O GALEGO FUNDAMENTAL: DICCIONARIO DE FRECUENCIAS

María Xesús Bugarín López

Instituto da Lingua Galega

0. INTRODUCCIÓN

O galego fundamental: diccionario de frecuencias é o título da miña tese de doutoramento presentada na Universidade de Santiago de Compostela (Bugarín 2007)¹. Aínda que o obxectivo inicial era o establecemento do léxico galego fundamental, a magnitude do traballo, as limitacións humanas, cronolóxicas, tecnolóxicas, etc. levaron a acoutar o campo de traballo ao léxico galego de frecuencia, un dos dous elementos base para a construción do léxico fundamental.

A denominación de *Diccionario de frecuencias* vén dada pola presentación das formas obtidas do corpus, xa que foron agrupadas baixo o seu lema correspondente.

1. LÉXICO, LÉXICO FUNDAMENTAL E LÉXICO DE FRECUENCIAS. QUE SON?

Unha das acepcións recollidas no *DRAG* di que léxico é “o conxunto de palabras dunha lingua”. Se falamos dunha lingua viva, este conxunto caracterízase por ser aberto, susceptible de cambios. As palabras nacen, cambian no seu uso (amplían os seus significados, perden algunhas das acepcións...), deixan de ser usadas, etc.

Así a todo, a tendencia xeral que se observa en todas as linguas é que a maioría dos seus usuarios, a pesar da súa maior ou menor riqueza de vocabulario, fan uso unicamente dunha pequena parte dos seus recursos léxicos. Ese sería o léxico fundamental: un núcleo de palabras que todo

1. Agradézolle ao Dr. Manuel González González a oportunidade que me brindou ao darme a coñecer esta liña de investigación, os seus ánimos e a dirección da tese.

membro dunha comunidade debe coñecer para establecer unha comunicación elemental.

Pero, como se chega ao establecemento do léxico fundamental dunha lingua? Xeralmente na construción do léxico fundamental adoitase recorrer á combinación do léxico de frecuencias e do léxico de dispoñibilidade. Do primeiro extraense as formas máis usadas polos falantes mentres que o segundo proporciona aquelas palabras que os falantes actualizan de modo máis espontáneo a través de estímulos axeitados, aínda que non alcancen unha determinada frecuencia.

2. UTILIDADE DOS VOCABULARIOS E DICCIONARIOS DE FRECUENCIAS

Coñecer a frecuencia do léxico dunha lingua é moi útil e o seu establecemento resulta un material básico para o desenvolvemento de novos estudos ou de aplicacións en campos tan diversos como os seis seguintes, por poñer algúns exemplos:

(1) O ensino de lingua para estranxeiros. Nesta área xurdiron as primeiras listas de frecuencias, coas que se buscaba o vocabulario que debían coñecer as persoas que se achegaban a unha lingua distinta á súa establecendo unha gradación para a súa posterior aprendizaxe.

Os estudos de frecuencia coinciden en sinalar que son as palabras gramaticais as que aparecen sempre nos máis altos índices de frecuencia, como tamén en que hai un número moi reducido de palabras que se repiten ao longo de todo o corpus cunha determinada frecuencia, mentres que as formas denominadas *hápax* son as maioritarias. Estes coñecementos son os que permiten o establecemento dunha gradación: elemental ou introdutorio, medio e superior.

(2) Relacionado co campo anterior está a elaboración do material escolar, en lingua galega no noso caso, dun xeito racionalizado, gradual e progresivo. Cal é o léxico que deben aprender e fixar os cativos en cada unha das distintas etapas de escolarización debe ser algo previo ao deseño editorial.

(3) Este traballo tamén é útil na elaboración e corrección de dicionarios. As formas que integra calquera dicionario están expostas a unha serie de procesos que van desde a aparición do termo, a ampliación de significados, a restrición dos mesmos, ata a completa caída en desuso. Todos eses aspectos deben ser tidos en conta á hora da actualización das obras lexicográficas. Outro factor importante son as diferenzas internas entre os vocábulos que conforman o léxico dunha lingua. Algúns termos son

moi frecuentes por posuíren moitas acepcións mentres que outros son raros por apareceren en contextos de uso demasiado especializados ou técnicos. Ónde establezamos o corte fará que o dicionario sexa de tipo básico ou escolar, por exemplo, onde deben estar recollidas as formas máis usuais, fronte a outro concibido coma un tesouro da lingua, xa no outro extremo, onde se pretende recoller todo o léxico desa lingua.

(4) Outro campo de aplicación é o da recuperación de pacientes con perda ou problemas de fala. Cando tratamos de dicir unha palabra que sae do noso campo de léxico habitual e frecuente adoitamos atoparnos con certas dificultades chegando a tela... ‘na punta da lingua’. Ese fenómeno, aínda que máis agravado, afecta a persoas afásicas. Neste grupo tamén poderíamos incluír as persoas con dificultades na lectoescritura.

(5) O galego, como o resto das linguas, debe incorporarse ao mundo das novas tecnoloxías para poder ser útil aos seus falantes. Entre as súas múltiples aplicacións centrémonos, por un instante, no desenvolvemento de recoñecedores de fala. Hoxe en día o home pode establecer diálogos coas máquinas para que así elas leven a cabo as tarefas para as que foron concibidas. Así e todo, as comunicacións adoitan producirse en medios que non están exentos de ruídos e interferencias. Se a máquina dispón de ferramentas como listaxes de léxico acompañadas de datos estatísticos de frecuencia de uso pode chegar a recompoñer aquelas mensaxes dúbidas grazas á lóxica matemática.

(6) Se seguimos dentro do campo informático, o deseño de xogos tamén necesita dunha ferramenta de apoio que indique os distintos niveis de frecuencia e, coa mesma, de proximidade ou afastamento do xogador-falante fronte aos vocábulos-xogo. Un clásico *Aforcado* en soporte electrónico, por exemplo, ten que posuír distintos niveis de dificultade e non errar neles se quere ter éxito e non perder o xogador xa no nivel principiante por exceso de dificultade.

3. LÉXICO DE FRECUENCIA DA LINGUA GALEGA

3.0. Antecedentes

Para a lingua galega en 1986 fíxose un traballo deste tipo que nunca chegou a ver a luz e do que souben por un dos seus autores. Trátase do *Vocabulario de frecuencias da lingua oral*, elaborado por Dolores Sánchez Palomino e Manuel González González. Esta obra, froito dun proxecto de investigación financiado pola Xunta de Galicia, consistiu na creación dun

corpus de 300.000 palabras que permitiu a elaboración dunha lista coas 1.000 máis frecuentes.

