355	4.037	102,404645
9. Història i Geografia.	6.026,600074	5.928	98,600074-
0. Correspondència.	233,929150	291	57,070850

Freqüència total: 67.123

Freqüència NO LITERARI: 51.546

Dispersió: 0,97053674

Ús: 50.027,28680004

Figura 12. Distribución de la frecuencia del lema DONAR VTP («dar») en el subcorpus de lengua no literaria, resultado de una consulta interactiva de la base de datos. En la primera columna puede verse la frecuencia que tendría este lema en cada grupo tipológico si estuviera uniformemente repartido, es decir, proporcionalmente a la extensión del texto en cada uno de ellos (frecuencia prevista, esperada o teórica); en la segunda columna tenemos la frecuencia que presenta el lema en cada grupo (frecuencia real), y en la tercera columna la diferencia, positiva o negativa, entre las dos. En la misma pantalla se nos ofrece la frecuencia total del lema en el corpus y en la lengua no literaria, el coeficiente de dispersión calculado según la fórmula mencionada y el uso (resultado de multiplicar la frecuencia por el coeficiente de dispersión).

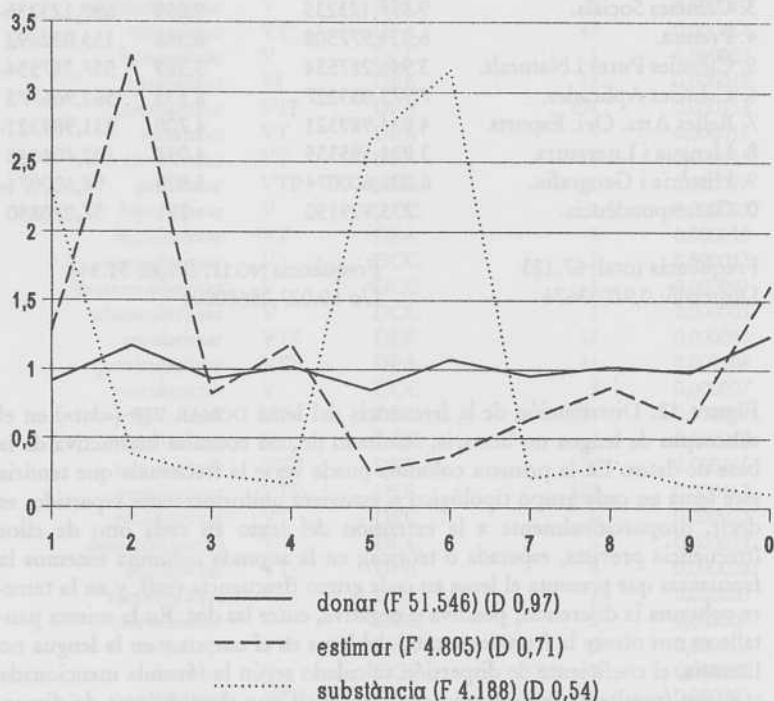


Figura 13. Representación de la frecuencia proporcional de tres lemas de distribución diferente. DONAR, que tiene un repartición muy uniforme ($D = 0.97$), presenta una línea casi horizontal con valores muy cercanos a 1 en todos los grupos (0.91, 1.17, 0.93, 1.02, 0.85, 1.07, 0.95, 1.02, 0.98, 1.24). ESTIMAR, con una distribución no tan uniforme ($D = 0.71$), se presenta con una frecuencia superior a la teórica en los grupos 1 (1.27), 2 (3.26) y 0 (1.61), mientras que tiene valores por debajo de los previstos en los grupos 5 (0.26), 6 (0.37), 7 (0.64) y 9 (0.62); en 3 (0.82), 4 (1.18) y 8 (0.87) son muy próximos a los esperados. En el caso de SUBSTÀNCIA, de repartición muy irregular ($D = 0.54$), tenemos una frecuencia superabundante en los grupos 1 (2.3), 5 (2.66) y 6 (3.16), mientras que en el resto de grupos la frecuencia es muy deficitaria (0.38 en 2, 0.23 en 3, 0.17 en 4, 0.21 en 7, 0.3 en 8, 0.16 en 9, 0.1 en 0).

