

## **A extracción de información fraseolóxica no CORGA: 3 actividades iniciáticas e 10 propostas para levar á aula<sup>1</sup>**

*The extraction of phraseological information in CORGA: 3 initiatory activities and 10 proposals to take to the classroom*

Eva María Domínguez Noya

Centro Ramón Piñeiro para a investigación en humanidades /  
Universidade de Santiago de Compostela  
edomín@cirp.gal

María Sol López Martínez

Centro Ramón Piñeiro para a investigación en humanidades /  
Universidade de Santiago de Compostela  
msol.lopez@usc.es

Resumo: Nesta contribución describimos polo miúdo o funcionamento da ferramenta de recuperación de información do *Corpus de Referencia do Galego Actual* (CORGA) mediante a realización de 3 actividades iniciáticas deseñadas especificamente para a extracción de unidades fraseolóxicas: *a fume/lume de carozo*, *falar os sete falares* e similares e mais *ir/vir ó/ao pelo*. Ofrecemos logo dez propostas para levar ás aulas nas que se explota o corpus cunha dobre vertente: facilitar que o alumno/usuario adquira as destrezas precisas para a consulta de corpus, en relación coas TIC, e a didáctica de aspectos lingüísticos como son variación, sinonimia, eufemismo, rexistros e niveis de lingua, estruturas atípicas, sentido literal e sentido figurado etc., onde as expresións empregadas se rexistran en textos naturais inmersas na secuencia completa, o que resulta instrutivo e ameno. A modo de conclusión, os corpus constitúen unha fonte valiosa para a obtención de material fraseolóxico, así como para a didáctica da lingua, tanto para a elaboración de exercicios como para a explicación de múltiples fenómenos localizados en textos reais.

Palabras clave: corpus, lingua, didáctica.

*Abstract: In this contribution we briefly describe the operation of the Corpus de Referencia do Galego Actual (CORGA) information retrieval tool through 3 introductory activities specifically designed for the*

---

<sup>1</sup> Data de recepción: 10.05.2019. Data de aceptación: 13.09.2019.

*extraction of phrasemes: a fume/lume de carozo, falar os sete falares and similar, and also ir/vir ó/ao pelo. Afterwards we suggest ten proposals to be practised in the language classrooms in which the corpus is used with a twofold purpose: to enable the student/user to acquire the necessary skills to consult the corpora with regard to ICT, as well as the didactics of linguistic aspects such as variation, synonymy, euphemisms, language registers and levels of usage, atypical structures, literal and figurative sense, etc. Here, the expressions used are found in natural texts embedded in the complete sequence, which seems to be instructive and enjoyable. As a conclusion, corpora represent an important source for the collection of phraseological material, as well as for language didactics, both in the elaboration of exercises and in the explanation of multiple phenomena occurring in real texts.*

*Keywords: corpus, language, didactics.*

## 1. Introducción

Os avances tecnolóxicos que se veñen producindo desde mediados do século XX cambiaron de xeito moi importante a investigación filolóxica e os estudos lingüísticos, xa que permiten dispor de infinidade de recursos para o seu estudo, xa for procurando a perspectiva analítica xa for procurando a didáctica. Entre estas ferramentas, sen dúbida, os corpus lingüísticos ocupan un lugar destacado.

Un corpus, sintetizando a Rojo (2016), é un conxunto de textos naturais, almacenados en formato electrónico, representativos no seu conxunto dunha variedade lingüística e reunidos co propósito de facilitalo seu estudo científico. Trátase, xa que logo, dun recurso electrónico que pon á nosa disposición a consulta de millóns de palabras nun breve intervalo de tempo. Non obstante, os resultados das procuras van depender i) do tipo de textos que se inclúan (non se poderán obter termos propios da economía se o corpus non contén documentos desa área temática), ii) de como se codifique a información neles presente (será imposible procurar unha determinada cadea nun titular se na estrutura interna das novas estes non constan individualizados) e iii) do sistema de consulta que se habilite para realiza-las buscas (por moito que o corpus estea anotado, se o sistema de consultas non permite a recuperación por etiquetas, non poderá procurarse a través deste criterio).

A utilización de corpus na aula é relativamente recente e tense centrado sobre todo no ensino da lingua como L2 (Bailini 2015). Porén, ó noso modo de ver, resulta tamén moi interesante para o ensino da lingua como L1 pois, como indican Sevilla e Barbadillo (2004: 199), “toda lingua no se puede comprender sin la cultura que lleva aparejada”; ou Ferro Ruibal (2016: 14): “o manexo atinado dos refráns e dos ditos (colocacións, locucións e fórmulas) é o nivel máis elevado de dominio dunha lingua; é algo que está por riba do nivel C2; algo que só teñen ordinariamente os nativos; e nativos con moita aldea ou con moito mar”.

Neste sentido, Tolchinsky (2014) establece tres maneiras de aproveita-lo uso de corpus no ensino: de forma indirecta, de forma directa e a través de corpus de aprendices<sup>2</sup>. O noso traballo englobase nos dous primeiros usos pois, por un lado, no uso pedagóxico indirecto, dáselle importancia á aproximación léxica, isto é, ó ensino de colocacións,

<sup>2</sup> É o caso, por exemplo, de CORTEGAL, para o galego, ou de CAES, para o castelán.

expresións idiomáticas e as chamadas unidades prefabricadas, cuxo coñecemento é fundamental para o desenvolvemento de habilidades lingüísticas, tanto en L1 coma en L2, e polo outro, nun uso directo, “se propoñe que profesores y alumnos miren juntos los datos reales, los discutan e intercambien opinións sobre ellos y traten de descubrir juntos las razones que subyacen a los usos detectados” (2014: 14).

Amosando como recuperar datos que conduzan logo á fase de investigación, a nosa achega persegue tamén incentiva-la elaboración de estudos sobre fraseoloxía en corpus galegos, escasos como demostra a consulta en BILEGA coa combinación do tema fraseoloxía (e calquera dos subapartados nos que este se clasifica) e maila palabra CORGA ou TILG como dato adicional.

Así pois, deseguido imos realizar tres actividades deseñadas especialmente para coñecer-lo funcionamento da aplicación de consulta do *Corpus de Referencia do Galego Actual* (CORGA) e mailas posibilidades de recuperación e extracción de información das que dispón o usuario. O obxectivo é coñecer-las opcións que ofrece o sistema para a recuperación de información, especialmente orientada cara ás unidades fraseolóxicas, co fin de adquirirlas destrezas básicas precisas para logo poder realiza-los exercicios que propoñemos ou formular outros. Na segunda parte preséntanse dez exercicios nos que se vincula a consulta ó corpus, a fraseoloxía e en xeral a didáctica de diferentes aspectos lingüísticos, pois inda que a finalidade do corpus é fornecer datos para o estudo da lingua galega dende múltiples perspectivas, é evidente o seu potencial aproveitamento para a didáctica da lingua.

A modo de apéndice, incluímos unha táboa que esquematiza os exercicios propostos, indicando a modalidade de consulta que se emprega, o tipo de resultado, os aspectos computacionais que se tratan coa procura e mailos aspectos lingüísticos que se poden traballar. Engadimos, así mesmo, unha indicación sobre o nivel de dificultade que presenta a actividade, o cal establecemos en función do tipo de información que se precisa para realiza-la procura: consulta por formas ortográficas, dificultade baixa; consulta por lemas, dificultade media; e consulta por etiquetas, dificultade alta.

## 2. O CORGA

### 2.1. Características xerais e composición

O deseño e o desenvolvemento de recursos informáticos que axuden á incorporación da lingua galega ó mundo das novas tecnoloxías da información destaca como un dos obxectivos fundacionais do Centro Ramón Piñeiro para a investigación en humanidades. É neste contexto onde se debe sitúa-lo *Corpus de Referencia do Galego Actual* (CORGA), accesible de balde e sen necesidade de rexistro previo no enderezo <http://corpus.cirp.gal/corga>; un corpus documental aberto, equilibrado e representativo que abrangue cronoloxicamente dende 1975 ata a actualidade e cuxo obxectivo é fornecer datos para o estudo da lingua galega actual dende múltiples perspectivas: léxica, morfolóxica, sintáctica, fraseolóxica, terminolóxica, comunicativa etc.

Na súa versión 3.1 de abril de 2019 o CORGA contén 40.178.271 palabras ortográficas (48.184.012 elementos gramaticais) pertencentes maioritariamente a distintos tipos de textos escritos representativos do galego actual —xornais, revistas, blogs, ensaios e textos de ficción (novelas, relatos curtos, obras de teatro e guións)—, mais tamén inclúe 25 horas de transcricións ortográficas de programas de radio dos 90 —informativo, tertulia, variedades ou programa cultural— nas que se alíña o texto coa voz; transcricións que paulatinamente se irán estendendo ós demais períodos cronolóxicos establecidos. Cómpre salientar, así mesmo, que o corpus, codificado na linguaxe estándar XML (*eXtensible Markup Language*), enriqueceuse coa lematización e etiquetaxe morfosintáctica automáticas dos textos que integra.