3.1. Elaboración do Dicionario de frecuencias da lingua galega

3.1.1. Deseño e creación do corpus

A elaboración do vocabulario de frecuencias requiría primeiramente a creación dun corpus-mostra como punto de partida ou base da nosa investigación.

En primeiro lugar consultáronse e analizáronse distintas obras relacionadas co noso ámbito de traballo. Esas obras versaban maioritariamente sobre linguas románicas –castelán, catalán, francés, italiano, portugués... –, xa que os problemas que o galego podería presentar serían máis facilmente predicibles. Logo procedemos ao deseño do corpus e á posterior recompilación do material que o integraría e que pretendería ser unha representación da lingua galega actual en todas as súas manifestacións. De cara a conseguir o obxectivo proposto, creamos un corpus de carácter textual, con documentos completos, subdividido en dous subcórpora: un de lingua escrita (65%) e outro de lingua oral (35%). O subcorpus de lingua escrita estaba formado por obras literarias (novela, teatro, poesía), científico-técnicas, divulgativas e ensaísticas, en distintas proporcións. Á súa vez, o subcorpus de lingua oral recollía exemplos de intervencións formais (no ámbito do Parlamento ou na Universidade) e non formais (conversas, contos, etc.). Os materiais lingüísticos que o integraban pertencían ao período cronolóxico comprendido entre os anos 1975 e 1999, é dicir, aos últimos 25 anos do pasado século XX.

Na elaboración do corpus ideal hai moitos atrancos, algúns difíciles de resolver, aos que o investigador se debe enfrontar e que converten esta tarefa, segundo os estudosos, nunha utopía. Referímonos a dificultades como o nivel de lingua en que basear o traballo –lingua estándar, lingua común, lingua corrente, dialectos...–, o distinto valor e a porcentaxe dos materiais lingüísticos que se utilizan –escrito/oral, prosa/verso...–, os problemas técnicos –adquisición do material, uso de programas...–, e por qué non, a inexperiencia do deseñador, etc. A precaución ante tales avisos levounos a tentar contrarrestar os previsibles erros de selección de material cun tamaño global do corpus o suficientemente grande para que, dalgún xeito, conseguise facelos mínimos. O noso labor, individual e sen apoio informático, fixo que situáramos o límite ao redor dos 2 millóns de ocorrencias.

Na escolla do material empregado, das proporcións establecidas e do período temporal pesaron, *grosso modo*, tres razóns:

(1) A través das distintas porcentaxes tentamos solucionar unha serie de problemas que podían interferir na ‘calidade’ dos materiais recollidos.

A linguaxe literaria está fortemente influída por dúas variables: (a) a cultura castelá dos escritores (ben sexa a simple escolarización recibida en territorio galego, ben a realización de estadías en universidades de fóra de Galicia, por exemplo), e (b) a acción dos correctores nas editoriais, acción que explicarei un pouco máis adiante.

Do mesmo xeito pode verse afectada a expresión oral formal, en ocasións froito dunha tradución simultánea desde un pensamento formulado orixinariamente en castelán (por exemplo, a lingua dalgúns parlamentarios).

Para contrarrestar a posible ‘desviación’ e empobrecemento lexical, os textos de recolla oral de rexistro non-formal permiten o acceso a un léxico máis propio e variado, xa que non podemos deixar de lado que a cultura rural foi a gran depositaria durante séculos da lingua galega.

(2) A selección cronolóxica veu dada en primeiro lugar pola intención de ‘actualidade’. A primeira característica que buscou sempre esta investigación foi a súa practicidade e posterior aplicabilidade aos problemas que sofre a lingua galega como entidade minoritaria. En segundo lugar, preferiuse que a selección de textos fose dun período non excesivamente entrado no mundo das novas tecnoloxías editoriais. Así, aínda que os materiais chegan ao ano 1999, a maioría son anteriores a 1990. E isto para evitar a excesiva corrección lingüística provocada polo lema “o que non está recollido en dicionarios e correctores non é válido, non existe”, que está ocasionando tan graves eivas no noso patrimonio lingüístico.

(3) Finalmente, e debido á amplitude do corpus que perseguíamos, foinos de esencial axuda valernos de material recompilado noutros proxectos² (xa en formato electrónico), aínda que debeu ser completado ata conseguir o deseño global do noso ‘universo lingüístico’ ideal (idea de representatividade).

2. Quixera agradecer o apoio prestado polo Dr. Antón Santamarina, con quen tiveron a sorte de traballar durante seis anos no seu equipo de Lexicografía, no Instituto da Lingua Galega. Unha parte importante dos textos manexados neste corpus proceden do seu proxecto *Base de datos lexicográfica do galego moderno*.

3.1.2. Tratamento do material do corpus

O respecto á lingua dos textos na súa substancia fonética, morfosintáctica e léxica foi a norma xeral básica. Así e todo, houbo que sometelos a unha serie de adaptacións de cara a unificalos e a poder procesalos despois mediante aplicacións informáticas. Basicamente as modificacións efectuadas afectaron á corrección de erros tipográficos; á eliminación de topónimos, antropónimos, estranxeirismos, etc. a efectos de cómputo; á uniformización no uso das grafías *b/v*, *h/ø*, *i/y*, seguindo criterios etimolóxicos; á representación dos pronomes enclíticos ou ás segundas formas do artigo unidas a eles; á eliminación do uso do guión en formas compostas cando iso era factible ou á súa substitución por un guión baixo que soldaba as partes integrantes como unha única entidade; á unión, tamén con guión baixo, de palabras formantes dun lema único, etc.

Unha vez que foron uniformizados todos os materiais lingüísticos foi posible o emprego de programas informáticos para extraer o léxico do corpus. Primeiro, con axuda de *WordSmith Tools* elaborouse a lista de todas as formas gráficas distintas aparecidas no corpus coa ferramenta *WordList*.

N	Word	Freq.	%	Lemmas
118	ABANDONASEN	2		
119	ABANDONE	1		
120	ABANDONEDES	1		
121	ABANDONEI	4		
122	ABANDONÉMO	1		
123	ABANDONEN	1		
124	ABANDONES	1		
125	ABANDONO	72		
126	ABANDONOS	1		
127	ABANDONOU	12		

Despois, coa ferramenta *Concord*, extraéronse todas as concordancias de cada unha desas formas. Coas concordancias obtivéronse os contextos de uso, fundamentais para a posterior desambiguación.

De seguido, toda esa información foi adaptada e introducida nunha base de datos do programa *Microsoft Access*. Nas diferentes táboas de traballo establecéronse os seguintes campos: contexto anterior, forma, contexto posterior, lema, categoría gramatical, tipo de lingua (escrita/oral), rexistro (coloquial/formal/científico, etc.) e fonte (documento orixe da forma). Unha vez que se revisou e corrixiu a transferencia da información de *Wordsmith* a *Access*, xa todo estaba preparado para comezar o proceso de lematización manual coa introdución dos datos referentes a lema e a categoría nos campos correspondentes.