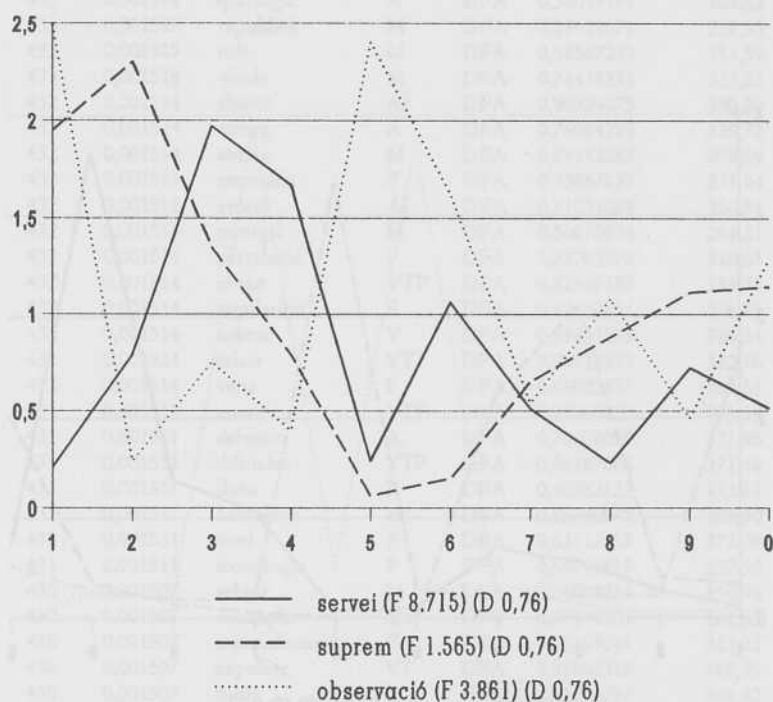


Figura 14. Representación de la frecuencia proporcional de tres lemas que tienen una distribución equivalente ($D = 0.76$), pero que reparten sus ocurrencias de manera muy distinta. SERVEI es superabundante en los grupos 3 (1.97) y 4 (1.67), en el grupo 6 su frecuencia es sensiblemente igual a la teórica (1.06) y es deficitario en los grupos restantes (0.21 en 1, 0.79 en 2, 0.24 en 5, 0.51 en 7, 0.22 en 8, 0.72 en 9, 0.51 en 0). SUPREM tiene sus máximos valores relativos en los grupos 1 (1.95), 2 (2.30) y 3 (1.35); en los grupos 9 (1.11) y 0 (1.14) está ligeramente por encima de la previsión, en 5 (0.06) y 6 (0.15) está muy por debajo y en el grupo 4 (0.79) ligeramente por debajo. OBSERVACIÓ tiene sus máximos valores en los grupos 1 (2.49), 5 (2.4) y 6 (1.62), está ligeramente por encima de la frecuencia teórica en los grupos 8 (1.08) y 0 (1.29) y en los grupos 2 (0.24), 3 (0.75), 4 (0.44), 7 (0.54) y 9 (0.46) está por debajo. Las cumbres más destacadas son filosofía y ciencias (puras, naturales y aplicadas) para *observació*, filosofía y religión para *suprem*, y ciencias sociales y prensa para *servei*.

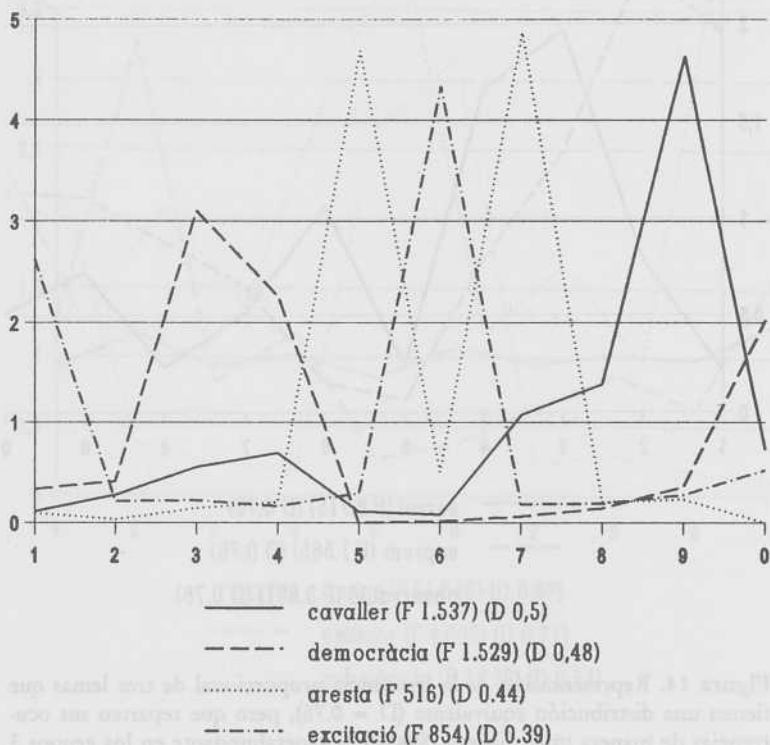


Figura 15. Representación de la frecuencia proporcional de lemas con repartición muy irregular (coeficiente de dispersión igual a 0.5 o inferior). Los valores más elevados de la frecuencia están concentrados en unos grupos característicos para cada uno de ellos: geografía e historia para *cavaller*, con una frecuencia 4.63 veces la prevista; ciencias sociales, prensa y correspondencia para *democràcia*, con una frecuencia 3.1, 2.26 y 2.02 veces la prevista, respectivamente; ciencias puras y naturales y bellas artes, ocio y deportes para *aresta*, con una frecuencia 4.69 y 4.87 veces la prevista, respectivamente; y filosofía y ciencias aplicadas para *excitació*, con una frecuencia 2.62 y 4.33 veces la prevista, respectivamente.

