

do dominio lingüístico castelán. Bascuas explica formulando a hipótese de que o sistema paleoeuropeo sen vogal /ō/ sería o orixinario, e que os hidrónimos con esa vogal representarían unha innovación, idéntica á que tivo lugar, de xeito independente, nas linguas con cinco vogais (o latín, o grego, o celta...).

Os dous libros de Edelmiro Bascuas sobre hidronimia (2002 e 2006), aos que podemos sumar numerosos traballos breves aparecidos en publicacións periódicas, venñen a proxectar unha luz formidable sobre a toponimia e o léxico prelatinos do noroeste hispánico ao tempo que proporcionan valiosa información para o coñecemento da prehistoria lingüística de Europa.

¹ Bascuas cita o pola segunda edición (Bern-Stuttgart, 1989).

² O autor tamén contradi a opinión de Villar sobre o vocalismo prerromano en E. Bascuas. (2002).

REFERENCIAS BIBLIOGRÁFICAS

Alinei, Mario / Francesco Benozzo [2006]: "Alcuni aspetti della Teoria della Continuità Paleolitica applicata all'area gallega", *Actas do II Congreso Internacional de Onomástica Galega* (Pontevedra, 19-21 de outubro de 2006) (no prelo), (<http://www.continuitas.com>).

Bascuas, Edelmiro (2002): *Estudios de hidronimia paleoeuropea gallega*. Santiago de Compostela: USC (Anexo 51 de *Verba*).

Pokorny, Julius (1994³): *Indogermanisches Etymologisches Wörterbuch*, 2 vols. Tübingen – Basel: Francke.

Untermann, Jürgen (1985-1986): "Lusitanisch, Keltiberisch, Keltisch", *Veleia* 2-3, 57-76 (número dedicado a *Studia Paleohispanica*. *Actas del IV Coloquio sobre lenguas y culturas paleohispánicas* [Vitoria – Gasteiz, 6-10 maio 1985], editado por J. Gorrochategui / J. L. Melena / J. Santos).

Untermann, Jürgen (1992): "Anotaciones al estudio de las lenguas prerromanas del Noroeste de la Península Ibérica (1)", en *Galia: da romanidade á xermanización* (*Actas do encontro científico en homenaxe a Fermín Bouza Brey*). Santiago de Compostela: Museo do Pobo Galego, 367-397.

Villar Liébana, Francisco (1995): *Estudios de celtibérico y de toponimia prerromana*. Salamanca: Ediciones Universidad de Salamanca.

BUGARÍN LÓPEZ, M^a XESÚS (2007): O GALEGO FUNDAMENTAL: DICCIONARIO DE FRECUENCIAS. A CORUÑA: FUNDACIÓN PEDRO BARRIÉ DE LA MAZA, 1109 PP.

O contido desta obra de M^a Xesús Bugarín consiste fundamentalmente nun dicionario de frecuencias, froito da investigación desenvolvida pola autora para a súa tese de doutoramento. É, que saibamos, o primeiro traballo destas características publicado en lingua galega que se centra no léxico adulto, que incorpora material en lingua oral e que aparece en forma de libro.

O libro consta de sete capítulos, sendo os catro primeiros de carácter fundamental, conceptual. Os capítulos centrais son o 5 e o 6, os cales constitúen a contribución máis persoal porque se centran na creación dun corpus, na elaboración do dicionario de frecuencias e na análise dos datos obtidos. O texto inclúe un último capítulo titulado "Conclusións", que basicamente consiste nun resumo. Finalmente, inclúe referencias bibliográficas.

Bugarín no capítulo 1 explica as abreviaturas que emprega e no capítulo 2 ofrece unha breve xustificación da lingüística de corpus.

O capítulo 3, que se titula "Os córpora. Ferramentas metodolóxicas de análise lingüística", trata o concepto de corpus e fai referencia a algunhas dificultades para a súa elaboración. Sostén que non existe unha definición unanimemente aceptada de corpus. Con todo, pódese entender que é unha mostra de material lingüístico procedente fundamentalmente de producións escritas, pero tamén de producións orais. Esta mostra debe caracterizarse por recoller materiais lingüísticos o máis variados posible, tender á representatividade da lingua de que se trate, ter en conta unha determinada finalidade e permiti-la xestión informatizada. Por outra parte, a autora comenta os diferentes tipos de *córpora* (con finalidade xeral versus con finalidades específicas, orais versus escritos, de linguaxe xeral versus de linguaxes específicas e íntegros versus de referencia). Continúa o capítulo coa exposición das súas posibles utilidades (en lexicografía, semántica

Gonzalo Navaza

ca, fonética, lingüística computacional etc.). O último punto do capítulo céntrase na descrición dalgúns exemplos de *córpore*, orais e escritos en castelán, portugués e galego.