O proceso de lematización, ou adxudicación dun lema ou forma léxica única a cada unha das ocorrencias aparecidas no corpus, foi o último paso da elaboración do corpus, xunto coa adxudicación dunha categoría gramatical. Para iso seguimos a tendencia maioritaria neste tipo de traballos e tomamos a palabra, ou elemento que aparece separado por espazos en branco, como a unidade de traballo. Para enfrontármonos aos problemas derivados de fenómenos coma a polisemia e homografía, a división e agrupación de cadeas gráficas, e a asignación de palabras a clases debe-

mos establecer previamente unha serie de criterios básicos que guiasen a asignación das ocorrencias aos seus respectivos lemas. A desambiguación viuse facilitada por ser este un traballo totalmente manual. A consulta de todo tipo de obras lexicográficas, principalmente en linguas románicas, axudou na resolución dalgunhas situacións problemáticas.

contexto1	forma	contexto2	lema	cate	lingua	rexis	font
derradeiros anos, baixou a produción polo	abandono	das terras e a forte presión da	ABANDONO	s.m.	escrita	divulg: posvin.	
o rural cara as cidades e o conseguinte	abandono	das defensas (cerres e muros) c	ABANDONO	s.m.	escrita	divulg: cbdrap	
ación neste ámbito amosa o descrédito e	abandono	dos mitos mobilizadores e funde	ABANDONO	s.m.	escrita	literari maider	
sa, xa era bastante para que aceptase o	abandono	da tobeira? Ali tiñan un refuxi	ABANDONO	s.m.	escrita	literari cidfou.	
l. Máis ben dominaba a desintegración, o	abandono	de postulados nouro tempo defi	ABANDONO	s.m.	escrita	literari cidfou.	
persoas cara as cidades co concomitante	abandono	da utilización na agricultura d	ABANDONO	s.m.	escrita	divulg: cbdrap	
an cativos accidentes. As causas dun	abandono	A desaparición da poboación c	ABANDONO	s.m.	escrita	divulg: ferill.tx	
or terra certas teorías sobre o suposto	abandono	do sector público neste campo	ABANDONO	s.m.	oral	formal parlam	
ga dunha rosa e con soño me entrega ao	abandono	[33] A VISIÓN DO GUERREIF	ABANDONO	s.m.	escrita	literari lncor.t	
e marina, esas cousas que se miran, ese	abandono	espantoso, de que as lavadeiras	ABANDONO	s.m.	oral	coloqu gale.tx	
ca dentro da economía neoricardiana que	abandono	dende que me dedico á política	ABANDONAR	v.	escrita	xormal te0002	

3.2. Resultados

3.2.0. Introducción

Un corpus informatizado coma este permite distintas vías de consulta:

(a) polo lema:

ABANDONO		
s.m.	abandono	66
s.m.	abandonos	1
s.m.	abandoo	2
suma:		69

(b) pola forma:

ABANDONO		
s.m.	abandono	66
v.	abandono	6
suma:		72

- (c) pola categoría: listas de substantivos, de adxectivos, de verbos etc.
 (d) outras posibilidades: pódense obter listas froito do cruzamento das distintas variables establecidas no corpus.

3.2.1. O dicionario de frecuencias

A información preséntase ao posible usuario a través de dúas vías distintas:

- (1) Unha lista na que por frecuencia decrecente aparecen os lemas resultantes do proceso de lematización. Cando existe máis dun lema coa mesma frecuencia hai unha ordenación secundaria de carácter alfabético.

6,28% O (<i>art.</i>)	1,45% NON	0,59% SEU
3,76% E	1,28% Ó	0,58% HABER
3,74% DE	1,19% A (<i>prep.</i>)	0,56% ESTAR
3,09% DO	0,94% EN	0,54% DICIR
2,19% QUE (<i>pron.</i>)	0,83% TER	0,53% FACER
1,94% SER (<i>v.</i>)	0,78% PARA	0,53% ME
1,89% UN	0,71% POR	0,52% MOITO
1,77% QUE (<i>conx.</i>)	0,69% O (<i>pron.</i>)	0,51% MÁIS (<i>adv.</i>)
1,56% NO	0,64% LLE	0,51% TODO
1,48% SE (<i>pron.</i>)	0,63% IR	0,50% OU (<i>conx.</i>)

- (2) Outra lista que inclúe, seguindo unha orde alfabética, o lema coas súas ocorrencias, así como a(s) categoría(s) diferenciada(s) xunto coas cifras tanto da(s) ocorrencia(s) como do total de formas agrupadas baixo ese lema.

CARO		
adx.	cara	23
adx.	caras	11
adx.	carísimo	1
adx.	carísimas	1
adx.	carísimo	1
adx.	carísimos	1
adx.	caro	54
adx.	caros	11
		suma: 72

3.2.2. Outras aproximacións

A pesar de que neste punto remataría a elaboración do vocabulario de frecuencias ou, neste caso, dicionario de frecuencias, o feito de traballar sobre un corpus destas características permite múltiples posibilidades de estudo do léxico galego. A continuación, imos facer unha breve demostración dalgunhas aproximacións de carácter inicial aos datos que poden parecer máis relevantes desde o punto de vista da frecuencia:

(1) Abordar os datos de carácter xeral que se extraeron do corpus, é dicir, o número de ocorrencias totais, das formas gráficas distintas, dos lemas resultantes, dos hápax, etc.

(2) Extraer os cen lemas máis frecuentes e comentar algúns aspectos desprendidos da súa observación e da súa comparación cos cen lemas máis frecuentes noutras linguas románicas.

(3) Facer unha aproximación segundo a categoría gramatical. Para exemplificalo imos traballar coas categorías substantivo, adxectivo, verbo e adverbio. Procederemos a elaborar a lista dos cen lemas máis frecuentes de cada categoría, acompañada dalgúns comentarios xerais.

(4) Facer, por exemplo, unha cala dentro dos substantivos buscando lemas relacionados con algúns campos léxicos.

3.2.2.1. Algúns datos de carácter xeral

Un primeiro achegamento aos datos estatísticos totais proporcionou a seguinte información:

- As 2.064.847 ocorrencias que integran o corpus corresponden a 87.273 formas gráficas distintas.
- Despois de lematizadas esas 87.273 formas gráficas distintas, obtivéronse un total de 32.529 lemas diferentes.
- O número de hápax, ou lemas cunha única ocorrencia, ascende a 10.117, ou sexa, un 31,1%, o que indicaría que case un terzo dos lemas ten carácter ocasional.
- A suma das ocorrencias correspondentes aos 53 primeiros lemas en orde de frecuencia representa un 49,98%, o que vén ser case a metade das ocorrencias do corpus.
- Os 1.000 lemas máis frecuentes cobren coas súas ocorrencias o 80,32% do corpus.