freqüència		lema	c.g.	proc.	dispersió	ús
absoluta	relativa					
433	0,001518	ira	F	DFA	0,68000854	294,44
433	0,001518	olivera	F	DFA	0,78222604	338,70
433	0,001518	quirúrgic	A	DFA	0,38019195	164,62
433	0,001518	república	M	DFA	0,59431071	257,33
433	0,001518	solc	M	DFA	0,58567243	253,59
433	0,001518	vincle	M	DFA	0,74418533	322,23
432	0,001514	absent	AI	DFA	0,90324075	390,20
432	0,001514	amarg	A	DFA	0,76094299	328,72
432	0,001514	ample	M	DFA	0,69142282	298,69
432	0,001514	anomalia	F	DFA	0,53667639	231,84
432	0,001514	ardent	AI	DFA	0,71031083	306,85
432	0,001514	cirurgia	M	DFA	0,56670834	244,81
432	0,001514	correlació	F	DFA	0,73762772	318,65
432	0,001514	esvair	VTP	DFA	0,82349187	355,74
432	0,001514	regularitat	F	DFA	0,82614854	356,89
432	0,001514	sobrar	V	DFA	0,89431583	386,34
432	0,001514	teixir	VT	DFA	0,74713219	322,76
432	0,001514	vena	F	DFA	0,61022337	263,61
431	0,001511	asseure	VTP	DFA	0,84360235	363,59
431	0,001511	defensiu	A	DFA	0,74676059	321,85
431	0,001511	difondre	VTP	DFA	0,86187066	371,46
431	0,001511	flora	F	DFA	0,40282622	173,61
431	0,001511	hectolitre	M	DFA	0,65730972	283,30
431	0,001511	sord	A	DFA	0,63117559	272,03
431	0,001511	tecnologia	F	DFA	0,59745859	257,50
430	0,001507	arbust	M	DFA	0,36038238	154,96
430	0,001507	eixampla	MF	DFE	0,37571201	161,55
430	0,001507	especialització	F	DFA	0,75195081	323,33
430	0,001507	expulsar	VT	DFA	0,85802358	368,95
430	0,001507	mitja	F	DFA	0,80214732	344,92
430	0,001507	ocult	A	DFA	0,76618870	329,46
430	0,001507	òrbita	F	DFA	0,60401966	259,72
430	0,001507	alpí	A	DFA	0,44274682	189,93
429	0,001504	detingut	A	*	0,70175968	301,05
420	0,001504	erigir	VTP	DFA	0,82661013	354,61
429	0,001504	imaginari	A	DFA	0,81751667	350,71
429	0,001504	sepultura	F	DFA	0,65110689	279,32
429	0,001504	verbigràcia	AV	DEC	0,45597309	195,61
428	0,001500	contràriament	AV	DFA	0,87268725	373,51
428	0,001500	criminal	AI	DFA	0,67870421	290,48
428	0,001500	esquizofrènia	F	DFA	0,01967756	8,42
428	0,001500	liquidació	F	DFA	0,58007615	248,27
428	0,001500	multiplicació	F	DFA	0,68212841	291,95
428	0,001500	ordenança	F	DFA	0,67711505	289,80
428	0,001500	plantejat	A	*	0,79316729	339,47

Figura 16. Fragmento de un índice de lemas clasificados por orden decreciente de frecuencias, con el valor del coeficiente de dispersión y el uso. En él puede observarse cómo varía el uso en función de la dispersión para una misma frecuencia.

<i>ús</i>	<i>lema</i>	<i>c.g.</i>	<i>proc.</i>	<i>frequència</i>	<i>dispersió</i>
433,89	promoció	F	DFA	611	0,71014658
433,65	convidar	VT	DFA	580	0,74768915
433,62	jo	M	DFA	956	0,45358410
433,54	clos	A	DFA	489	0,88658964
433,52	gramàtica	F	DFA	938	0,46218138
433,45	mer	A	DFA	558	0,77679827
433,40	pacífic	A	DFA	549	0,78945224
433,25	enfilat	VVP	DFA	514	0,84291179
433,21	rostre	M	DFA	562	0,77083650
432,98	prejudici	M	DFA	534	0,81083218
432,83	anhel	M	DFA	533	0,81207160
432,57	edificació	F	DFA	570	0,75889719
432,49	productor	M	DFA	555	0,77927381
432,49	taxa	F	DFA	697	0,62051305
432,42	excitar	VTP	DFA	535	0,80827316
431,79	divuit	AN	DFA	470	0,91870423
431,33	navegació	F	DFA	565	0,76341702
431,18	perfil	M	DFA	630	0,68442493
431,08	circular	AI	DFA	652	0,66118056
431,07	productor	A	DFA	597	0,72207391
430,83	argila	F	DFA	736	0,58537558
430,59	solitari	A	DFA	481	0,89521488
430,49	protagonista	M	DFA	621	0,69323185
430,27	entusiasta	AI	DFA	537	0,80126593
430,05	mudar	VVP	DFE	504	0,85327512
429,92	discret	A	DFA	484	0,88828185
428,98	contradir	VTP	DFA	532	0,80636959
428,71	vaca	F	DFA	651	0,65855263
428,68	invertir	VT	DFA	503	0,85226589
428,60	trajectòria	F	DFA	532	0,80564919
428,46	genoll	M	DFA	562	0,76239631
428,44	sobirà	A	DFA	558	0,76781485
428,41	suficientment	AV	DFA	502	0,85340749
427,84	menester	M	DFA	516	0,82915484
427,83	reverend	A	DFA	695	0,61558525
427,64	endins	AV	DFA	536	0,79783729
427,64	forçat	A	DEC	455	0,93987237
427,43	invitar	VT	DFA	571	0,74857979
427,41	còmode	A	DFA	476	0,89793982
427,39	intensament	AV	DFA	472	0,90549466
427,38	vàlua	F	DFA	495	0,86340313
427,29	casat	A	DEC	542	0,78837215
427,11	renaixença	F	DFA	882	0,48425668
427,02	futbol	M	DFA	781	0,54677150
426,93	oblit	M	DFA	496	0,86076316
426,84	mortalitat	F	DFA	711	0,60034523
426,66	pasqua	F	DFA	725	0,58850538