Respecto da súa composición, o CORGA distribúese como segue:

- 641 xornais que representan 12.181.778 palabras distribuídas en 41.122 noticias. Inclúe exemplares completos d'*A Nosa Terra*, *A Peneira*, *De Luns a Venres*, *Diario Oficial de Galicia*, *Galicia Hoxe*, *Novas do Eixo Atlántico*, *O Correo Galego*, *O Xornal de Galicia* e *Sermos Galiza*.
- 293 revistas que supoñen 5.514.372 palabras repartidas en 8502 noticias. Engloba números d'*A Voz de Vilalba*, *Cerna*, *Código Cero*, *Consumer Eroski*, *Disquecool*, *Díxitos*, *Entregas de Comunicación Cultural*, *Feiraco*, *Galicia Internacional*, *Gciencia*, *Luzes*, *Man Común*, *Petroglifo*, *SCQ Basket Magazine*, *Sprint motor*, *Teatro do Noroeste*, *Teima*, *Tempos Novos* e *Revista Galega de Economía*.
- 569 libros distribuídos en 2809 documentos que totalizan 21.447.180 palabras repartidas do seguinte xeito<sup>3</sup>:
  - 189 novelas que suman 9.671.594 palabras.
  - 161 ensaios que supoñen 7.832.692 palabras distribuídas en 7 libros de texto (511.898 palabras), 10 memorias (474.923 palabras), 52 artigos científicos (321.505 palabras) e 617 documentos de divulgación (6.524.366 formas).
  - 129 coleccións de relatos que supoñen 2.829.955 formas ortográficas.
  - 91 obras de teatro que conforman un global de 1.312.261 palabras.
- 7 guións televisivos (*A vida por diante*, *As leis de Celavella*, *Matalobos*, *Padre Casares*, *Pazo de familia*, *Serramoura* e *Terra de Miranda*) e 1 incluído nunha colección de relatos (*A noite dos cans doentes*) que suman 78.082 palabras.
- 157 entradas constituídas por 65.864 palabras pertencentes ós blogs *Natureza Dixital*, *Se o vello Sinbad tivese fillos* e *Está en galego!*
- 25 horas de transcricións dos programas radiofónicos *Camiño de Volta*, *Crónica das dúas*, *Diario Cultural*, *Informativo*, *O miradoiro*. *Paco Feixó*, *Pensando en ti* e *Xamón Xamón*. En total supoñen 263.407 palabras.

<sup>3</sup> Ante a imposibilidade de relaciona-los títulos, remitimos á listaxe dos documentos recollidos na base de datos que fornece a aplicación de consulta no enderezo <http://corpus.cirp.gal/corga/datos?subcorpus=etiquetado+automaticamente>.

## 2.2. A aplicación de consulta do CORGA

Para facilitar a recuperación e extracción de datos do corpus deseñouse unha aplicación de consulta, dispoñible en <http://corpus.cirp.gal/corga><sup>4</sup>, que *grasso modo* permite buscar:

- Palabras ortográficas: *cancela, falar, disllo, nel, ao redor do...*
- Elementos gramaticais: *cancela, falar, dis* (incluíndo tódolos casos con pronomes enclíticos nos que entra *dis*: *dilo, disllo, disnolo* etc.), *el* (incluíndo as contraccións *nel* e *del*), *ao redor de* (incluíndo os casos de *ao redor desta, ao redor duns, ao redor das* etc.).
- Lemas: *cancela* (inclúe os casos de *cancela* e *cancelas*), *falar* (tódalas formas do paradigma do verbo *falar*), *el* (tódolos casos do pronome de terceira, masculinos, femininos ou xenéricos, singulares ou plurais, e mailas concorrencias da forma arcaica do artigo determinado), *amigo* (inclúe *amigo, amiga, amigos, amigas, amiguiño, amiguiña, amiguiños, amiguiñas, amiguísima, amiguísimo, amiguísimos, amiguísimas, amig@, amig@s, amigo/a, amiga/o, amigos/as, amigas/os, amigo/amiga, amiga/amigo, amigos/amigas, amigxs*).
- Hiperlemas<sup>5</sup>: *ata* (agrupa os lemas *ata, até, ate, hasta* e *hastra*), *esixir* (agrupa os lemas *esixir* e mais *exixir*, polo que devolve tódalas ocorrencias con calquera das formas verbais de calquera deses dous lemas).
- Clases de palabras: *preposición*, por exemplo.
- Valores das subcategorías gramaticais aplicables en cada caso: *grao, xénero* e *número* no caso dos adxectivos, por exemplo.

A continuación imos desenvolver tres actividades iniciáticas que amosan o funcionamento do sistema de consultas do CORGA, co fin de exemplificar algunhas das opcións de recuperación

<sup>4</sup> As pestanas *Información, Guía* e *Frecuencias* conteñen diversa documentación de utilidade para o usuario. Da presente en *Información* destacámo-los criterios de selección dos documentos que se incorporan ó corpus, a listaxe de documentos que este contén, o historial de versións, os datos globais do corpus (número de versión, data desta e número de palabras ortográficas e mais elementos gramaticais que constitúen o corpus), as frecuencias en función dos criterios de clasificación (data, tipo de texto, medio, orixe, bloque, xénero, área temática e subtipo de documento), a listaxe de preguntas frecuentes, ou mesmo as publicacións que empregaron o CORGA. Pola súa banda, a pestana *Guía* acolle un manual de uso detallado do sistema de consultas e maila documentación de base do proxecto relativa á estruturación, codificación e anotación do corpus, en tanto que baixo a pestana *Frecuencias* ofrécense as listaxes de formas, elementos gramaticais, lemas, hiperlemas e etiquetas, desagregados segundo correspondan ó CORGA xeral ou ó subcorpus desambiguado manualmente que serviu para adestrar-lo etiquetador automático.

<sup>5</sup> O hiperlema nace no noso sistema para agrupar lemas e minimizar así parte dos problemas que xorden pola alta variación ortográfica existente nos textos do CORGA que ten a súa orixe en tres factores: i) a tardía normativización (lémbrese que as primeiras normas oficiais para o galego datan de 1982 e que o corpus contén textos que abranguen dende o ano 1975 ata a actualidade); ii) a convivencia no mesmo territorio de dúas linguas, galego e castelán, o que orixina numerosas interferencias; e iii) a existencia de duplas opcións que posibilita a norma oficial, como poden ser-los sufixos *-aría, -ería* ou *-ble* e *-bel, alcohol* e *alcol, osixeno* e *oxixeno* etc. Deste xeito, o hiperlema agrupa os lemas que presentan variación gráfica no mesmo sufixo (*cafetería-cafetería; amable-amábel*), cambio de conxugación (*combaten-combatir, discorren-discurrir*), presenza de grupo culto (*dictar, construción, rector*) ou supresión (*ditar, construción*) e mesmo vacilacións no vocalismo (*adobo-adubo, entroido-antroido*). Porén, o hiperlema non unifica sufixos diferentes (*alargo* e *alargamento; canadense* e *canadiano*), formas patrimoniais e formas pertencentes a outras linguas (*coxin de seguridade* vs. *airbag; xeonllo, xionllo* vs. *rodilla*) nin sinónimos sen relación de semellanza gráfica (*croucheira, nogueira*). Para máis información véxase Domínguez e López (2017).

que acabamos de enumerar e, asemade, capacita-lo usuario coas destrezas básicas que lle permitan logo levar a cabo os exercicios que se propoñen no apartado último. A primeira, a máis sinxela, persegue mostra-la información que pode obterse, para calquera consulta que se realice, mediante unha procura na modalidade de palabra ortográfica a través da recuperación do modismo *a fume de carozo*. A segunda, coa extracción das variantes que presentan a estrutura tipo do dito *falar os sete falares*, quere evidencia-la importancia dos datos de tipo gramatical para supera-las procuras baseadas en formas gráficas, e a terceira, por fin, contrapón contigüidade e proximidade mediante a consulta da frase verbal *ir/vir ó pelo*.

### 2.2.1. Actividade 1: a modalidade de procura por palabras ortográficas

O sistema de consultas, no que imos centrarnos deseguido e ó que se accede a través da pestana *Buscas*, ofrece o modelo de captación de datos que se observa na figura 1:

Figura 1. Pantalla inicial de captación de datos

Para recupera-las ocorrencias dunha unidade dada o usuario simplemente ten que se dirixir á caixa de texto que amosa a figura 1 e cubrir alí coa expresión cuxo uso desexa consultar, ata un máximo de 5 palabras ou segmentos, pois por defecto a busca realízase na totalidade do corpus. Neste caso cubrimos con *a fume de carozo*.

En cuestión de segundos, modificando a opción *Tipo* do bloque *Resultado*, obtemos:

1) as concordancias con tódolos casos do idiotismo (figura 2) no formato KWIC (*Key Word in Context*):

Figura 2. Pantalla con algunhas das concordancias iniciais

Como vemos, indícase para cada concordancia, así mesmo, o ano e medio no que se localiza (se se sitúa o cursor enriba deste dato emerxe a cabeceira que contén a información bibliográfica completa: *título, autor, editorial* ou *url, ano, medio, orixe, bloque, xénero, subtipo e sección no texto*), e mailo número de ocorrencia que supón no total de resultados (premendo nel accédese a un pequeno contexto).

O sistema ofrece tamén a posibilidade de visualiza-las concordancias resultantes da procura precedidas da cabeceira correspondente (ítem *Cabeceira* do bloque *Resultado* da parte superior da pantalla), mais este é un formato visualmente pesado, polo que por defecto non se amosan e debe se-lo usuario quen o escolla.