3.2.2.2. Os cen primeiros lemas

1	O (art.)	129.256	34	ESTE	9189	67	BO	3493
2	E	77.624	35	CO	9090	68	AQUÍ	3435
3	DE	77.220	36	PERO	8561	69	AGORA	3410
4	DO	63.735	37	EU	7631	70	PÓR	3292
5	QUE (pr.)	45.200	38	XA	7584	71	ONDE	3213
6	SER (v.)	40.150	39	PODER (v.)	7539	72	PASAR	3183
7	UN	38.985	40	OUTRO	7228	73	CHE	3173
8	QUE (cx.)	36.553	41	SE (cx.)	6967	74	CASA	3155
9	NO	32.311	42	ESE	6955	75	POUCO	3044
10	SE (pr.)	30.487	43	DAR	6683	76	COMA	3042
11	NON	30.026	44	EL	6395	77	CHEGAR	3023
12	Ó	26.402	45	DUN	5693	78	TEMPO	3011
13	A (pp.)	24.503	46	MEU	5436	79	GALEGO	2996
14	EN	19.339	47	TAMÉN	5310	80	ENTRE	2958
15	TER	17.199	48	VIR	5250	81	ALÍ	2929
16	PARA	16.017	49	CANDO	5235	82	AQUEL	2878
17	POR	14.573	50	SABER	5220	83	SI (av.)	2854
18	O (pr.)	14.247	51	VER	5123	84	NIN	2806
19	LLE	13.319	52	PORQUE	4906	85	DESTE	2718
20	IR	13.109	53	BEN	4814	86	AÍNDA	2700
21	SEU	12.245	54	MESMO	4463	87	PRIMEIRO	2692
22	HABER	12.061	55	ANO 'tp'	4441	88	SOBRE (pp.)	2680
23	ESTAR	11.636	56	DESPOIS	4425	89	SEÑOR	2668
24	DICIR	11.065	57	ASÍ	4406	90	CHAMAR	2658
25	FACER	11.022	58	QUERER	4216	91	GRANDE	2623
26	ME	10.918	59	POIS	3990	92	FALAR	2608
27	MOITO	10.655	60	NUN	3924	93	TANTO	2587
28	MÁIS (av.)	10.562	61	DOUS	3882	94	COUSA	2543
29	TODO	10.438	62	NOS	3772	95	ATA	2520
30	OU (cx.)	10.418	63	VEZ	3714	96	CUN	2509
31	COMO	10.268	64	DÍA	3660	97	PARTE	2507
32	POLO	9807	65	LEVAR	3633	98	DEIXAR	2502
33	CON	9673	66	SEN	3548	99	HOME	2501
						100	NOSO	2470

Da observación dos cen lemas máis frecuentes no idioma galego pódense tirar, entre outras, as seguintes conclusións:

- Os lemas máis usuais en lingua galega son formas curtas, na súa maioría bisílabas.
- Os lemas máis frecuentes correspóndense coas denominadas palabras gramaticais: o artigo determinado, a conxunción copulativa *e*, e as preposicións *de* e *a*, etc.
- As primeiras palabras léxicas son verbos: *ser, ter, ir, haber, estar, dicir, facer...*
- O primeiro substantivo que aparece é *ano* ‘intervalo de tempo’. Outros substantivos frecuentes son: *día, casa, tempo, señor, vez, cousa, home* e *parte*.
- Os adxectivos máis frecuentes son *grande, galego* e *bo*.
- Os adverbios están a medio camiño entre as formas gramaticais e as léxicas. Os máis frecuentes en orde decrecente son: *non, máis, xa, moito, tamén, ben, así, aquí, despois, agora, onde, alí, si, sempre, aí, aínda*³.

A elaboración desta lista permitiunos confrontar os datos referidos ao idioma galego con outros idiomas da Romania. En 1994 Henri Guiter publicaba o traballo “Rang et dispersion des fréquences lexicales romanes” (Guiter 1994). Nese artigo atópase unha táboa, a número 1, na que, en orde decrecente, aparecen os cen termos máis frecuentes de portugués, castelán, catalán, francés, provenzal, italiano e romanés. Para a correcta comparación foi necesario someter a lista galega a unha ‘corrección’: as contraccións e amálgamas foron desdobradas nos seus formantes orixinais e computáronse con cada un dos lemas base correspondentes.

	gal.	port.	cast.	cat.	fr.	prov.	it.	rom.
1	o (art.)	o - a	el - la	el - la	le - la	lou - la	il -la	de
2	de	de	de	de	de	de	di + da	și (conx.)
3	e	que	y	i	un	que	e	el - ea (pron.)
4	en	e	que	a	être	e	a	avea
5	un	ser	en	un	et	à	essere	în (prep.)
6	a (prep.)	um	a	que (conx.)	à	un	che	un
7	que (pron.)	se	él	se	avoir	èstre	un	fi

3. É interesante comprobar que os adverbios máis habituais fan referencia á situación cronolóxica ou locativa. Esas referencias revélanse como un motivo habitual en calquera produción lingüística reflexo dunha necesidade de concreción situacional no tempo ou no espazo.