Figura 17. Fragmento de un índice de lemas clasificados por orden decreciente de uso, con el valor de la frecuencia y el coeficiente de dispersión.

<i>dispersió</i>	<i>lema</i>	<i>c.g.</i>	<i>proc.</i>	<i>frequència</i>	<i>ús</i>
0,76605560	ubicació	F	DEC	66	50,55
0,76604562	ambient	AI	DFA	356	272,71
0,76603934	ressorgiment	M	DFA	180	137,88
0,76602642	breu	AV	DCC	21	16,08
0,76602306	puja	F	DFA	149	114,13
0,76602302	embenat	A	DCC	25	19,15
0,76601934	viari	A	DFA	55	42,13
0,76597808	foment	M	DFA	790	605,12
0,76597066	rialler	A	DFA	147	112,59
0,76596302	gregarisme	M	DFA	13	9,95
0,76596122	observació	F	DFA	3.855	2.952,78
0,76595137	context	M	DFA	965	739,14
0,76592962	portador	A	DFA	213	163,14
0,76592232	content	A	DFA	710	543,80
0,76590477	sensat	A	DFA	63	48,25
0,76588357	esquarterar	VT	DFA	44	33,69
0,76585282	desmantellar	VT	DFA	24	18,38
0,76584678	sext	A	DCC	83	63,56
0,76576273	fascinar	VT	DFA	58	44,41
0,76574357	competitiu	A	DFA	108	82,70
0,76571770	embastar	VT	DFA	12	9,18
0,76571565	servei	M	DFA	8.703	6.664,02
0,76569190	gall	M	DFA	537	411,17
0,76567561	atractívol	A	DFA	57	43,64
0,76563981	suplementari	A	DFA	179	137,04
0,76562690	tórtora	F	DFA	64	49,00
0,76562088	suprem	A	DFA	1.562	1.195,89
0,76560028	provador	A	DFA	8	6,12
0,76559572	arravatar	VP	DFA	58	44,40
0,76556890	passada	D	DFA	953	729,58
0,76554373	escarmentar	VVP	DFA	40	30,62
0,76553920	doble	M	DFA	315	241,14
0,76551333	suburbi	M	DFA	89	68,13
0,76551260	abatiment	M	DFA	72	55,11
0,76549296	commoure	VT	DFA	352	269,45
0,76548766	militant	AI	DFA	118	90,32
0,76545967	desatendre	VT	DFA	86	65,82
0,76545883	seguidor	M	DFA	343	262,55
0,76545849	competidora	F	DFA	12	9,18
0,76542800	maltractar	VT	DFA	139	106,39
0,76541997	tristesia	F	DFA	508	388,83
0,76540934	arrest	M	DFA	20	15,30
0,76535983	indiscret	A	DFA	47	35,97
0,76535158	irresponsabilitat	F	dfa	50	38,26
0,76533153	flotant	AI	DFA	97	74,23
0,76532102	isolar	VTP	DFE	213	163,01
0,76531716	repartidor	A	DFA	11	8,41

Figura 18. Fragmento de un índice de lemas clasificados por coeficiente de dispersión en orden decreciente, con expresión de la frecuencia y el uso.

	COMPLETO	-BLE	-MENT
CORPUS	129.007	1.547	3.497
DICCIONARIO	85.373	1.276	2.390
C-D	74.037	691	1.687
C-D %	57,38	44,66	48,24
$C \cap D$	54.970	856	1.810
$C \cap D$ % /C	42,62	55,33	51,75
$C \cap D$ % /D	64,39	67,08	77,35
D-C	30.403	420	530
D-C %	35,61	32,91	22,64

Figura 19.