2) os datos relativos á súa frecuencia simple: 78 casos nun total de algo máis de 39 millóns e medio de palabras, distribuídos en 52 documentos dos 52.650<sup>6</sup> que integran o corpus, o que supón unha frecuencia normalizada de 2/millón (figura 3):

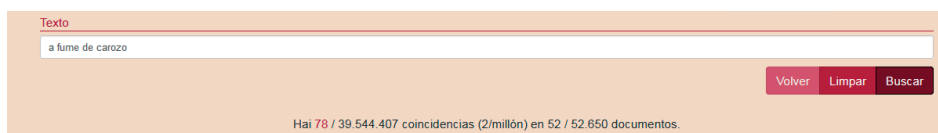


Figura 3. Información sobre a frecuencia simple

3) as frecuencias completas segundo os distintos parámetros de clasificación do corpus (lustros, área temática, medio, orixe, bloque, xénero e subtipo de documento), ben visualizándoas mediante gráficas (o que proporciona unha idea rápida da súa distribución) ben mediante táboas (valores numéricos específicos de coincidencias e documentos sobre o total incorporado por cada un dos criterios, ademais da frecuencia normalizada). Por cuestións de espazo deseguido amosamos só a información relativa ó parámetro temporal (figuras 4 e 5):

Lustro			
	Coincidencias	Documentos	Frec. norm.
1975-1979	1 / 811.519	1 / 117	1/millón
1980-1984	1 / 1.345.949	1 / 143	1/millón
1985-1989	1 / 1.646.519	1 / 120	1/millón
1990-1994	4 / 4.393.605	4 / 386	1/millón
1995-1999	10 / 8.069.491	8 / 12.423	1/millón
2000-2004	18 / 6.058.547	8 / 9.479	3/millón
2005-2009	3 / 7.936.959	3 / 13.877	0/millón
2010-2014	34 / 8.474.477	20 / 14.478	4/millón
2015-2019	6 / 807.341	6 / 1.627	7/millón

Figura 4. Frecuencias completas por lustro

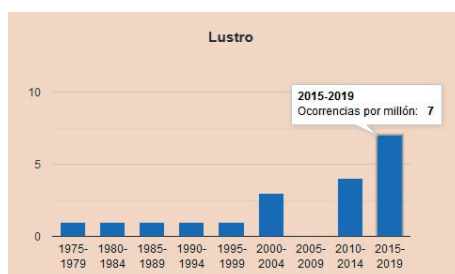


Figura 5. Gráfica por lustro

<sup>6</sup> O alto número de documentos débese a que cada noticia dun xornal ou revista, cada entrada dun blog, cada relato curto dunha colección ou cada prólogo de autoría diferente á da obra xeral son considerados no sistema documentos distintos, pois cada un pode pertencer a un autor diferente, tratar temáticas variadas, ou en definitiva presentar características que os individualizan.



Aquí, se se contrasta a información que proporcionan as dúas figuras para o lustro 2005-2009, obsérvase que a gráfica amosa un baleiro, mentres que a táboa rexistra 3 casos en case 8 millóns de palabras distribuídos en 3 documentos dos case 14.000 que conforman ese período temporal, o que provoca que a frecuencia normalizada non chegue sequer a 1/millón e non estea representada na gráfica por lustros.

Agora ben, hai dous factores que non tivemos en conta na procura realizada e que están impedindo que contemos con datos exactos para analiza-lo modismo. Por un lado, está activada a sensibilidade a maiúsculas e acentos nos parámetros de captación de datos<sup>7</sup>. Trátase sen dúbida dunha funcionalidade moi útil para calquera corpus, e máis se cabe no caso que nos ocupa ante a enorme variación que presentan os textos e que evidencian o proceso de normativización do galego. Porén, está afectando a nosa procura impedindo que se recuperen os casos nos que o dito presenta a preposición con maiúscula por localizarse ó comezo de enunciado. Se deshabilitamos este parámetro, os resultados incrementáanse ata os 83.

Polo outro lado, non tivemos en consideración que existe unha variante, non recollida por exemplo no DRAG, nin en Losada *et alii* (2016) nin tampouco no *Tesouro* do proxecto *Fraseoloxía galega* que se desenvolve no CIRP<sup>8</sup>, en cuxa difusión cremos que inflúen dous feitos: un de cariz fónico e outro de tipo etnográfico. Trátase da variante *a lume de carozo* para a que o CORGA ofrece 11 casos en 10 documentos diferentes, pertencentes a 9 autores distintos, que abranguen cronoloxicamente dende o ano 2000 ata o 2015, e cuxa primeira documentación se rexistra no ano 2000<sup>9</sup>. En primeiro lugar, a semellanza formal entre *fume* e *lume*, para as que cambia só a consoante inicial, provoca a confusión no modismo e, en segundo lugar, poderían contribuír a diluí-la diferenza varios cambios culturais e sociais: i) a despoboación do rural; ii) a perda do contacto directo cos elementos que orixinan o dito: xa non se ven carozos nin se observa, en consecuencia, cando se botan ó lume, como estes emiten un brevísimo fume e inmediatamente prende neles o lume despedindo unha calor rápida, o que leva a construí-la imaxe da rapidez; iii) cambios apreciables nas tarefas agrícolas: cada vez máis o millo ensíllase, proceso no que se tritura toda a planta e non se separa a espiga da palla, e cada vez menos consérvase en espiga, co que si se obterían carozos tras debullalo; e iv) modificacións nos hábitos domésticos xa que, fóra casos moi esporádicos, non se cociña con lume senón con gas ou electricidade, e a fonte de calor na época do frío non é a lareira senón os variados tipos de calefacción.

Pois ben, independentemente das causas que poidan orixina-la segunda variante, é posible recupera-las ocorrencias de ambas, ó tempo, se no lugar do carácter consonántico discrepante utilizámo-lo signo de peche de interrogación ( ? ), dado que ‘?’ actúa como comodín que substitúe un carácter, de xeito que permite que na posición que ocupa apareza calquera letra ou cifra. Se só quixésemos recuperar *fume* e *lume* non valería

<sup>7</sup> Naturalmente, a distinción entre formas con acentos ortográficos e sen eles non é relevante nesta busca, mais si o sería para contrasta-lo emprego de, poñamos por caso, *elite* e *élite*.

<sup>8</sup> <https://www.cirp.es/w3/proxectos/proxecto-fraseoloxia-galega.html>

<sup>9</sup> O *Tesouro Informatizado da Lingua Galega* (TILG) recolle 3 ocorrencias que anticipan o rexistro deste dito en corpus do galego ó ano 1998 (2 casos) e 1999 (1 caso).



cubrir con *?ume*, posto que obteríamos tamén os casos de *cume*, *Eume*, *gume*, *Hume* e *zume*, ou sexa, o que se coñece como *ruído*: resultados que cumpren co criterio de busca pero non responden ó que o usuario desexa localizar. A restrición dos parámetros de consulta e, sobre todo, a fixación da expresión que se procure diminuírá ou anulará o ruído, como ocorre con a *?ume de carozo*, xa que as palabras contiguas restrinxen a *fume* e *lume* as posibilidades de aparición do substantivo inicial.

### 2.2.2. Actividade 2: a modalidade de procura por *elementos gramaticais*

Malia a axuda que nos poden prestar comodíns e operadores booleanos para realiza-las buscas, a modalidade por palabra ortográfica ofrece serias limitacións por canto non pode subtraerse dunha forma gráfica e devén pouco práctica cando na información que se pretende obter existe variación, flexión e irregularidade. Pois ben, para supera-las limitacións que imponen as consultas fundamentadas en formas ortográficas concretas creouse a modalidade de consulta por *elementos gramaticais*, a cal, grazas á etiquetaxe morfosintáctica do corpus, facilita dar un salto cualitativo na recuperación de información ó permitir introducir nos parámetros de consulta información de carácter gramatical. Así, a segunda das actividades de demostración, moito máis complexa, persegue exemplificar a potencialidade do sistema desenvolvido, especialmente útil para a obtención de datos de tipo fraseolóxico<sup>10</sup>, cando prescindimos dun xeito parcial ou total das formas e recorremos só a información de carácter gramatical. Para iso imos recuperar a través da modalidade de elementos gramaticais os fraseoloxismos do tipo *falar os sete falares* que indican que a acción se realiza en cantidade grande.

Tanto o DRAG coma Losada *et alii* (2016) recollen *falar os sete falares* e mais *chorar os sete chorares*, non obstante, aquí o obxectivo é comprobar se o CORGA enriquece estes resultados con outras variantes. Para iso imos prescindir das formas gráficas, ou sexa da expresión, e reducir o dito a etiquetas. Imponse, polo tanto, a selección da modalidade *Elementos gramaticais* no formulario de consulta, o que implica a inmediata aparición, como amosa máis abaixo a figura 6, das 4 caixiñas que dan conta da información gramatical que se asocia a cada unidade do corpus: *elemento gramatical*, *etiqueta*, *lema/hiperlema* e *unidade*.

Antes de avanzarmos cómpre ter presentes dous feitos: 1) o CORGA está etiquetado morfoloxicamente dun xeito automático (Domínguez 2013), polo que pode presentar erros, factor co que se debe contar á hora de formula-las buscas por etiqueta; e 2) o etiquetario<sup>11</sup> definiuse tendo en conta as recomendacións de EAGLES (1996), o camiño percorrido xa na etiquetaxe de corpus en linguas coma o portugués, o catalán ou o

<sup>10</sup> Non hai máis ca pensar no que supón poder reduci-los paralelismos, repeticións ou estruturas bimembres, característicos de numerosas unidades fraseolóxicas, a padróns estruturais abstractos, sen concreción de forma gráfica que limite as procuras.