	gal.	port.	cast.	cat.	fr.	prov.	it.	rom.
8	ser (v.)	não	ser	ser	il	soun	si	se (pron.)
9	que (conx.)	a (prep.)	su	anar	ne	se (pron.)	in	să (conx.)
10	se (pron.)	com	haber	haver	que (conx.)	avé	non	nu (adv.)
11	non	lhe	con	no	je	dins	avere	la (prep.)
12	por	ter	un	que (pron.)	ce (adx.)	pèr	con	pe (pron.)
13	con	seu	por	per	se	en	per	al (art.)
14	para	em	yo	en	qui	faire	lo	care (pron.)
15	ter	para	este	amb	pas (adv.)	lou (pron.)	mi	mai (adv.)
16	o (pron.)	me	no	fer	ce (pron.)	tout	suo	din
17	lle	dizer	todo	com	en (prep.)	mai	quello	ca (conx.)
18	este	por	como	me	dans	coume	gli	eu
19	ir	eu	más	seu	son	sus	fare	ce (pron.)
20	seu	mais	para	dir	vous	emé	tutto	că (conx.)
21	haber	mas	estar	li	pour	mai	ma	acest
22	estar	como	tener	més	on	me	dire	putea
23	me	estar	ir	hi	plus	vèire	sù	tu
24	dicir	ir	ese	tot	nous	dire	piu	cel, cea
25	facer	haver	hacer	tenir	par	moun	io	face
26	moito	meu	otro	ell	me	ié (pron.)	mio	pentru
27	máis (adv.)	ele	poder	ne (pron.)	mais	veni	questo	cînd (adv.)
28	todo	ver	pero	si (conx.)	sur	ana	altre	noi
29	ou	fazer	decir	però	tout	bèu	come	dar (conx.)
30	como	muito	o	aquest	le (pron.)	en	egli	între
31	ese	todo	ver	poder	pouvoir	toun	ne	a
32	pero	ja	mi	estar	faire	se (conx.)	potere	tot
33	outro	dar	sin	altre	dire	aquéu	vedere	voi
34	el	saber	aquel	aquell	mon	vous	lui	da
35	eu	querer	hombre	ja	comme	quand	vi	vedea
36	xa	outro	ya	veure	avec	pas	volere	prin (prep.)
37	poder (v.)	sem	vida	ho	bien	te	o	om
38	se (conx.)	poder	dar	saber	lui	ié (adv.)	se	cum (conx.)
39	dar	homem	sobre	voler	en (pron.)	éu	sapere	mare
40	meu	este	si	quan	tu	íeu	poco	veni
41	tamén	esse	cuando	perquè	y	noun	stare	aşa (adv.)

	gal.	port.	cast.	cat.	fr.	prov.	it.	rom.
42	vir	quem	mismo	jo	ou	nous	cosa	numai (adv.)
43	cando	aquele	dos	o	si	grand	andare	nici (conx.)
44	saber	casa	nuestro	donar	leur	ounte	venire	după (prep.)
45	ver	vir	vez	home	voir	tant	anche	zice
46	porque	bem	mucho	què	cela	voulè	perchè	spune
47	ben	nem	grande	cosa	où	poudé	me	ști
48	mesmo	lá	tu	molt	sans	bon	ci	dacă
49	ano 'tempo'	então	entre	passar	savoir	autre	cosi	trebui
50	despois	palavra	día	senyor	monsieur	entre	quale	său
51	así	olho	querer	bé	deux	tu	senza	decît
52	querer	quando	poco	te	moi	bèn	ti	doi
53	pois	onde	muy	ara	notre	teni	tanto	meu
54	nos	algun	tan	tornar	même	acò	due	unul (pron.)
55	aquel	ficar	porque	fins (prep.)	vouloir	sant	dovere	iar
56	dous	mesmo	usted	ni	non	nostre	parlare	tot
57	che	vida	ni	mà	grand	o	giorno	lua
58	vez	até	uno	sense	tout	tèms	parere	acel
59	día	ou	creer	día	venir	diéu	ora	zi
60	levar	vez	primer	meu	très	amour	trovare	ochi
61	sen	assim	llegar	casa	encore	falé	qualche	nostru
62	bo	deixar	tiempo	tan	petit	long	quando	acum (adv.)
63	aquí	entrar	mujer	deixar	là	cop	dare	parte
64	agora	isso	año	venir	homme	dous	sentire	domn
65	pór	tão	algún	dos	quelque	plus	tu	unde
66	onde	primeiro	parecer	mateix	peu	pièi	uomo	însă
67	pasar	so	hasta	trobar	devoir	canta	mai	sau
68	casa	também	español	semblar	falloir	jour	cui	mînă
69	pouco	tu	mundo	any	jour	ome	ancora	apoi (adv.)
70	coma	coisa	vivir	arribar	donner	souto	occhio	spre (prep.)
71	chegar	conhecer	pasar	gran	aussi	léu	guardar	bine (adv.)
72	tempo	día	casa	bo	autre	aquí	mano	trece
73	galego	dois	después	pensar	dont	man	parola	loc

	gal.	port.	cast.	cat.	fr.	prov.	it.	rom.
74	entre	ouvir	saber	mig	te	crida	chi	a(i)-ci (adv.)
75	alí	pai	siempre	nos	aller	mèstre	grande	atunci (adv.)
76	algún	pensar	cosa	encara	oui	vièi	amare	pîna (prep.)
77	si (adv.)	sentir	hablar	ull	rien	quau	mettere	cap
78	nin	passar	desde	mai	quand	jouine	parte	an
79	aínda	tempo	sino	parlar	si (adv.)	rèi	poi	sub(t) (prep.)
80	primeiro	coração	donde	tot (adv.)	premier	soulèu	quanto	bun
81	sobre (prep.)	mão	nuevo	això	même	vostre	sempre	rămine
82	señor	olhar	también	sentir	trouver	car	vecchio	sta
83	chamar	te	sólo	entrar	heure	tres	ogni	altul (pron.)
84	grande	velho	pueblo	pas (adv.)	croire	vers	tempo	vreme
85	falar	ainda	ahora	tant	prendre	encaro	prendere	mult
86	tanto	ano	mano	sortir	votre	saupre	dio	pune
87	cousa	hora	mil	hora	moins	rèn	signore	singur (adx.)
88	ata	nada	tres	on	bon	sèuse	vita	urmă
89	parte	pouco	llevar	obrir	celui	blanc	pensare	începe
90	deixar	sobre	bien	vegada	temps	ne	primo	lăsa
91	home	noite	dejar	posar	chose	uei	tuo	cît
92	noso	paixão	ojo	algun	tout ⁴	paure	certo	părea
93	só	pois	encontrar	sempre	alors	plen	solo	viață
94	nada	senhor	salir	vida	vie	passa	molto	peste (prep.)
95	te	grande	quien	així	toujours	toumba	nè	pămînt
96	quedar	porta	nada	damunt	femme	ansin	voi	uita
97	ti	sim	pensar	mirar	parler	ni	allora	trei
98	sempre	alí	aquí	temps	donc	terro	amore	mult
99	medio	aquí	poner	sobre	entre	courre	bene	tot
100	andar	parecer	sentir	entre	jamais	aigo	casa	vorba

Da análise das distintas listas pódense tirar as seguintes conclusións:

- Canto máis frecuente é un lema máis posibilidades hai de coincidencia entre as distintas linguas románicas comparadas.