No capítulo 4 estúdanse os vocabularios e dicionarios de frecuencia. Comeza cun sucinto achegamento ó concepto de léxico fundamental, que sería a parte dos recursos léxicos que habitualmente usan os falantes dunha lingua. O capítulo continúa coa discusión acerca do criterio máis adecuado para establecer este léxico. Nesta investigación, seguindo a A. G. Sciarone (1977), adóptase a frecuencia como o criterio máis adecuado para este fin (p. 48). A frecuencia refírese ó número de ocasións en que se produce unha unidade lingüística nun determinado corpus e está unida ás nocións de utilidade e dispoñibilidade. O uso deste criterio permite elaborar listas ou dicionarios de frecuencias a partir dos que se constrúen os vocabularios fundamentais, básicos ou usuais. Tamén se abordan no capítulo 4 as posibles aplicacións destes dicionarios (deseño de programas de ensinanza de linguas e de intervención en trastornos da linguaxe e da fala, elaboración de textos escolares, aplicación en tecnoloxías da fala etc). Finalmente, descríbense algúns traballos sobre frecuencias en francés, italiano, español, catalán, asturiano e galego.

O capítulo 5 está adicado ó proceso de construción dun corpus de carácter textual e á elaboración do dicionario de frecuencias. A elaboración do corpus comporta solucionar unha serie de dificultades, entre as que se encontra conseguir que sexa o máis representativo posible da lingua que se trate. Para solucionar esta dificultade, Bugarín utilizou un número o suficientemente amplo de materiais, fixando o límite de ocorrencias en torno a 2 millóns. Traballou con mostras tanto de lingua escrita (112 textos) coma oral (25 textos) do período cronolóxico entre 1975 e 1999. A mostra de lingua escrita representa o 65% do total e está formada con diferentes porcentaxes por obras literarias (ensaio, ensaio-etnografía, novela, teatro e poesía), científico-técnicas (bioloxía, astroloxía, química, medicina, economía e arquitectura), divulgativas (basicamente sobre cooperativismo e medicina popular), prensa

(tres números de 1997 da revista mensual *Tempos novos*) e dúas traducións (narrativa e ensaio). Empregou documentos completos, agás nas científico-técnicas en que seleccionou as 9000 primeiras ocorrencias de cada texto. A mostra de lingua oral, que representa o 35% do total, está composta por intervencións formais e non-formais.

Para prepara-los textos mantivo a substancia fonética, morfosintáctica e léxica, pero introduciu algunhas modificacións para uniformalos na maior medida do posible. Estas modificacións consistiron basicamente en corrixir erratas tipográficas; excluír topónimos, antropónimos, estranxeirismos, citas e prólogos de autor diferente; uniformar as normas ortográficas con b/v, h/ø, i/y; separar do verbo os pronomes enclíticos; eliminar-lo máximo número posible de guións nas formas compostas; desenvolver as formas grafadas con apóstrofo e excluír los números, agás os romanos de I a X.

Os materiais foron recollidos en 114 arquivos informáticos, un por cada documento-fonte. Extraeu tódalas formas gráficas do corpus por medio do programa informático *WordSmith Tools*, que permite ordenalas alfabeticamente e por frecuencia. O número total de ocorrencias é de 2.064.847, que se corresponden con 87.273 formas distintas. A continuación obtivo os contextos en que aparecían estas e que estaban formados polas oracións en que estaban incluídas. Unha vez determinados na base de datos os campos para cada palabra relativos ós contextos, á forma, á categoría gramatical, ó tipo de lingua, ó rexistro e á fonte, pasou a realiza-la lematización de forma manual. Esta consiste “na adxudicación dun lema ou forma léxica única a cada unha das ocorrencias” (p. 84). Dito doutro modo, trátase de contabilizar formas básicas, unha vez descontadas as formas que se repiten total ou parcialmente.