<sup>11</sup> O tagset ofrécese en dúas vertentes: i) a modo de esquema, en <http://corpus.cirp.gal/xiada/etiquetario/taboa>, nunha táboa que recolle as clases de palabras, os atributos e os valores que xulgamos pertinentes para cada unha delas, xunto cun número que indica a posición que o valor ocupa na cadea resultante; e ii) mediante a exemplificación, accesible en <http://corpus.cirp.gal/xiada/etiquetario/exemplos>, de cada unha das etiquetas, agrupadas por clase de palabra.

castelán e, principalmente, fundamentándonos nos estudos gramaticais dispoñibles para o galego. En total, trala adaptación para dar cabida ás graffias innovadoras que promoven unha linguaxe non sexista<sup>12</sup> (Domínguez e Barcala 2018), o tamaño do tagset sitúase en 453 etiquetas diferentes, resultado de combina-las 18 clases de palabras recoñecidas cos valores dos subtipos e categorías gramaticais pertinentes en cada unha delas: substantivo (18), adxectivo (25), verbo (77), preposición (1), conxunción (2), adverbio (4), artigo (12), demostrativo (13), relativo (15), posesivo (80), indefinido (22), numeral (25), pronome (96), interrogativo-exclamativo (18), locución (4), interxección (1), sinal de puntuación (20) e 20 para a categoría periférica (distínguese aquí a través do subtipo entre abreviatura, sigla, fórmula, símbolo e outros tipos).

Dito isto, premendo no + que está xusto enriba de *Buscar* procedemos a crear 4 liñas de análise, unha por elemento dos que consta o dito *falar os sete falares*, e logo pensamos en como reducir este á súa estrutura morfolóxica. Cómpre, por un lado, deixar aberta a opción de que apareza calquera verbo para indica-la acción, polo que non podemos facer uso da información que nos proporciona o lema (no caso dos verbos a forma en infinitivo); por outro, hai que ter tamén presente que o verbo non ten por que aparecer en infinitivo, pode aparecer unha forma conxugada, o que provoca que non poidamos introducir unha etiqueta completa, mais si podemos quedar co trazo que é común a tódolos casos: trátase dun verbo, polo que no campo *Etiqueta* introducímola etiqueta correspondente: *V*<sup>13</sup>. A continuación, na liña seguinte, completamos coa etiqueta do artigo determinado masculino plural, e logo na fila correspondente á análise do elemento 3 introducimos na caixa do elemento *sete*<sup>14</sup>. Por último, na caixiña do elemento 4, botando man do asterisco ( \* ), comodín que substitúe un, ningún ou varios caracteres, e do operador booleano NON, representado na ferramenta de consulta polo signo de peche de admiración ( ! ), escribimos *\*res!mares*, co que estamos pedindo que ese cuarto elemento remate en *-res* mais que se exceptúen dos resultados os casos nos que ese elemento sexa *mares* (en previsión do ruído que introducirían as probables aparicións de *navegar/percorrer/sucar os sete mares*). En definitiva, a procura descrita formalizaríase como amosa a figura 6:

<sup>12</sup> Deste xeito pretendemos afrontar un problema de etiquetaxe que afecta arredor de 2500 formas e propoñemos unha solución para recoñecelas automaticamente e facilita-la súa recuperación.

<sup>13</sup> Diseñouse un menú amigable que guía o usuario nas alternativas dispoñibles, primeiro segundo a clase de palabra e, logo, segundo os trazos subcategoriais posibles en función das eleccións previas. Así mesmo, na observación dos resultados, xa for nas concordancias xa for no contexto destas, sempre que esta explícita unha etiqueta, a ferramenta desenvólvea con só situar enriba o punteiro do rato, proporcionando a caracterización morfosintáctica da unidade en cuestión nunha lingua comprensible para calquera usuario. Esta dobre actuación, menú despregable para formula-las consultas e cadro de texto emerxente co desenvolvemento da etiqueta nas concordancias ou na análise completa da secuencia visible dende o contexto, permite que un usuario non familiarizado co sistema de etiquetaxe aplicado poida realizar consultas a través da modalidade *elementos gramaticais* e comprender facilmente os resultados.

<sup>14</sup> A procura coa etiqueta relativa a numeral cardinal, por se puidese aparecer un numeral distinto de *sete*, incrementa os casos ata os 222, mais entre eles non se rexistra ningunha variante do noso dito cun numeral diferente, polo que na actividade directamente restrinximos a consulta a este.

**Corpus de Referencia do Galego Actual**

CORGA Información - Búsquedas - Guía - Frecuencias - Contacto - Equipo

**Busca**

Corpus: Etiquetado automa  
 Tipo: Elem. gramaticais

**Sensibilidade**

Acentos: Si  
 Maiúsculas: Si

**Resultado**

Tipo: Concordancias  
 Cabeceira:   
 Ordenación: Coincidencia  
 Tamaño páxina: 50

**Filtros**

Tipo de texto: Calquera  
 Medio: Calquera  
 Área temática: Calquera  
 Sección: Calquera

Dende: 1975  
 Ata: 2016  
 Subárea: Calquera  
 Buscar en: Calquera (12)

**Elementos gramaticais**

Elemento gramatical	V*	Lema / Hiperlema	Unidade
Elemento gramatical	Ddmp	Lema / Hiperlema	Unidade
sete	Etiqueta	Lema / Hiperlema	Unidade
*reslmares	Etiqueta	Lema / Hiperlema	Unidade

Volver Descargar Limpar Buscar

Figura 6. Pantalla de captación de datos gramaticais para o dito tipo *falar os sete falares*

A aplicación devólvenos 17<sup>15</sup> resultados nos que aparecen, como pode observarse na figura 7, os verbos *beber* (*bebendo os sete beberes*), *chorar* (*choraba/chorando/chorar/chorei/chorou os sete chorares*) *dormir* (*durmiu os sete durmires*), *falar* (*falar/falou os sete falares*), *porfiar* (*porfiou os sete porfiarés*) e *pregar* (*pregou os sete pregarés*).

Resultados 1 a 19 de 19

1	1996/Libro	esliveramos en a vila	<b>bebendo os sete beberes</b> e a as dúas de a mañá xa ningún de o fallo podíamos con o cu nin eramos capaces de remener la
2	2000/Libro	non queres morrer de morte matada , desangrado en un camiño ou desfachecándose literalmente o corpo en unha cama mentras	<b>berras os sete matares</b> .
3	1996/Libro	pos como dicía , o Fermoso ficou un día de chuvia e Xoana non puido soportar a súa desaparición.	<b>choraba os sete chorares</b> , choraba tanto que foron necesarias quince almofas para recoller as súas bágoas.
4	2007/Libro	un mes endeus pasei lambendo espaldas de palabras non mensaxarías, unha semana	<b>chorando os sete chorares</b> , un día con día de cubanos, de lentos liras palloletes, esclares, pádicos resendos consorciados para ver
5	2007/Libro	pompón ou burlilla cobiñado de os eledicos que os suscitaban a os permes. notou cada charro de a encalencia	<b>chorando os sete chorares</b> , que nunca a castigarán tanto , que en o Sañón lleve leira en o colo de mensé un moestallio , que
6	2008/Libro	e eu , entretentidos ,	<b>chorando os sete chorares</b> .
7	2010/Libro	endaba a papera de dar a volta cando vin a dona Pelba , a prole de Coñecemento de o Medio ,	<b>chorando os sete chorares</b> a carón de a Mincha.
8	2013/Libro	do como con o mataxado trato quero , amegados sesan todos ! , urtaba o castrate	<b>chorando os sete chorares</b> tal carpoeira en un enteno ...
9	2012/Libro	_ primero	<b>chorar os sete chorares</b> .
10	2013/Libro	e a o peor aínda acabo por vos apesarrar máis e tacer vos	<b>chorar os sete chorares</b> , o preto ver vos nr coma deca agora.
11	2006/Libro	somenin me en o meu cuarto e all	<b>chorei os sete chorares</b> .
12	2008/Libro		<b>chorei os sete chorares</b> aquela noite porque me marchaba a vida enteira en as súas palabras.
13	2002/Libro	qué mal o pasou Lucrécia de aquela vez que	<b>chorou os sete chorares</b> porque querían que o Ciobaneni demostrase en publico que era home abondo para cotar casado con ela , des
14	2006/Libro	o futuro pescador de cina	<b>durmiu os sete durmires</b> e en un trío cufiro de pendor a volada.
15	2007/Libro	daba xeriu ver nos as tarfilas de os soves , trabado de festa para	<b>falar os sete falares</b> .
16	2000/Comal		<b>falou os sete falares</b> e cando , homilde mente , din a minha opinión , puao se seriu , mirou me adverso e desou claro con e
17	1996/Comal	¡ aquí moi ben a conto poso segundo se está a edir en estes días despois de a súa morte.	<b>porfiou os sete porfiarés</b> para agachar a enfermidade que o levou de este mundo .
18	1995/Libro	¡ e li	<b>prega os sete pregarés</b> para que veña lustroso e con a galaña con que se foi !
19	2015/Revista	estes	<b>son os sete investigadores</b> señor de Biocaps e as súas fillas de investigación actuais

Figura 7. Pantalla de resultados para o dito tipo *falar os sete falares*

<sup>15</sup> Cómpre descartar *son os sete investigadores* por ser ruído e *berras os sete matares* por esixir un estudo máis a fondo. Pola súa banda, o TiLG ofrece exemplos con *aguantar* (1), *cavilar* (2), *chorar* (10), *falar* (3), *maquinar* (2), *ponderar* (1), *porfiar* (3), *querer* (1), *revolver* (1), *saber* (1) e *xogar* (1), o que incide na fortaleza do padrón e na variabilidade do verbo base.