4. O artigo manexado inclúe tres veces a forma *tout* na columna francesa nas posicións 29, 58 e 92.

- Os lemas máis frecuentes son palabras gramaticais. Salientan de xeito especial os artigos *o* e *un*, a conxunción copulativa *e* e as preposicións *de* e *a*.
- As primeiras palabras léxicas en aparecer son verbos. Os máis frecuentes son *ser*, *haber*, *ter* e *estar*, xusto os máis desementizados.
- Os substantivos comúns ás oito linguas comparadas son: *día*, *tempo* e *home*. Os dous primeiros lemas responden á necesidade humana dunha referencia situacional no tempo ou no espazo na súa produción lingüística (a forma *ano* rexístrase en seis das linguas comparadas). O terceiro lema aglutina distintas acepcións que van desde a máis xenérica ‘ser humano’ ata a máis específica ‘ser humano adulto de xénero masculino’ ou ‘cónxuxe de xénero masculino’.

3.2.2.3. Os cen primeiros lemas por categoría

(a) Os cen substantivos máis frecuentes:

1	ANO 'tp'	4441	34	ESTADO	1018	67	PALABRA	705
2	DÍA	3540	35	EXEMPLO	1001	68	VECÍÑO	700
3	CASA	3152	36	LEI	997	69	VACA	696
4	TEMPO	2889	37	ZONA	979	70	SITUACIÓN	696
5	SEÑOR	2665	38	DEUS	970	71	AMIGO	689
6	VEZ	2552	39	RAPAZ	951	72	LINGUA	668
7	COUSA	2543	40	SOCIEDADE	941	73	ESPECIE	661
8	HOME	2478	41	MES	936	74	HISTORIA	652
9	PARTE	2281	42	NOME	934	75	LIBRO	627
10	CASO	1822	43	PORTA	920	76	VERDADE	625
11	AUGA	1772	44	CONTA	902	77	LUZ	624
12	MULLER	1698	45	ILLA	884	78	LIÑA	624
13	TERRA	1639	46	MOMENTO	881	79	AMOR	621
14	NOITE	1625	47	OBRA	879	80	PÉ	617
15	MAN	1624	48	HORA	877	81	IRMÁN	616
16	VIDA	1602	49	CABEZA	867	82	CAMPO	615
17	XENTE	1545	50	PAI	866	83	SISTEMA	607
18	TRABALLO	1490	51	PUNTO	854	84	RELACIÓN	605
19	LUGAR	1451	52	PROBLEMA	848	85	RAZÓN	604
20	VIÑO	1448	53	DON 'trat.'	830	86	PESETA	599
21	FORMA	1325	54	POBO	829	87	NAI	598
22	FILLO	1246	55	CORPO	829	88	CIDADE	594
23	MEDIO	1238	56	MANEIRA	828	89	SITIO	593
24	OLLO	1230	57	SÉCULO	824	90	EMPRESA	589
25	XEITO	1189	58	LADO	812	91	VOLTA	585
26	NENO	1180	59	PEDRA	808	92	PAN	576
27	MAR	1162	60	PERSOA	796	93	GALEGO	570
28	MUNDO	1156	61	FESTA	790	94	MEDO	568
29	MOZO	1129	62	PRODUCCIÓN	774	95	ALETA	564
30	PAÍS	1070	63	NÚMERO	763	96	A 'letra'	563
31	CAMIÑO	1070	64	VELLO	758	97	TÍO	558
32	TIPO	1069	65	MONTE	715	98	IGREXA	557
33	GRUPO	1024	66	GOBERNO	713	99	POUCO	550
						100	FEITO	549

Se nos achegamos aos datos desde o punto de vista semántico é posible establecer unha serie de grupos atendendo a campos nocionais⁵:

5. En ocasións algún dos lemas pode aparecer en máis dun grupo xa que a ampla frecuencia está en relación directa coa posibilidade de posuír varias acepcións.

- Tempo cronolóxico: *ano, día, tempo, vez, noite, mes, momento, hora, século*.
- Ser humano: *señor, home, muller, xente, neno, mozo, grupo, rapaz, sociedade, don, persoa, vello, veciño, amigo*.
- Familia. Este campo está moi relacionado co anterior e case funcionando coma un subcampo: *fillo, pai, irmán, nai, tío*. Hai que incluír aquí tamén *home* ‘esposo’, *muller* ‘esposa’, *nenos* ‘fillos’, *mozo* ‘noivo’, *rapaz* ‘fillo’ e *vello* ‘avó’.
- Espazo e medio físico: *casa, parte, auga, terra, lugar, mar, mundo, camiño, zona, illa, lado, monte, situación, campo, cidade, sitio*.
- Outros campos que ocupan un lugar importante nesta listaxe son os referidos ao corpo humano (*ollo, cabeza, corpo, lingua*), aos alimentos (*auga, viño, pan*), ás relacións laborais (*traballo, obra, produción, sistema, empresa*), á política (*país, estado, lei, pobo, goberno*), aos sentimentos (*verdade, amor, medo*), á relixión (*deus, igrexa*).

Os campos nocionais documentados permiten afirmar que o mundo xira ao redor do ser humano, que se describe a si mesmo física e psicoloxicamente, que se sitúa no espazo e no tempo, e que se organiza socialmente.

(b) Os cen adxectivos máis frecuentes:

1	GRANDE	2623	34	NEGRO	456	67	EUROPEO	287
2	GALEGO	2150	35	PÚBLICO	452	68	PRINCIPAL	276
3	BO	2096	36	ACTUAL	449	69	PRÓXIMO	268
4	NOVO	1715	37	FORTE	429	70	COÑECIDO	268
5	PEQUENO	1590	38	NATURAL	423	71	DURO	267
6	PROPIO	1260	39	SUPERIOR	418	72	LOCAL	264
7	VELLO	979	40	NECESARIO	407	73	AGRARIO	263
8	MAIOR	892	41	CERTO	399	74	NORMAL	259
9	SOCIAL	877	42	CHEO	396	75	MÍNIMO	259
10	ECONÓMICO	847	43	ESPAÑOL	383	76	LIBRE	257
11	POLÍTICO	843	44	DETERMINADO	368	77	MODERNO	256
12	IMPORTANTE	831	45	SECO	360	78	DEREITO	253
13	MEDIO	776	46	DORSAL	349	79	RICO	253
14	ALTO	763	47	MORTO	348	80	TRADICIONAL	252
15	SANTO	721	48	COMÚN	345	81	XUNTO	252
16	BRANCO	720	49	INDUSTRIAL	339	82	URBANO	249
17	MALO	691	50	SEMELLANTE	339	83	FUNDAMENTAL	248
18	MELLOR	670	51	IGUAL	322	84	MENOR	248
19	SEGUINTE	661	52	PRESENTE	316	85	TOTAL	247
20	ÚLTIMO	652	53	REAL 'res'	314	86	CURTO	236
21	XERAL	651	54	DIFÍCIL	314	87	CONCRETO	232
22	ÚNICO	639	55	DIFERENTE	313	88	FAMILIAR	232
23	LONGO	618	56	VIVO	310	89	PARLAMENTARIO	230
24	RURAL	614	57	ESCURO	307	90	DOADO	228
25	SÓ	580	58	INFERIOR	307	91	MÁXIMO	222
26	POPULAR	571	59	CULTURAL	305	92	POSTERIOR	220
27	CLARO	547	60	HUMANO	303	93	PASADO	220
28	ANTIGO	544	61	DERRADEIRO	301	94	FRÍO	219
29	DISTINTO	524	62	VERDADEIRO	297	95	ABSOLUTO	211
30	POSIBLE	522	63	CIVIL	294	96	CAPAZ	209
31	ANTERIOR	509	64	VERDE	291	97	PRIVADO	208
32	BAIXO	484	65	ESPECIAL	291	98	FINO	205
33	POBRE	479	66	HISTÓRICO	291	99	ENTEIRO	204
						100	CENTRAL	203