	CORPUS	DICCION.	C-D	$C \cap D$	D-C
ACET-	30	49	<i>13</i>	17	32
ANGIO-	27	32	18	9	23
BENZO-	4	17	<i>1</i>	3	14
BIBLIO-	15	29	3	12	17
BIO-	133	99	91	42	57
BRADI-	10	16	6	4	12
CARDIO-	29	22	20	9	13
CITO-	29	28	17	12	16
HELIO-	18	28	7	11	17
HEMO-	57	37	31	26	11
HIPER-	257	88	209	48	40
HIPO-	209	108	149	60	48
MACRO-	72	44	56	16	28
MICRO-	187	128	132	55	73
NEFRO-	24	20	18	6	14
QUIMIO-	12	36	2	10	26
SERO-	22	29	9	13	16
ULTRA-	117	68	81	36	32
XANTO-	5	18	3	2	16
ZOO-	38	46	15	23	33

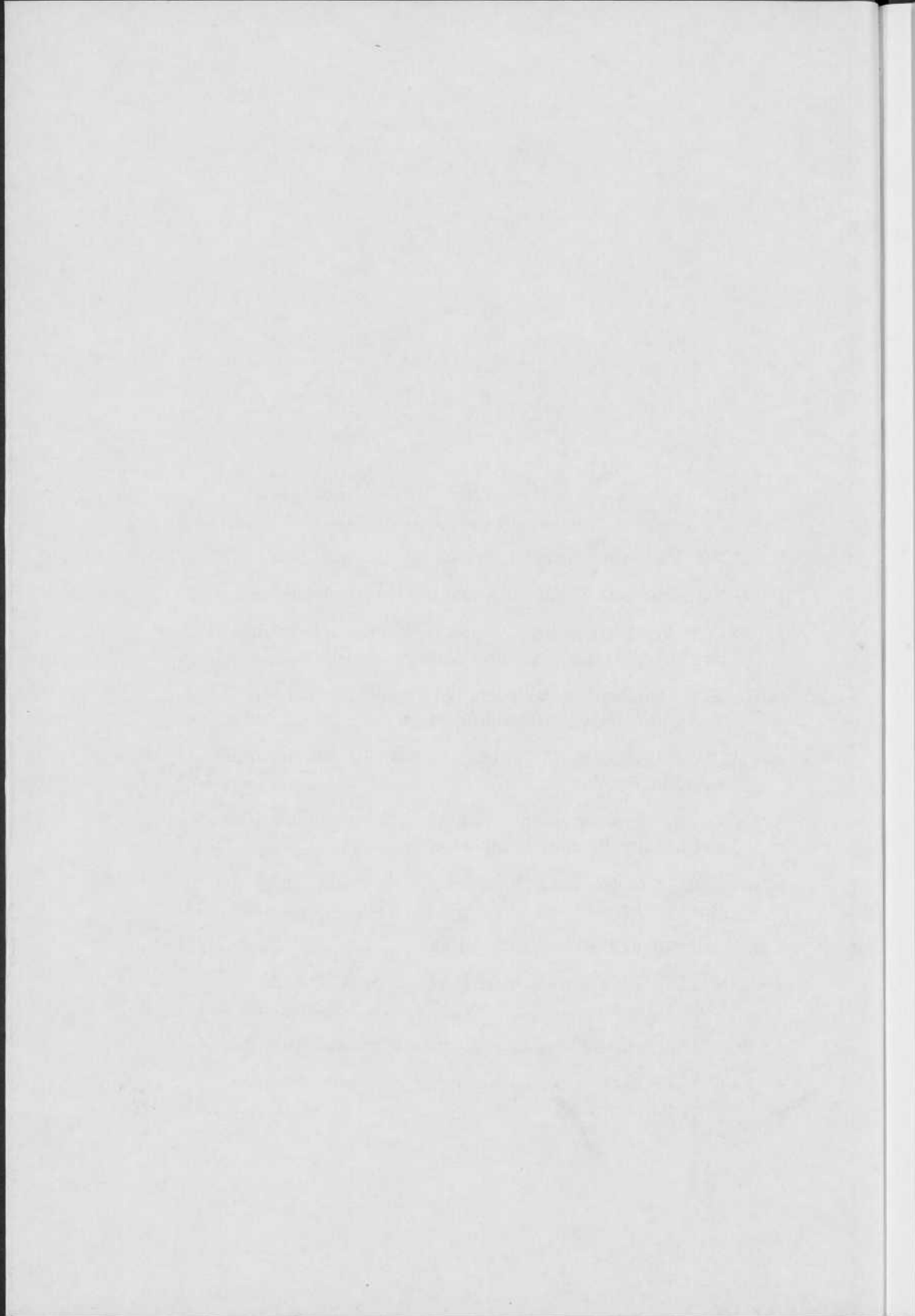
Figura 20. En las dos primeras columnas vemos el número de lemas que aparecen en el corpus y en el diccionario con cada uno de los prefijos cultos o formas prefijadas que figuran en el cuadro; la cifra superior en cada caso se ha puesto en negrilla; como puede observarse, en unos casos predomina el corpus y en otros el diccionario. En las tres columnas restantes aparecen el número de lemas exclusivos del corpus (que no figuran en el diccionario de base), el número de lemas comunes (que aparecen en el corpus y en el diccionario) y el número de lemas exclusivos del diccionario (que no aparecen en el corpus); podemos observar que la cifra más elevada (en negrilla) no corresponde nunca a los lemas comunes, sino que en diez casos corresponde a los lemas exclusivos del corpus y en los otros diez a los lemas exclusivos del diccionario; la cifra menor (en cursiva), en cambio, unas veces corresponde a los lemas exclusivos del corpus (siete), otras corresponde a los lemas exclusivos del diccionario (cuatro) y otras a los lemas comunes a los dos (nueve).