Neste proceso, salvo algunhas excepcións que afectan a 454 ocorrencias, adoptou a palabra como unidade mínima de traballo, entendendo por palabra o elemento que aparece separado por espazos en branco. Utilizou unha serie de criterios para establecer-los lemas: computar separadamente as formas homógrafas cando pertencían a categorías gramaticais distintas, cando proce-

dían de étimos diferentes ou cando os significados eran moi distintos; sinala-los lemas en masculino/singular con carácter xeral; sinala-las formas verbais en infinitivo e as locucións locativas como unha forma etc.

No capítulo 6 da publicación preséntase o dicionario de frecuencias. Este dicionario contén soamente os lemas cunha ocorrencia maior de 2, é dicir, que aparecen máis de dúas veces no corpus de 2.064.847 ocorrencias. Deste modo, os 32.529 lemas iniciais vense reducidos a uns 18.000, segundo un cálculo aproximado (a cifra exacta non aparece no texto). Está presentado en dúas listas para a consulta: unha por frecuencia decrecente (cando hai máis dun lema coa mesma frecuencia, aparecen ordenados alfabeticamente) e outra lista por orde alfabética, na que aparecen baixo cada lema as ocorrencias totais do mesmo, ademais das diferentes formas dese lema e as ocorrencias correspondentes.

No mesmo capítulo 6, a autora efectúa unha serie de análises dos datos desde o punto de vista da frecuencia. Comeza cun resumo dos datos xerais extraídos do corpus na súa totalidade. Entre estes datos que non expuxera anteriormente están que, dos 32.529 lemas que obtivo, 10.117 (31,1%) presentan unha única ocorrencia. Por outro lado, os 53 lemas máis frecuentes constitúen o 49,98% do total das ocorrencias do corpus (os primeiros son palabras gramaticais, igual que sucede noutras linguas románicas). Outro dato que chama a atención é que os 1000 lemas máis frecuentes representan o 80,32% das devanditas ocorrencias. A continuación comenta aspectos relacionados cos 100 lemas máis frecuentes e ofrece unha comparativa destes lemas "... cos 100 primeiros noutras sete linguas románicas. Despois analiza de forma xeral os 500 lemas máis frecuentes.

Remata o capítulo cunha análise atendendo á categoría gramatical dos lemas. Achega datos relevantes sobre os substantivos, adxectivos, verbos e adverbios. Tamén presenta a lista dos 100 lemas máis frecuentes en cada categoría, con comentarios de carácter xeral en cada unha delas. No caso das palabras gramaticais, aborda datos referidos ós artigos, ós pronomes, ós numerais,

ás preposicións e locucións prepositivas e ás conxuncións e locucións conxuntivas. Estes datos consisten fundamentalmente nunha relación das formas que existen e das súas frecuencias.

Respecto ó contido deste libro, hai unha serie de cuestións que merecen algunha consideración. En primeiro lugar, non queda suficientemente claro se se continuará coa informatización do dicionario. Unha ferramenta deste tipo facilitaría a obtención de información por lema ou por forma e posibilitaría igualmente conseguir información de lemas e formas por categorías e por sílabas (que conteñan, que rematen ou que empecen por unha determinada sílaba). Algúns exemplos de materiais neste formato son o de Martínez e García (2004) en castelán e o de Teijido, Pérez e Fraga (2001) en galego.

Outro aspecto que sería interesante desenvolver é a posibilidade de obter información acerca das sílabas, bigramas e trigramas máis frecuentes. Esta información resultaría de utilidade cando fose preciso contemplar as regras da fonotáctica e da grafotáctica, por exemplo, á hora de construír tests de lectura e escritura ou en investigación básica sobre recoñecemento de palabras. Isto é así porque a mencionada información permitiría tanto seleccionar termos atendendo á frecuencia dos compoñentes sinalados como xerar pseudopalabras conforme ó mesmo criterio.

En terceiro lugar, no dicionario de Bugarián non se diferencia entre o vocabulario recollido de fontes orais e de fontes escritas, mentres que, se ocorre en galego coma en inglés, non necesariamente hai unha correspondencia entre ámbolos dous vocabularios (Alabama State Department of Education, 2001; Carlo et al., 2004; Carrol, Davies e Richman, 1971; Corson, 1997; Schonell, Meddleton e Shaw, 1956). Por esta razón, pode ser conveniente actuar con cautela ó seleccionar material destinado a unha aplicación, dependendo de se se centra no plano da lingua oral ou no plano da lingua escrita.