Despois de observar os cen adxectivos máis frecuentes pódense facer as seguintes observacións:

- Os adxectivos que teñen moción xenérica son o 61,58%. Os únicos que non responden ao modelo masculino *-o* / feminino *-a* son: *bo*, *só* e *español* (fem. + *-a*).
- Os adxectivos sintéticos poden ser positivos, comparativos ou superlativos. Na relación obsérvase que, ao igual que aparecen as formas positivas *bo*, *grande* e *pequeno*, entre as máis frecuentes tamén están a ese nivel as súas correspondentes comparativas: *mellor*, *maior* e *menor*.
- Importante é tamén o número de adxectivos que atopan na mesma lista outro(s) lema(s) co(s) que poden establecer distintos tipos de relación semántica: *grande/pequeno*; *bo/malo*; *novo/vello*; *maior/menor*; *alto/baixo*; *branco/negro/verde*; *longo/curto*; *rural/urbano*; *claro/escuro*; *antigo/moderno*; *distinto/igual*; *anterior/posterior*; *pobre/rico*; *público/privado*; *actual/histórico*; *superior/inferior*; *morto/vivo*; *industrial/agrario*; *semellante/diferente*; *difícil/doado*; *galego/español*.

(c) Os cen verbos máis frecuentes:

1	SER	39.901	34	METER	1291	67	BAILAR	681
2	TER	17.199	35	CONTAR	1176	68	VENDER	681
3	IR	13.109	36	TRABALLAR	1144	69	DURMIR	672
4	HABER	12.057	37	MORRER	1113	70	ESCRIBIR	669
5	ESTAR	11.636	38	TRAER	1107	71	NACER	664
6	DICIR	11.047	39	VIVIR	1082	72	MANTER	661
7	FACER	11.022	40	EMPEZAR	1052	73	MATAR	657
8	PODER	7477	41	ENTRAR	1043	74	MARCHAR	657
9	DAR	6683	42	COÑECER	972	75	RESULTAR	651
10	VIR	5250	43	TRATAR	927	76	SUPOR	638
11	SABER	5137	44	RIR	927	77	MANDAR	633
12	VER	5123	45	EXISTIR	917	78	CONSIDERAR	631
13	QUERER	4201	46	CRER	914	79	PAGAR	626
14	LEVAR	3633	47	ACABAR	907	80	VALER	624
15	PÓR	3292	48	TIRAR	904	81	SUBIR	617
16	PASAR	3183	49	ABRIR	880	82	PREGUNTAR	615
17	CHEGAR	3023	50	TOMAR	869	83	LER	603
18	CHAMAR	2658	51	CASAR	869	84	COMEZAR	600
19	FALAR	2592	52	TOCAR	854	85	CORRER	599
20	DEIXAR	2502	53	CAER	840	86	ACORDAR	587
21	QUEDAR	2355	54	APARECER	830	87	GUSTAR	584
22	ANDAR	2207	55	SENTIR	822	88	UTILIZAR	567
23	DEBER	2050	56	PRESENTAR	813	89	PERMITIR	560
24	BOTAR	1879	57	FORMAR	802	90	OÍR	556
25	SAÍR	1853	58	CANTAR	800	91	SERVIR	554
26	MIRAR	1844	59	REMATAR	797	92	RECOLLER	548
27	PARECER	1752	60	PERDER	796	93	BAIXAR	542
28	COMER	1662	61	PEDIR	764	94	ESCOITAR	499
29	VOLVER	1542	62	BUSCAR	762	95	REFERIR	499
30	COLLER	1455	63	ENTENDER	762	96	BEBER	498
31	SEGUIR	1447	64	SACAR	729	97	SENTAR	486
32	PENSAR	1420	65	PRODUCIR	708	98	EMPREGAR	471
33	ATOPAR	1304	66	CUMPRIR	685	99	COMPRAR	464
						100	LEMBRAR	457

Unha vez examinada a lista de frecuencia pódense tirar as seguintes observacións:

- O verbo copulativo *ser* ocupa a primeira posición. A pouca distancia sitúase o verbo *estar*.
- Nas primeiras posicións están aqueles verbos que teñen, entre as súas posibles funcións, a de poder desempeñar o papel de verbo auxiliar en perífrases: *ser, ter, ir, haber, estar, poder, dar, vir, querer, levar, chegar, deixar, andar, deber, botar, volver, seguir, empezar~comezar, acabar*. Nas ocorrencias do verbo *ser* inflúe ademais a voz pasiva.
- Moitos dos verbos da listaxe manteñen entre si relacións semánticas de oposición, de complementariedade, etc.: *ir/vir; saír/entrar; vender/comprar; levar/traer; quedar/marchar; nacer/vivir/morrer; coller/deixar; subir/baixar; empezar~comezar/rematar~acabar; buscar/atopar; meter/pór/sacar; preguntar/dicir; etc.*

(d) Os cen adverbios⁶ máis frecuentes:

1	NON	29809	34	ALÁ	660	67	MAÑÁ	181
2	MÁIS	10431	35	FÓRA	615	68	NAMAI	177
3	XA	6998	36	SOAMENTE	602	69	ESPECIALMENTE	169
4	MOITO	6702	37	ARRIBA	590	70	PERFECTAMENTE	164
5	TAMÉN	5310	38	MAL	554	71	ABSOLUTAMENTE	163
6	BEN	3941	39	TAMPOUCO	540	72	SIMPLEMENTE	160
7	ASÍ	3551	40	ALÓ	527	73	CEDO	159
8	AQUÍ	3435	41	ADEMAIS	488	74	EVIDENTEMENTE	156
9	DESPOIS	3403	42	ABAIXO	488	75	NATURALMENTE	152
10	AGORA	3312	43	CANDO	459	76	ENRIBA	141
11	ONDE	3213	44	TANTO	457	77	SEICA	136
12	ALÍ	2929	45	LONXE	440	78	PRACTICAMENTE	136
13	SI	2844	46	ADIANTE	437	79	DISQUE	131
14	SEMPRE	2231	47	ALGO	411	80	CAL	131
15	AÍ	2158	48	IGUAL	386	81	UNICAMENTE	129
16	AÍNDA	1888	49	BASTANTE	349	82	FINALMENTE	127
17	SÓ	1837	50	PRECISA+	331	83	SEGURAMENTE	124
18	COMO	1613	51	INCLUSO	323	84	FUNDAMENTAL+	124
19	HOXE	1557	52	DIANTE	321	85	XERALMENTE	123
20	ENTONCES	1472	53	SEQUERA	304	86	DEMASIADO	117
21	TAN	1369	54	PRIMEIRO	278	87	TALVEZ	113
22	NUNCA	1246	55	ATA	277	88	PROBABLEMENTE	112
23	CASE	1243	56	QUIZAIS	266	89	XAMAI	106
24	LOGO	1161	57	ATRÁS	245	90	CERTAMENTE	105
25	ANTES	1090	58	REALMENTE	242	91	PRONTO	104
26	MESMO	1072	59	TARDE	237	92	VELAQUÍ	103
27	MENOS	1046	60	EFFECTIVA+	234	93	EXACTAMENTE	103
28	DENTRO	1008	61	BAIXO	223	94	CLARAMENTE	101
29	INDA	876	62	TOTAL+	208	95	VELAÍ	97
30	ENTÓN	771	63	AXIÑA	207	96	PRINCIPALMENTE	96
31	DAQUELA	714	64	NORMAL +	194	97	EXCLUSIVAMENTE	95
32	MELLOR	681	65	DIRECTA+	182	98	ABONDO	91
33	POUCO	668	66	ONTE	181	99	MEDIO	91
						100	XUSTAMENTE	90

6. Na lista dos cen adverbios a aparición do símbolo (+) indica que a palabra remata en *-mente*. Esta solución veu motivada por razóns de espazo.

Dos adverbios da lista pódense facer algunhas observacións:

- O adverbio máis frecuente é *non*. As súas ocorrencias supoñen un 23,27% sobre os cen adverbios máis frecuentes, e o 21,89% sobre o total do corpus.
- Contrasta fortemente a alta porcentaxe de aparición do adverbio de negación *non* (29.809) en relación ao adverbio afirmativo *si* (2.844). *Si* e *non* poden actuar como adverbios substitutivos dun enunciado completo: *Vés canda min? –Si* (ou *–Non*). Con todo, mentres que se utiliza o adverbio *non* para indicar o carácter negativo dun predicado (*o neno non come ben*) ou de calquera dos elementos (*unha vida non fácil*), non se utiliza *si* para indicar o carácter afirmativo.
- O segundo adverbio da lista que destaca sobre o resto polas ocorrencias no corpus é o de cantidade *máis*. Ao seu alto índice de aparición contribúe sen dúbida o feito de ser o procedemento habitual de formación do grao comparativo de superioridade.
- Dentro dos adverbios con frecuencia superior a 1.000, o grupo máis numeroso é o formado polos adverbios de tempo: *xa, despois, agora, sempre, aínda, hoxe, entonces, nunca, logo, antes*.
- Dentro dos adverbios de lugar os predominantes son os que pertencen á serie rematada en *–í*: *aquí, alí e aí*, nesta orde. Das series en *–á* (*acá, alá*), en *–ó* (*acó, aló*), só se rexistran entre os cen primeiros lemas as formas correspondentes á 3ª persoa: *alá, aló*, pero non as que indican un maior grao de proximidade: *acá, acó*.
- Dos cen lemas máis frecuentes 26 corresponden a adverbios rematados en *–mente*. O primeiro en aparecer é *soamente*.
- Se eliminamos os adverbios que rematan en *–mente* entrarían na lista os seguintes lemas: *endexamaís, peor, nada, tras, cerca, dabondo, embaixo, denantes, canto, detrás, acolá, quizabes, cecaís, seguro, acaso, xusto, total, acó, alomenos, preto, recén, decote, amais, debaixo, encima, lexos*⁷.

7. Neste traballo non foron eliminados aqueles castelanismos que o autor usa inconscientemente coma se fose palabra galega: *carretera, jueves* etc. Estes casos adoitan rexistrarse maioritariamente nos traballos que recollen material oral, rexistro non-formal.

3.2.2.4. Algúns campos léxicos

Un corpus deste tipo permite tamén o estudo de campos léxicos. Ao falar dos substantivos despuntaron xa os relacionados co tempo cronolóxico, co ser humano e a familia, co espazo e o medio físico. Vexamos, por exemplo, o léxico que xorde, con máis de 100 ocorrencias, ao traballar os seguintes campos: os animais e os alimentos.

- Os animais: *vaca, peixe, besta, can, burro, animal, cabalo, lobo, gando, paxaro, ovella, gato, ave, raposo, porco, boi, galo, galiña, egua, gaivota, becerro, corvo, sardiña, cabra, pito*. Neste apartado habería que ter en conta as variantes dialectais de etimoloxía distinta (*golpe, cuxo, xato*, etc.) con frecuencias inferiores e sumalas para coñecer a realidade.
- Os alimentos: *auga, viño, pan, peixe, leite, carne, millo, uva, ovo, patata, [froito], comida, café, aceite, gran, augardente, azucre, castaña, sal, caldo, trigo, fariña, sardiña, pera, mazá*. Neste campo tamén sería conveniente incluír algúns dos elementos do campo anterior (*ave, porco, galo, galiña* ou *becerro*).

Unha vez extraídos os datos, depurados e ordenados despois de engadir as variantes dialectais con étimos distintos, poderase establecer a maior ou menor frecuencia de uso. Unha lista con esa información serviría de base para a elaboración, por exemplo, de exercicios de aprendizaxe de léxico segundo campos semánticos.

4. CONCLUSIÓN

A través desta breve exposición quixemos dar a coñecer o *Galego fundamental. Dicionario de frecuencias*, sen deixar de lado o corpus-base e a información que del se pode extraer. Cando afrontamos o traballo de investigación o obxectivo que se perseguía era, en suma, contribuír ao mellor coñecemento da vitalidade do léxico da lingua galega, así como poder establecer un punto de partida sólido sobre o que basear a elaboración de material didáctico nos distintos niveis educativos, a extracción de datos útiles para o desenvolvemento de ferramentas de enxeñaría lingüística e de tecnoloxías da fala, etc. Se o conseguimos, aínda que só fose por encetar novos camiños, daríamos por bo o presente labor.