ÍNDICE

INDICE	3
A. SIMONE, <i>Verbi etimologici in Gallico</i>	9
M. SISO, <i>La rievocazione del dialetto in un reperto arial</i>	25
J. RIV-GERONT, <i>Les dialectes romans de l'Estrope et de la caudina dans les dialectes de langue italienne</i>	37
E. G. BELTACH, <i>Notizie dell'opera del cardinale italiano di Tosca della lingua italiana dell'origine</i>	49
M. ALVAR ESQUERRA, <i>Compendio de las diferencias etimológicas de</i>	65
J. BOGHO, <i>Lexicographie historique et appliquée de l'italien des mots de haute fréquence</i>	87
M. GONZÁLEZ GONZÁLEZ, <i>O Diccionario de lingua gallega de Real Academia Gallega</i>	111
J. GONZÁLEZ, <i>Diccionario etimológico gallego</i>	121
I. GONZÁLEZ, <i>O diccionario etimológico gallego de Real Academia Gallega</i>	131
I. BALESCA, <i>Lexicografía etimológica de los dialectos gallegos</i>	141
J. BARRI I FORTINATA, <i>Diccionariu y corpus etimològic Prosopopeyas de català</i>	157

	COMPLETO	EN F	EN F/2
COMPLETO	10000	1000	500
EN F	1000	100	50
EN F/2	500	50	25
EN F/4	250	25	12,5
EN F/8	125	12,5	6,25
EN F/16	62,5	6,25	3,125
EN F/32	31,25	3,125	1,5625
EN F/64	15,625	1,5625	0,78125
EN F/128	7,8125	0,78125	0,390625
EN F/256	3,90625	0,390625	0,1953125
EN F/512	1,953125	0,1953125	0,09765625
EN F/1024	0,9765625	0,09765625	0,048828125
EN F/2048	0,48828125	0,048828125	0,0244140625
EN F/4096	0,244140625	0,0244140625	0,01220703125
EN F/8192	0,1220703125	0,01220703125	0,006103515625
EN F/16384	0,06103515625	0,006103515625	0,0030517578125
EN F/32768	0,030517578125	0,0030517578125	0,00152587890625
EN F/65536	0,0152587890625	0,00152587890625	0,000762939453125
EN F/131072	0,00762939453125	0,000762939453125	0,0003814697265625
EN F/262144	0,003814697265625	0,0003814697265625	0,00019073486328125
EN F/524288	0,0019073486328125	0,00019073486328125	0,000095367431640625
EN F/1048576	0,00095367431640625	0,000095367431640625	0,0000476837158203125
EN F/2097152	0,000476837158203125	0,0000476837158203125	0,00002384185791015625
EN F/4194304	0,0002384185791015625	0,00002384185791015625	0,000011920928955078125
EN F/8388608	0,00011920928955078125	0,000011920928955078125	0,0000059604644775390625
EN F/16777216	0,000059604644775390625	0,0000059604644775390625	0,00000298023223876953125
EN F/33554432	0,0000298023223876953125	0,00000298023223876953125	0,000001490116119384765625
EN F/67108864	0,00001490116119384765625	0,000001490116119384765625	0,0000007450580596923828125
EN F/134217728	0,000007450580596923828125	0,0000007450580596923828125	0,00000037252902984619140625
EN F/268435456	0,0000037252902984619140625	0,00000037252902984619140625	0,000000186264514923095703125
EN F/536870912	0,00000186264514923095703125	0,000000186264514923095703125	0,0000000931322574615478515625
EN F/1073741824	0,000000931322574615478515625	0,0000000931322574615478515625	0,0000000465661287307739278125
EN F/2147483648	0,000000465661287307739278125	0,0000000465661287307739278125	0,00000002328306436538696390625
EN F/4294967296	0,0000002328306436538696390625	0,00000002328306436538696390625	0,000000011641532182693481953125
EN F/8589934592	0,00000011641532182693481953125	0,000000011641532182693481953125	0,0000000058207660913467409765625
EN F/17179869184	0,000000058207660913467409765625	0,0000000058207660913467409765625	0,00000000291038304567337048828125
EN F/34359738368	0,0000000291038304567337048828125	0,00000000291038304567337048828125	0,000000001455191522836685244140625
EN F/68719476736	0,00000001455191522836685244140625	0,000000001455191522836685244140625	0,0000000007275957614183426220703125
EN F/137438953472	0,000000007275957614183426220703125	0,0000000007275957614183426220703125	0,00000000036379788070917131103515625
EN F/274877906944	0,0000000036379788070917131103515625	0,00000000036379788070917131103515625	0,000000000181898940354585655517578125
EN F/549755813888	0,00000000181898940354585655517578125	0,000000000181898940354585655517578125	0,0000000000909494701772928277587890625
EN F/1099511627776	0,000000000909494701772928277587890625	0,0000000000909494701772928277587890625	0,00000000004547473508864641387939453125
EN F/2199023255552	0,0000000004547473508864641387939453125	0,00000000004547473508864641387939453125	0,000000000022737367544323206939697265625
EN F/4398046511104	0,00000000022737367544323206939697265625	0,000000000022737367544323206939697265625	0,0000000000113686837721616034696986328125
EN F/8796093022208	0,000000000113686837721616034696986328125	0,0000000000113686837721616034696986328125	0,00000000000568434188608017323484931640625
EN F/17592186044416	0,0000000000568434188608017323484931640625	0,00000000000568434188608017323484931640625	0,000000000002842170943040086617174208203125
EN F/35184372088832	0,00000000002842170943040086617174208203125	0,000000000002842170943040086617174208203125	0,0000000000014210854715200433085871041015625
EN F/70368744177664	0,000000000014210854715200433085871041015625	0,0000000000014210854715200433085871041015625	0,00000000000071054273576002165429355205078125
EN F/140737488355328	0,0000000000071054273576002165429355205078125	0,00000000000071054273576002165429355205078125	0,000000000000355271367880010827146776025390625
EN F/281474976710656	0,00000000000355271367880010827146776025390625	0,000000000000355271367880010827146776025390625	0,0000000000001776356839400054135733880126953125
EN F/562949953421312	0,000000000001776356839400054135733880126953125	0,0000000000001776356839400054135733880126953125	0,00000000000008881784197000270678669000634765625
EN F/1125899906842624	0,0000000000008881784197000270678669000634765625	0,00000000000008881784197000270678669000634765625	0,000000000000044408920985001353393345003173828125