Unha cuarta consideración ten que ver coa impresión que podería desprenderse do título, no sentido de que este dicionario sexa un vocabulario fundamental en lingua

escrita. Non parece que se poida entender como tal vocabulario fundamental para tódolos niveis socioculturais, debido a que non se recolleron datos dos niveis máis desfavorecidos. Por outro lado, é discutible que unha lista que contén en torno a 18.000 lemas, e unha parte importante dos mesmos con frecuencias moi baixas, se poida definir coma un vocabulario fundamental.

Por último, pensamos que caben dúas reflexións finais. A primeira é que este é un traballo de gran interese porque, tendo en conta a relevancia da frecuencia do léxico, resulta necesario dispoñer en galego de dicionarios de frecuencias que sirvan de base para desenvolver materiais destinados a ámbitos coma os que se sinalaron anteriormente, ademais do seu valor no campo da lexicografía. Cómpre salientar que a obra que comentamos cobre un oco existente ó respecto en canto ó léxico adulto e complementase co único dicionario que xa existe relativo ó léxico infantil (Teijido, Pérez e Fraga, 2001). Por isto, o traballo constitúe unha referencia para profesionais vinculados á investigación en léxico, pero tamén para profesionais doutras disciplinas como a pedagogía, a psicoloxía e a logopedia que se adican á didáctica da lingua escrita ou ó tratamento de trastornos da linguaxe, por exemplo. Entre outras calidades, ofrece a de incluír termos non estritamente normativos que responden ó uso en distintos niveis socioculturais, polo menos no plano oral, distintos sectores xeográficos e diferentes rexistros, o que facilita axusta-la intervención a persoas de diferentes estratos en dificultades de linguaxe, por exemplo.

A última reflexión final ten que ver co mérito engadido que supón que se desenvolve de forma individual un traballo que esixe un proceso moi laborioso de execución, desde a selección de textos e creación do corpus ata a lematización.

REFERENCIAS BIBLIOGRÁFICAS

Alabama State Department of Education (2001): *Oral Language and Vocabulary Development: Grades 2-3*. Alabama Reading Initiative. Montgomery, AL: Alabama Reading Initiative (ERIC ED 464.374).

Carlo, Maria S. et al. (2004): "Closing the gap: Addressing vocabulary needs of English-language learners in bilingual and mainstream classrooms", *Reading Research Quarterly* 29 (2), 188-215.

Carroll, John B. / Peter Davies / Barry Richman (1971): *The American heritage Word frequency book*. Boston: Houston Mifflin.

Corson, David (1997): *Using English Words*. New York: Kluwer.

Martínez, Jesús A. / Emma García (2004): *Diccionario. Frecuencias del castellano escrito en niños de 6 a 12 años*. Salamanca: Publicaciones de la Universidad Pontificia.

Sciarone, Abondio G. (1977): *Vocabulario fondamentale della lingua italiana*. Perugia: Guerra Edizioni.

Schonell, Fred J. / Ivor G. Meddleton / B. A. Shaw (1956): *A Study of Oral Vocabulary of Adults*. Brisbane: University of Queensland Press.

Teijido, M^a Carmen / M^a Luz Pérez / Isabel Fraga (2001): *Lexical Primaria. Diccionario de frecuencias léxicas do galego escrito para nenos*. Santiago de Compostela: Servizo de publicacións da USC.

Antonio Estévez Álvarez

FEIXÓ CID, XOSÉ (COORD.) (2007): *DICIONARIO FRASEOLÓXICO CASTELÁN/GALEGO E DE CORRESPONDENCIAS GALEGO/CASTELÁN*. VIGO: EDITORIAL GALAXIA / EDICIONS DO CUMIO, 504 PP.

A obra que aquí se presenta é a segunda das súas características en lingua galega: un dicionario bilingüe castelán-galego específico de fraseoloxía. A primeira foi o *Diccionario Cumio de expresións e frases feitas: castelán-galego* de Xosé A. Pena (2001), mais a diferenza de formato e concepción da fraseoloxía achéganos a de Xosé Feixó Cid máis aos dicionarios bilingües xerais recentemente publicados, que sempre procuran integrar algunha información fraseolóxica nas entradas monolexicais'. Este dicionario enmárcase dentro da colección "Proxecto século21", unha escolma de materiais de lingua galega para uso didáctico e de consulta cunha clara orientación á aprendizaxe, que conxuga o labor de tres editoriais galegas (Editorial Galaxia, Edicións do Cumio e Edi-