Antes de dar por finalizada esta actividade é preciso salientar dous aspectos que se desprenden dos resultados anteriores: i) nas buscas con información gramatical tódalas contraccións e conglomerados de forma verbal con pronome enclítico ou segunda forma do artigo desagréganse nos seus elementos constituintes (por exemplo *en a* na concordancia 1 e *somexín me* na 11); e ii) as marcas de codificación que aplicamos para sinalar fenómenos que entendemos son relevantes dende o punto de vista lingüístico<sup>16</sup>, por exemplo o texto que forma parte dun poema ou aparece nunha táboa, un segmento que está escrito nunha lingua distinta do galego, unha palabra cortada etc., reflíctense nos resultados amosando as palabras afectadas pola marca en cuestión sobre fondo amarelo. Ó pasa-lo cursor sobre algún destes elementos actívanse indicacións aclaratorias sobre a marca en cuestión (*verso*, para a concordancia número 4).

### 2.2.3. Actividade 3: a modalidade de procura por *elementos gramaticais próximos*

A pesar de que as posibilidades de recuperación se incrementan exponencialmente ó poder empregar información de carácter gramatical, a modalidade de procura exemplificada na Actividade 2 caracterízase por requirir-la contigüidade dos elementos implicados, é dicir, estes deben ser sucesivos. Pois ben, para supera-lo atranco que implica a linearidade nas consultas, sexan estas por palabras ortográficas sexan por elementos gramaticais, no sistema de consultas do CORGA habilitouse a modalidade por proximidade. É por isto que, como terceira e última actividade para amosa-las utilidades da ferramenta que serve para extraer datos do CORGA, propoñemos agora recuperar tódolos casos das frases verbais *ir* ou *vir ó/ao pelo* a través da modalidade de procura por elementos gramaticais próximos, para o que debe seleccionarse esa alternativa na sección *Tipo* do bloque *Busca*.

Os datos lingüísticos que cómpre ter en conta para levar a cabo a consulta son: i) o verbo que pode aparecer é tanto *ir* coma *vir*; ambos sumamente irregulares; ii) a forma verbal pode mudar de tempo e persoa; iii) o pronome átono pode estar enclítico, proclítico ou mesmo non aparecer; iv) en caso de aparecer, o pronome átono pode variar e v) a contracción pode grafarse *ó* ou *ao*.

A consulta por *elementos gramaticais* non dá solución ó problema que se formula en iii) e iv): a posible aparición dun pronome en énclice ou en próclise, dado que a linearidade esixe a presenza do elemento que se indica. Para solucionar esa pexa establécense as consultas por proximidade, tanto na modalidade de palabras ortográficas (dúas palabras ou dous segmentos destas) coma de elementos gramaticais (dous lemas, dous hiperlemas, dúas etiquetas, dous elementos gramaticais ou calquera combinación destes), nas que o elemento 1 (e aquí elemento é calquera das opcións que acabamos de mencionar) está con respecto ó elemento 2 á distancia que especifique o usuario (un valor entre 1 e 10, exacto ou inferior ó indicado). A figura 8 amosa o modelo de captación de datos desta modalidade e mailos termos nos que formulamos a consulta para a recuperación dos fraseoloxismos mencionados:

<sup>16</sup> A relación está dispoñible para a súa consulta no enderezo [http://corpus.cirp.gal/corga/etiquetas\\_codificacion](http://corpus.cirp.gal/corga/etiquetas_codificacion), xunto cunha pequena descrición de cada unha e mais un exemplo de uso.

The screenshot shows the 'Corpus de Referencia do Galego Actual' (CORGA) search interface. The search criteria are as follows:

- Busca:** Etiquetado automat
- Corpus:** Elem. gramaticais p
- Sensibilidade:** Acentos: Si; Maiúsculas: Si
- Filtros:** Tipo de texto: Calquera; Medio: Calquera; Área temática: Calquera; Sección: Calquera
- Resultado:** Tipo: Concordancias; Ordenación: Coincidencia; Cabeceira: ; Tamaño páxina: 50
- Elementos gramaticais:** Elemento gramatical: pelo; Etiqueta: Etiqueta; Lema / Hipertema: ir|vir; Unidade: Unidade; s 5

Buttons at the bottom: Volver, Descargar, Limpar, Buscar.

Figura 8. Consulta gramatical por proximidade para a expresión *ir|vir ó/ao pelo*

Esta alternativa permite que se interpoles unidades entre os dous elementos introducidos, co que procurámo-las ocorrencias dos lemas *ir|vir* no elemento 1 que concorran co elemento gramatical *pelo* no elemento 2 a unha distancia máxima de 5 ou menos elementos. Na caixa do lema recorreremos ó operador booleano equivalente a OU, a barra vertical ( | ), de xeito que imos obter tanto os casos correspondentes ó verbo *ir* como a *vir*. Por último, cubrimos con *pelo* na caixa do elemento, dado que neste sentido figurado sempre se emprega en singular.

Entre outras, recuperámo-las seguintes concorrecias: *\_Non, de veras que eses calificativos che van ó pelo. / \_A roupa vénche ó pelo, que andas feito un zalapastro\_ terciou o Nacho\_* / *“A Mariano Rajoy e os seus secuaces do Partido Popular vénlles ao pelo a crise económica. / A canción viña ó pelo para explicarlle a Álvaro un xentilicio que tiña algo confuso. / E digo aguilloadas porque alí falouse moito de touros e vénme moi ó pelo. / E a expedición organizada polos xornalistas de Verín viñéralle ó pelo para as súas intencións*. En contrapartida, recuperáanse tamén resultados que non corresponden ó que queríamos obter, inda que si cumpren cos criterios da busca, como son: *Tiña razón, pero como ía saber eu isto e ademais onde ía deixa-lo pelo. / Vaiche caer o pelo, rapaz, ata o da barba. / Aínda agora aborrezo ir cortar o pelo, co pouco que me queda*. Agora ben, como xa advertimos, canto maior sexa a fixación da estrutura que se persegue, menos ruído se introducirá nos resultados.

### 3. Propostas para levar á aula

Deseguido propoñemos dez exercicios<sup>17</sup>, relacionados con unidades fraseolóxicas ou con estruturas galegas típicas, que serven tanto para que os alumnos<sup>18</sup> adquiran as destrezas precisas para consultar un corpus, como para que os docentes reparen na utilidade dos recursos electrónicos como fonte para a preparación e apoio das súas aulas, e mesmo adquiran eles as destrezas necesarias para a explotación deses recursos. Estas propostas sitúanse no marco educativo da LOMCE<sup>19</sup>, o cal fomenta a interdisciplinabilidade e mailo emprego das TIC, e evidencian a utilidade dos corpus para facilita-la ardua tarefa de creación de materiais orixinais adaptados ás necesidades específicas de cada grupo de alumnos.

Cada exercicio inclúe indicacións para axudar a realizalo e información que debe terse en conta á hora de formula-la consulta. Por último, ofrécense pequenos comentarios sobre os aspectos lingüísticos a tratar que suxiren as buscas, mais só están esbozados, non desenvolto na súa totalidade. O docente deberá adapta-la explicación lingüística de cada aspecto suxerido ó nivel do curso no que aplique as actividades propostas, e tamén, loxicamente, pode adapta-las actividades en función das necesidades do alumnado. Nós suxerímo-lo Bacharelato porque o traballo con unidades fraseolóxicas corresponde ó nivel máis elevado de dominio dunha lingua.

Por último, a selección das unidades fraseolóxicas non responde a un criterio único, senón que combina a necesidade de ser variadas, requirir modalidades de busca e tipo de resultados diverxentes, ofrecer distintos niveis de dificultade e amosar riqueza lingüística nos datos recuperados. Velaquí as veñen.

#### EXERCICIO 1

1. Localiza no corpus os casos da locución *ser uña e carne*
2. Temporalmente, os resultados establecen algunha diferenza en canto ó uso das variantes?

Pista:

- Utiliza a consulta por palabra ortográfica.
- Lembra que existe a variante *unlla*.

Para comezar propoñemos algo sinxelo: trátase de localizar no corpus as concorrencias da locución verbal *ser uña e carne* empregando a consulta por palabra ortográfica e tendo en conta que existe unha variante de *uña* que é *unlla*.

<sup>17</sup> O corpus ofrece a posibilidade de realizar infinidade de actividades posibles, como busca-lo sufixo *-eiro* e clasifica-las formas segundo indiquen lugar ou oficio, investiga-lo uso dos dous alomorfos do artigo, buscar un neoloxismo e indagar se se mantén coa grafía estranxeira ou se adapta á nosa lingua (*tablet/tableta, spoiler/espoiler/destripacantos...*), extrae-las características da lingua oral (repeticións, vacilacións, palabras cortadas, secuencias inacabadas etc.), confrontándoas logo coa elaboración do rexistro escrito etc.

<sup>18</sup> Os exercicios están especialmente orientados cara á aula, mais se se prescinde dos conceptos recollidos na última columna da táboa final, son igualmente válidos para que calquera usuario potencial adquira as habilidades necesarias para extraer información fraseolóxica do CORGA.

<sup>19</sup> Lei Orgánica 8/2013, do 9 de decembro, para a mellora da calidade educativa ([https://www.boe.es/boe\\_gallego/dias/2013/12/10/pdfs/BOE-A-2013-12886-G.pdf](https://www.boe.es/boe_gallego/dias/2013/12/10/pdfs/BOE-A-2013-12886-G.pdf)).

Dado que esta expresión pode darse sen estar precedida do verbo *ser* ou este pode aparecer en distintos tempos (*eran, son, foron...*), para non restrinxírmola recuperación dos resultados debe cubrirse no campo *Texto* co segmento *u\*a e carne*, no cal o asterisco, dado que substitúe un, varios ou ningún carácter, facilita que se extraian os casos tanto de *uña* coma de *unlla*.