EN F/2251799813685248	0,00000000000044408920985001353393345003173828125	0,000000000000044408920985001353393345003173828125	0,0000000000000222044604925006766966725015869140625
EN F/4503599627370496	0,000000000000222044604925006766966725015869140625	0,0000000000000222044604925006766966725015869140625	0,00000000000001110223024625033834833625079345703125
EN F/9007199254740992	0,0000000000001110223024625033834833625079345703125	0,00000000000001110223024625033834833625079345703125	0,000000000000005551115123125169174168125396728515625
EN F/18014398509481984	0,00000000000005551115123125169174168125396728515625	0,000000000000005551115123125169174168125396728515625	0,0000000000000027755575615625845870840626983642578125
EN F/36028797018963968	0,000000000000027755575615625845870840626983642578125	0,0000000000000027755575615625845870840626983642578125	0,00000000000000138777878078127279354031349418212890625
EN F/72057594037927936	0,0000000000000138777878078127279354031349418212890625	0,00000000000000138777878078127279354031349418212890625	0,0000000000000006938893903906363967701672092091044545453125
EN F/144115188075855872	0,000000000000006938893903906363967701672092091044545453125	0,0000000000000006938893903906363967701672092091044545453125	0,000000000000000346944695195318198385083604604552222727265625
EN F/288230376151711744	0,000000000000003469446951953181983850836046045522227265625	0,0000000000000003469446951953181983850836046045522227265625	0,000000000000000173472347597659096925041802302276136326328125
EN F/576460752303423488	0,00000000000000173472347597659096925041802302276136326328125	0,000000000000000173472347597659096925041802302276136326328125	0,0000000000000000867361737988295484625209011511368161631640625
EN F/1152921504606846976	0,000000000000000867361737988295484625209011511368161631640625	0,0000000000000000867361737988295484625209011511368161631640625	0,00000000000000004336808689941477423125045057556808081631640625
EN F/2305843009213693952	0,0000000000000004336808689941477423125045057556808081631640625	0,00000000000000004336808689941477423125045057556808081631640625	0,0000000000000000216840434497073871156252252877780404081631640625
EN F/4611686018427387904	0,000000000000000216840434497073871156252252877780404081631640625	0,0000000000000000216840434497073871156252252877780404081631640625	0,0000000000000000108420217248536935578125112643889020204081631640625
EN F/9223372036854775808	0,000000000000000108420217248536935578125112643889020204081631640625	0,0000000000000000108420217248536935578125112643889020204081631640625	0,00000000000000000542101086242684677881255632194451010204081631640625
EN F/18446744073709551616	0,0000000000000000542101086242684677881255632194451010204081631640625	0,0000000000000000542101086242684677881255632194451010204081631640625	0,0000000000000000027105054312134423889062781607051010204081631640625
EN F/36893488147419103232	0,000000000000000027105054312134423889062781607051010204081631640625	0,000000000000000027105054312134423889062781607051010204081631640625	0,00000000000000000135525271560672119445313905070251010204081631640625
EN F/73786976294838206464	0,0000000000000000135525271560672119445313905070251010204081631640625	0,0000000000000000135525271560672119445313905070251010204081631640625	0,00000000000000000067762635780336059722656750251010204081631640625
EN F/147573952589676412928	0,0000000000000000067762635780336059722656750251010204081631640625	0,0000000000000000067762635780336059722656750251010204081631640625	0,000000000000000000338813178901680298613283750251010204081631640625
EN F/295147905179352825856	0,00000000000000000338813178901680298613283750251010204081631640625	0,00000000000000000338813178901680298613283750251010204081631640625	0,0000000000000000001694065894508401493066418750251010204081631640625
EN F/590295810358705651712	0,000000000000000001694065894508401493066418750251010204081631640625	0,000000000000000001694065894508401493066418750251010204081631640625	0,00000000000000000008470329472542007465332093750251010204081631640625
EN F/1180591620717411303424	0,0000000000000000008470329472542007465332093750251010204081631640625	0,0000000000000000008470329472542007465332093750251010204081631640625	0,000000000000000000042351647362710037326660468750251010204081631640625
EN F/2361183241434822606848	0,00000000000000000042351647362710037326660468750251010204081631640625	0,00000000000000000042351647362710037326660468750251010204081631640625	0,0000000000000000000211758236813550186633302343750251010204081631640625
EN F/4722366482869645213696	0,000000000000000000211758236813550186633302343750251010204081631640625	0,000000000000000000211758236813550186633302343750251010204081631640625	0,00000000000000000001058791184067750933166511718750251010204081631640625
EN F/9444732965739290427392	0,0000000000000000001058791184067750933166511718750251010204081631640625	0,0000000000000000001058791184067750933166511718750251010204081631640625	0,00000000000000000005293955920338754665832558593750251010204081631640625
EN F/18889465931478580854784	0,00000000000000000005293955920338754665832558593750251010204081631640625	0,00000000000000000005293955920338754665832558593750251010204081631640625	0,000000000000000000026469779601693773329162927968750251010204081631640625
EN F/37778931862957161709568	0,000000000000000000026469779601693773329162927968750251010204081631640625	0,000000000000000000026469779601693773329162927968750251010204081631640625	0,00000000000000000001323488980084688665808146139843750251010204081631640625
EN F/75557863725914323419136	0,00000000000000000001323488980084688665808146139843750251010204081631640625	0,000000000000000000013234889800846886658081461398437502510	

LIMIAR	5
R. SIMONE, Verbi sintagmatici in italiano	9
M. SECO, La microestructura del diccionario del español actual	25
J. REY-DEBOVE, Les domaines respectifs de l'exemple et de la citation dans les dictionnaires de langue actuels	39
P. G. BELTRAMI, Notizie dall'opera del vocabolario italiano: il <i>Tesoro della lingua italiana delle origini</i>	49
M. ALVAR EZQUERRA, Cincuenta años de diccionarios monolingües Vox	65
J. PICOCHÉ, Lexicographie théorique et appliquée: un dictionnaire des mots de haute fréquence	87
M. GONZÁLEZ GONZÁLEZ, O diccionario da lingua galega da Real Academia Galega	111
I. GONZÁLEZ, Diccionario italiano-galego	121
I. GONZÁLEZ, O diccionario castelán-galego da Real Academia Galega	131
I. SARASOLA, Lexicografía vasca: el diccionario general vasco	141
J. RAFEL I FONTANALS, Diccionarios y corpus textuales. Perspectivas para el catalán	157



R E A L
ACADEMIA
GALEGA



CONTIDO

- 5 C. GARCÍA, I. GONZÁLEZ, M. GONZÁLEZ,
Limiar
- 9 R. SIMONE,
Verbi sintagmatici in italiano
- 25 M. SECO,
La microestructura del
diccionario del español actual
- 39 J. REY-DEBOVE,
Les domaines respectifs de l'exemple
et de la citation dans les
dictionnaires de langue actuels
- 49 P. G. BELTRAMI,
Notizie dall'opera del vocabolario italiano:
il Tesoro della lingua italiana delle origini
- 65 M. ALVAR EZQUERRA,
Cincuenta años de
diccionarios monolingües Vox
- 87 J. PICOCHÉ,
Lexicographie théorique et appliquée:
un dictionnaire des mots de haute fréquence
- 111 M. GONZÁLEZ GONZÁLEZ,
O dicionario da lingua galega da
Real Academia Galega
- 121 I. GONZÁLEZ,
Diccionario italiano-galego
- 131 I. GONZÁLEZ,
O dicionario castelán-galego da
Real Academia Galega
- 141 I. SARASOLA,
Lexicografía vasca: el
diccionario general vasco
- 157 J. RAFEL I FONTANALS,
Diccionarios y corpus textuales.
Perspectivas para el catalán

A N E N O S . X I I I . C A D E R N O S D E L I N G U A