Observando os resultados, nunha análise rápida poderíase chegar á conclusión de que existe unha relación entre o uso de cada variante e o período cronolóxico no que se documenta, pois os datos amosan o predominio ata 1994 de *unlla* (5 casos), mentres que na primeira década do século XXI xa só se rexistra *uña* (3 casos). Naturalmente, para confirmar tal aseveración habería que estudar tódolos usos das dúas variantes, e non cinguirse só á locución verbal *ser uña e carne*.

## EXERCICIO 2

Empregando a información que proporciona o sistema de consultas, ou sexa, buscando no corpus e non acudindo á pestana *Frecuencias*, indíquese:

1. Cal é a frecuencia normalizada da locución prepositiva *por mor de*?
2. Documentábase en tódolos lustros representados no CORGA?
3. Hai algún caso cuxa orixe sexa oral?
4. Sen facer unha nova busca, cales son esas ocorrencias do oral?

Pista:

- Emprega as opcións do tipo *Frecuencia* dispoñibles para *Resultados*.

Para responder estas cuestións debemos seleccionar primeiramente a modalidade de *elementos gramaticais* e deseguido cubrir na caixiña do elemento gramatical con *por mor de*. Logo, en *Resultado* cómpre selecciona-la alternativa *Frecuencia simple*, pois aquí as concordancias non interesan, senón que o dato que se procura é o da frecuencia normalizada, ou sexa, o dato que nos permite ter unha visión xeral do peso estatístico dun elemento e que, dadas as baixas frecuencias dos elementos lingüísticos, adoita expresarse en porcentaxes por millón, que resulta ser neste caso de 43/millón.

Tamén se podería emprega-la consulta por palabra ortográfica, mais habería que ter en conta que a preposición da coda entra en numerosas contraccións, o que obriga a emprega-lo asterisco despois do *d* para cubri-las posibles combinacións: *por mor d\**. Nesta modalidade a frecuencia normalizada ofrece un resultado de 50/millón. A diferenza entre as dúas cifras estriba en que nun caso o cómputo realízase sobre 48 millóns e pico, os elementos gramaticais, e no outro sobre algo máis de 39 millóns, as palabras ortográficas susceptibles de ser recuperadas<sup>20</sup>.

Por outra banda e malia ser unha locución segura, isto é, unha unidade multipalabra que non presenta ambigüidade á hora de identificala, pois se aparece o segmento *por*

<sup>20</sup> Inda que a versión 3.1 do corpus consta de 40,2 millóns de formas ortográficas, para as frecuencias non se contabilizan unhas 600.000 palabras sobre as que non é posible facer buscas, ou sexa, aquelas que están caracterizadas como pertencentes a outras linguas ou as que aparecen en notas bibliográficas, de aí a cifra de 39 millóns.



*mor de* estamos sempre ante unha locución preposicional, os resultados poden variar lixeiramente en función da modalidade de consulta que empreguemos na procura. Así, por palabra ortográfica, se non está deshabilitada a distinción entre maiúsculas e minúsculas, recupéranse menos casos ca por elementos gramaticais, dado que na versión etiquetada calquera maiúscula a inicio de secuencia que non corresponda a un nome propio ou unha sigla transfórmase en minúscula no elemento gramatical, non así na unidade<sup>21</sup>. É dicir, as secuencias que comece con *Por mor de* só serán contabilizadas se a consulta se realiza por elementos gramaticais ou se na modalidade de palabras ortográficas se anula a sensibilidade a maiúsculas e minúsculas.

Para responde-la segunda pregunta que formulamos neste exercicio hai que cambia-la opción de *Frecuencia simple* pola de *Frecuencia completa*, e nas táboas que se nos ofrecen observa-las frecuencias por lustros e procedencia dos textos, co que obtémo-los outros dous datos que pediamos: si, rexístrase en tódolos lustros e hai dous casos no oral. Aquí debe facerse fincapé no diferente peso desta locución prepositiva no rexistro escrito (43/millón) fronte ó rexistro oral (7/millón).

Para a última das preguntas o lector non conta coa axuda de nada que viramos na demostración realizada ata o de agora, só coa súa intuición ou o coñecemento previo do emprego do corpus. Porén, chegar ás ocorrencias do oral desde a pantalla de frecuencias completas é sinxelo, posto que concordancias e frecuencias están interconectadas. Así, se prememos no 2 de *Oral* que aparece destacado na información das frecuencias para o parámetro *Orixe*, visualizámo-las dúas ocorrencias<sup>22</sup>, unha emitida por Carlos Blanco nun *Diario Cultural* (1) e a outra pronunciada nun informativo (2), ámbalas dúas pertencentes a un rexistro culto e formal:

- (1) Carlos Blanco: esta semana <pausa/> de novo <pausa/> con nós <pausa/> Emma Couceiro *por mor de* os festivos habidos a semana anterior [CORGA: Diario Cultural. 30/12/1996]
- (2) Xosé Lois Rúa: o chamamento a a doación <pausa/> a a doazón de sangue <pausa/> fai se máis importante en o verán <pausa/> cando descenden os voluntarios e sen embargo incrementan se as necesidades <pausa/> *por mor de* os accidentes [CORGA: Informativo 14:00. 28/07/1997]

### EXERCICIO 3

1. Que variantes se documentan no CORGA para a intensificación do fastidio coa expresión *estar ata* tipificada por *estar ata as orellas*?

<sup>21</sup> A diferenciación entre elemento gramatical (*por mor de*) e a unidade (*por mor dunha* ou *Por mor desta*) resulta de suma utilidade, por exemplo, para discriminar automaticamente nos usos dun marcador discursivo entre a posición inicial de secuencia e unha interior. O sistema ofrece tamén a alternativa de emprega-la consulta por palabras ortográficas e a sensibilidade a maiúsculas e minúsculas, mais esta posúe menos precisión por canto se depende da grafía da forma para a súa identificación: se se busca *por mor de* só recuperará eses casos e non os de *por mor da*, *por mor dunha*, *por mor destes* etc.

<sup>22</sup> Repárese en que nas buscas por elementos gramaticais as concordancias resultantes ofrécenno-las contraccións e amalgamas de formas verbais con enclíticos desagregados nos seus elementos constituintes: *por mor de os*, *a a*, *fai se*, *en o*, *incrementan se*. Para ve-lo texto en grafía convencional debe acudirse ó contexto.

2. Pódense distinguir entre os resultados variantes situacionais?
3. Que substantivos se sitúan na coda? Pódense relacionar con algún campo semántico concreto?

Pista:

- Combina a información de lema, elemento gramatical e clases de palabra e subtipos, respectivamente.

Empregaremos novamente a consulta por elementos gramaticais, pero agora serán 4 sucesivos, que cubriremos como segue:

Liña do elemento 1: lema *estar*

Liña do elemento 2: hiperlema */ata*

Liña do elemento 3: etiqueta Dd\*

Liña do elemento 4: etiqueta Sc\*

Con esta busca imos recuperar tódolos casos da expresión, independentemente do tempo e persoa no que apareza a forma verbal (*estou, estaba, están, estás, estar*), sen importar se a variante para a preposición é *ata, até* ou mesmo outra variante non normativa como *hasta, hastra* e *ate*, dado que se agrupan todas elas baixo o hiperlema *ata*, e por último, como mostra máxima da abstracción da consulta, a busca pola clase de palabra e o subtipo, artigo determinado primeiro e substantivo común despois, o que nos devolve tódolos casos que se acubillan baixo o padrón estrutural representado por *estar ata as orellas*. Extractámoslos deseguido ordenados polo substantivo coa indicación entre parénteses do número de ocorrencias que hai para cada caso reducindo as posibles variantes á súa estrutura base:

*estar ata o carallo* (1) | *estar ata a cintura* (1) | *estar ata o cocote* (1) | *estar ata os collóns* (6) | *estar ata o cu* (2) | *estar ata o curuto* (1) | *estar ata os fociños* (1) | *estar ata o gorro* (1) | *estar ata o moco* (1) | *estar ata o nabo* (2) | *estar ata as narices* (1) | *estar ata os ollos* (1) | *estar ata as orellas* (1) | *estar ata as pelotas* (1) | *estar ata a punta do paraugas* (1) | *estar ata os troles* (1)

Os resultados desta consulta constatan a riqueza léxica que posúe o galego con esta significación, posto que son numerosas as variantes que nos devolven os textos. Mais esta consulta non só pode servir para traballa-la sinonimia e o campo semántico (partes do corpo, por exemplo) senón tamén para ilustra-los distintos rexistros ou variedades situacionais, dado que fronte á sutileza demostrada en *estar ata a punta do paraugas*, podemos opoñe-la ordinariez de *estar ata os collóns* ou *ata o cu* e os vulgarismos *estar ata o carallo*, *ata as pelotas* ou *ata o nabo*, pasando polas zonas corporais que indican localización superior: *estar ata as orellas*, *ata os ollos*, *ata os fociños*, *estar ata o cocote*, *estar ata o curuto*; outras nas que se empregan eufemismos: *estar ata o nabo*; *estar ata as pelotas*, *estar ata os troles*; ou substantivos que, sen ser partes do corpo, empréganse na parte alta da cabeza: *estar ata o gorro*.

Así mesmo, o emprego do artigo feminino con *narices* permite abordar aspectos sociolingüísticos como son o contacto de linguas ou os distintos tipos de interferencias

que este orixina, e pode aproveitarse ademais como pretexto para suscitar un debate no que se contrapoña norma e uso.

#### EXERCICIO 4

1. Comproba a vitalidade da estrutura focalizada tradicional galega tipo *Aí a está, aí os veñen* etc.
2. Que problema supón para a sintaxe *aí os veñen*?
3. Sen realizar unha nova busca, amosa unha concordancia que se documente nun guión.

Pista:

- Combina respectivamente elementos gramaticais (*aí, alí, aquí...*) con etiqueta de terceira acusativo e lemas (*andar, estar, ir, vir*).
- Fai uso do operador booleano ( | ), equivalente a OU.

Neste exercicio propoñemos comproba-la vitalidade dunha estrutura que non podemos considerar fraseolóxica mais si tipicamente galega. É este ademais un caso problemático para a sintaxe pois nel atopamos pronomes de acusativo con verbos intransitivos. Trátase da aparición de pronomes de acusativo con verbos intransitivos en estruturas focalizadas con adverbios de lugar, tipo *aí a vén*.

Para extrae-los casos, introducimos na caixiña do lema ou do elemento gramatical do primeiro elemento os adverbios *aquí|aí|alí|velaí|velaquí|alá|acólá*<sup>23</sup> botando man do carácter que converte a busca en expresión regular, a barra vertical ( | ) que equivale a un OU, de xeito que poida aparecer calquera deles. No segundo elemento recorreremos ó menú de introdución de etiquetas e seleccionámo-los trazos correspondentes a pronome acusativo de terceira. Por último, na terceira liña introducimos no lema os verbos *andar|estar|ir|vir* e prememos en *Buscar*. O sistema devólvenos 111 concorrencias, en singular e plural, co verbo en presente, pero tamén en pasado, o cal nos indica que se trata dunha construción relativamente usual no corpus.

Para ve-la súa vitalidade imos acudir nesta ocasión ás gráficas, para o que optamos no bloque *Resultados* pola alternativa *Frecuencias completas* e logo seleccionamos *Gráficas*, onde cunha simple ollada constatamos que esa construción se usa ó longo de tódolos períodos, que a partir do lustro 90-94 aumenta o número de ocorrencias, e que estas se dan na escrita, mais non esquezamos que polo momento a parte oral non está representada en tódolos lustros. Se poñemos o cursor enriba dunha das columnas ou dun cacho da torta aparece o número de ocorrencias e a porcentaxe que representa sobre o total, pero podemos ir máis alá e ve-los exemplos concretos que se localizan nun dos parámetros: por exemplo, as dúas concordancias presentes no medio audiovisual pertencentes a un guión da serie *Padre Casares*.

- (3) Moncha: *Aí o vén*.
- (4) Miñato: *Quietos todos... aí os están*.

Porén, podemos ir máis alá e incluso ve-lo contexto e maila análise completa da secuencia seleccionada:

<sup>23</sup> E todos aqueles que creamos poden concorrer nesta construción.

<b>Título:</b> Pedro Casares. Tempada 1, episodio 1. O milagre de Santo Antón de Lourido Bloque: Ficción Sección no texto: corpo		<b>Autor:</b> Cabreira, Araceli G. / Portela, Carlos Medio: Audiovisual Xénero: Dramático		<b>Ano:</b> 2008 Orixe: Escrita Subtipo: Cución		
Ela: É eu coas ovelas sen gardar ... Acoutación: Óese un coche , póñense alerta . Miñato: Quietos todos ... aí os están .						
quietos	todos	...	aí	os	están	.
A0mp	Imp	Q...	Wh	Raa3mp	Vp130p	Q.
quieto	todo	...	aí	o	estar	.
Acoutación: Vese chegar o coche de Amelia . Acoutación: Eles recompóñense , e colócanse pra recibilos , estrados .						

Figura 9. Mostra do contexto e análise correspondente a unha concordancia

## EXERCICIO 5

1. Identifica as estruturas principais que se poden dar con *poñer/pór verde*.
2. Cales corresponden a usos literais e cales a usos figurados?

Pista:

- Emprega a consulta por elementos gramaticais próximos.

Propoñemos agora un exercicio para practica-la modalidade de consulta por elementos gramaticais próximos cuxo obxectivo é identifica-las diferentes estruturas que podemos atopar coa combinación de *poñer/pór* e mais *verde*, que nos servirán así mesmo para distinguir entre uso literal e uso figurado. Cómpre emprega-la busca por proximidade porque a estrutura non é lineal, xa que entre o verbo e o adxectivo caben diversos elementos, entre eles pronomes enclíticos, co que a consulta por elementos gramaticais sucesivos sería moi restritiva e non amosaría as distintas combinacións que poden darse.

Así, no elemento 1 cubriremos co lema *poñer|pór* ou co hiperlema */poñer*; e no elemento 2 co lema *verde*; a distancia escollida será de  $\leq 5$ .

As estruturas que se rexistran no corpus resúmense nas seguintes:

- (a) poñer algo de cor verde [unha prenda/un obxecto]
- (b) poñerse verde ~ poñerse en verde [o semáforo]
- (c) poñer verde a alguén (criticar)

## EXERCICIO 6

1. Localiza as variantes de expresión tipificadas por *onde comen dous comen tres*.
2. Que verbos se empregan?
3. Predomina algún numeral?

Pistas:

- Emprega a consulta por elementos gramaticais sucesivos.
- Ten en conta o paralelismo verbal e a posible variación do numeral.

Propoñemos localizar aquí tódolos casos da coñecidísima *onde comen dous comen tres*, para o que debemos ter en conta que pode varia-lo verbo empregado e mailo numeral. A consulta por elementos gramaticais sucesivos novamente volve na nosa axuda. Nela

imos aproveita-lo paralelismo verbal presente nas dúas partes e maila etiqueta de numeral cardinal para cubrir como segue:

Liña do elemento 1: elemento *onde*

Liña do elemento 2: etiqueta *V*\*<sup>24</sup>

Liña do elemento 3: etiqueta *Nc*\*

Liña do elemento 4: etiqueta *V*\*

Liña do elemento 5: etiqueta *Nc*\*

Unha vez máis, a fraseoloxía e o corpus sorpréndennos, pois a aplicación devólvenos seis casos que ilustran esta unidade fraseolóxica, mais cada un dos resultados ciméntase sobre un verbo diferente e un numeral distinto: *cabere, collere, comere, dormire, malcomere* e *sachere*. Recollémolos deseguido:

- (5) Roi: *Onde caben seis caben sesenta*, un cero a maiores... [CORGA: Xosé Luís Martínez Pereiro, *Carne de estatua*, Bahía Edicións, 2001]
- (6) Pero o concello parece compartir os principios fundamentais da malfadada tradición constructora viguesa, esa que di que *onde collen catro collen oito*, meténdolle un pouco máis por aquí e outro pouco por alá. [CORGA: Damián Villalaín, *Tempos Novos / TN1998-01/26*, Atlántica de Información e Comunicación de Galicia, 1998]
- (7) Gurgullo: *Xa sabes aquilo de que onde comen tres comen catro*. [CORGA: Agustín Magán, *Farsa Plautina. Sobre textos de Tito Maccio Plauto*, Sotelo Blanco, 1993]
- (8) *Fai caso, nena; onde dormen dez dormen quince ou vinte*; anda, ven e trainos. [CORGA: Xosé Vázquez Pintor, *A memoria do boi*, Edicións Xerais de Galicia, 2001]
- (9) Os albaneses, pobres de solemnidade, acollen sen fanfarra centos de miles de refuxiados nas súas casas: *onde malcomían catro malcomen oito*, pero malcomen todos. [CORGA: Carlos Mella, *O Correo Galego / CG1999-06-01/6*, Editorial Compostela, 1999]
- (10) E asemade as vosas mercedes, Don Bernaldino, Don Xaquín e vós, Sancho, que *onde sachan dous sachan cinco* e os meus serventes axudarán e malo será que non deamos cunha mina ou dúas para xantarmos uns centos desas deliciosas ostras montesas. [CORGA: Lalo Vázquez Xil, *A da alba sería*, Edicións do Cumio, 1996]

## EXERCICIO 7

1. Relaciona as seguintes frases verbais co significado axeitado e documenta o seu uso no galego actual mediante un exemplo extraído do CORGA.

a) *Disuadir, emborracharse levemente, marearse, coller xuízo*

---

<sup>24</sup> Se introducimos unicamente os lemas *comere, cabere* e *collere* estaríamos impedindo a aparición doutros verbos.

b) *Subir á cabeza, leva-la cabeza, quitar da cabeza, senta-la cabeza*

Pista:

- Emprega nas consultas a modalidade por elementos gramaticais próximos.
- Basea as buscas nos lemas dos modismos.

Aquí trataríase de que primeiramente o alumnado relacionase *disuadir* con *quitar da cabeza*, *emborracharse levemente* con *subir á cabeza*, *marearse* con *levar* (a alguén) *a cabeza* e por último *coller xuízo* con *senta-la cabeza*. Logo, mediante a consulta por elementos gramaticais próximos e facendo uso do lema verbal e substantivo respectivos a unha distancia igual ou inferior a 4 posicións (na liña do elemento 1, en cadansúa consulta, *quitar*, *subir*, *levar* e *sentar*; mentres que na liña do elemento 2 manteríase en tódalas ocasións *cabeza*), extraeríase un exemplo de cada significación. Serven de mostra as seguintes concorrencias:

- (11) Iso non hai quen mo *quite da cabeza* por moita terapia que me queiran dar. [CORGA: Francisco Galván, *As crendas de San Porfirio. Diario terapéutico dun esquizofrénico*, Edicións Morgante, 2012]
- (12) O viño préndeseme nas pernas, pero non me *sobe á cabeza*. [CORGA: Xosé Manuel Martínez Oca, *Beiramar*, Edicións Xerais de Galicia, 1983]
- (13) Como vos contaba, si, trae a cadeira, mellor sentado, é que me *leva a cabeza*, Irene e mais eu criámonos xuntos, viviamos porta con porta. [CORGA: Xesús Manuel Marcos, *Terra queimada / Febre*, Edicións Laiovento, 1997]
- (14) *Senta a cabeza*, Xavier, dinche os máis condescendentes, como se palmeasen na túa caluga co consello, monta unha familia, rapaz, que xa vas tendo idade. [CORGA: Manuel Portas, *Denso recendo a salgado*, Edicións Xerais de Galicia, 2010]

Este é un exercicio especialmente útil para estuda-lo réxime verbal e contrapoñe-las funcións sintácticas que desempeña *cabeza* en cada caso: Suxeito (*levar* [a alguén] *a cabeza*), CD (*senta-la cabeza*), Complemento preposicional (*quitar da cabeza*) e, por último, CCL (*subir á cabeza*).

## EXERCICIO 8

1. Localiza as ocorrencias de *meterse en camisa(s) de once varas* e ordena os resultados por data.

Pista:

- Usa a consulta por proximidade cos dous lemas substantivos.

Se nos parámetros de ordenación se mantén o que ofrece por defecto a aplicación, isto é, por coincidencia, debe terse en conta que ó realiza-la procura polos lemas, dado que nos resultados *camisa* amosa variación numérica (2 casos de *camisa* e 10 de *camisas*), primeiro ordénanse os resultados alfabeticamente (en primeiro termo os de *camisa* e posteriormente os de *camisas*), e logo como segundo criterio, ante formas iguais, as concordancias organízanse cronoloxicamente en orde descendente, das máis antigas ás máis recentes. É dicir, só unha vez que estamos ante casos iguais entra en xogo a ordenación por data que solicitamos. Naturalmente, os casos non son moitos e aínda

sen orde temporal observamos que se sitúan entre 1994 e 2002, polo que rapidamente emerxe unha conclusión inicial: nos últimos 12<sup>25</sup> anos non se rexistra ningún caso.

Por outra banda, o CORGA ofrécenos 11 casos nos que *camisa(s) de once varas* é o complemento preposicional de *meter*. Agora ben, é interesante así mesmo destaca-la concordancia que recollemos deseguido (15), na que *camisas de once varas* emprégase como sinónimo de *leas*, sen o verbo *meter*, presente nos restantes casos:

- (15) O resto sonlle lerias, *camisas de once varas*, adopcións fallidas, soños esnaquizados. [CORGA: Alfredo Conde, *O Correo Galego* / CG2002-05-31/25, Editorial Compostela, 2002]

## EXERCICIO 9

1. Partindo da estrutura bimembre *tal x tal z*, que variantes documentas para a unidade fraseolóxica castelá *de tal palo tal astilla*?

Pista:

- Emprega a consulta por proximidade cos dous lemas *tal*.
- Restrinxe a distancia a 3 posicións.

Nesta ocasión, empregando a consulta por elementos gramaticais próximos e tendo en conta a repetición de *tal* nunha estrutura bimembre de proximidade igual ou inferior a 3, propoñemos localiza-las expresións galegas que se empregan no CORGA para a expresión fraseolóxica castelá *de tal palo tal astilla*.

As variantes que se atopan son as seguintes:

*de tal bicho, tal cría | tal nai, tal filla | a tal señor, tal coche | de tal pano, tal saia | de tal semente, tal árbore | de tal árbore, tal semente | de tal pau, tal estela*

## EXERCICIO 10

1. Localiza no CORGA unidades fraseolóxicas que indiquen a *nula probabilidade de que algo suceda*, coa estrutura seguinte:

*cando* + artigo determinado + substantivo común + verbo en presente de subxuntivo + substantivo común

Pista:

- Emprega a consulta por elementos gramaticais sucesivos.
- Combina a información de lema ou elemento gramatical coas etiquetas que se indican na estrutura.

As variantes que rexistra o corpus para esta estrutura exemplificámolas a continuación extractando un caso de cada variante, inda que en realidade, se prescindimos da flexión xenérica, o esquema que se documenta límitase a só dúas, por un lado os exemplos (16) e (17) e polo outro as concordancias de (18), (19) e (20), dado que se mencionan tres árbores cuxos froitos (landras, piñas e castañas) é imposible que se convertan en uvas, cando menos, non sen mutacións xenéticas:

- (16) De momento todo era espera continente remachada polas reconvenções dos

<sup>25</sup> A data tope que temos en consideración é a do ano 2014, pois os documentos destinados a representa-lo lustro 2015-2019 están aínda en distintas fases do seu procesamento.



- adultos: ti fala *cando as galiñas mexen*; cando sexas pai, sentarás á mesa; ti fala *cando a ra críe pelo...*, e así. [CORGA: Xabier Paz, *As vidas de Nito*, Edicións Xerais de Galicia, 2013]
- (17) Pois simplemente quedaba reservada para a súa exclusividade a posibilidade de publicalo cando eles cresen oportuno. Ou sexa, *cando as ras críen pelo*. [CORGA: Pedro Feijoo Barreiro, *¡Viva o fu remol! Ópera en cinco actos sobre a música ó pé da rúa*, Edicións Xerais de Galicia, 2005]
- (18) Sancha: Teño de plantar un carballo en riba daquel tellado, *cando o carballo dea uvas* estarei ao teu mandado. [CORGA: Teresa Moure, *Unha primavera para Aldara*, Deputación Provincial da Coruña, 2008]
- (19) P: <inicio\_poema/>Teño de plantar un piñeiro, enriba do teu tellado, *cando os piñeiros dean uvas* estarei ó teu mandado<fin\_poema/>. [CORGA: Xosé Lois Ripalda, *O país dos mil ríos. Cultura fluvial en Galicia*, Galaxia, 2009]
- (20) R: <inicio\_poema/>teño de plantar un castiñeiro enriba da túa eira, *cando os castiñeiros dean uvas* andar hei ó teu rabo<fin\_poema/>. [CORGA: Xosé Lois Ripalda, *O país dos mil ríos. Cultura fluvial en Galicia*, Galaxia, 2009]

Non responde á estrutura que se especifica na procura a primeira variante destacada en itálica no exemplo (16), *cando as galiñas mexen*, inda que si corresponde á expresión de imposibilidade de levar adiante un determinado suceso. Como sabemos, as galiñas posúen un único orificio de saída, a cloaca, a través da cal expulsan o líquido de rexeitamento que produce o ril ó filtra-lo sangue mesturado cos excrementos procedentes do tubo dixestivo, o que devén nunha mestura pastosa, polo que *stricto sensu* as galiñas non miccionan.

#### 4. Conclusións

A modo de conclusión, queremos salienta, por unha banda, a potencialidade do sistema de consultas do CORGA, que permite empregar nunha mesma busca comodíns, operadores booleanos, sensibilidade a acentos ou maiúsculas e variables clasificatorias dos documentos combinándoos cos distintos tipos de modalidade de procura (por palabras ortográficas ou elementos gramaticais), ben sucesivos ben descontinuos, o que converte o CORGA nunha ferramenta moi útil para obter datos da lingua galega actual de tipo léxico, gramatical, terminolóxico, fraseolóxico, discursivo etc. Mais tamén, por outra banda, como demostramos nas actividades propostas e desenvoltas, o CORGA en particular, e os corpus en xeral, convértese nun valioso recurso para a obtención de fraseoloxía galega, serve ademais como fonte para a elaboración de materiais educativos con lingua real actual e promove a adquisición de competencias nas TIC mediante o seu emprego.

#### 5. Referencias bibliográficas

- BAILINI, Sonia (2015): “Los corpus como recursos didácticos para la enseñanza de las variedades diatópicas del español”, en *FIAPE. V Congreso internacional: ¿Qué español enseñar y cómo? Variadades del español y su enseñanza*. Biblioteca virtual redELE 16, 1-17.

- BILEGA = GARCÍA GONDAR, Francisco (dir.): *Bibliografía informatizada da lingua galega* [3.0]. Santiago de Compostela: Centro Ramón Piñeiro para a investigación en humanidades (<http://cirp.gal/bilega>) [Data de consulta: 6/9/2019]
- CAES = ROJO, Guillermo e M. PALACIOS, Ignacio (2016): “Learner Spanish on Computer. The CAES (Corpus de aprendices de español) Project”, en ALONSO-RAMOS, Margarita (ed.): *Spanish Learner Corpus Research. Current Trends and Future Perspectives*. Amsterdam: John Benjamins, 55-87.
- CORTEGAL = ÁLVAREZ DE LA GRANJA, María (2018): “Corpus de textos de estudantes galegos (CORTEGAL). Aspectos metodolóxicos” en DÍAZ, Marta *et alii* (eds.): *Actas do XIII Congreso Internacional de Lingüística Xeral*. Vigo: Universidade de Vigo, 55-62.
- DOMÍNGUEZ NOYA, Eva María (2013): *Etiquetaxe e desambiguación automáticas en galego: o sistema XIADA*. Tese de doutoramento (<http://hdl.handle.net/10347/9587>) [Data de consulta: 6/9/2019]
- DOMÍNGUEZ NOYA, Eva María e LÓPEZ MARTÍNEZ, María Sol (2017): “Tratamento da variación lingüística no CORGA”, en NEGRO ROMERO, Marta *et alii* (eds.): *Gallaecia. Estudos de lingüística portuguesa e galega*. Santiago: Universidade de Santiago de Compostela, 421-440.
- DOMÍNGUEZ NOYA, Eva María e BARCALA RODRÍGUEZ, Francisco M. (2018): “Grafías innovadoras na linguaxe non sexista: unha proposta para a súa etiquetaxe automática”, en DÍAZ, Marta *et alii* (eds.): *Actas do XIII Congreso Internacional de Lingüística Xeral*. Vigo: Universidade de Vigo, 291-298.
- DRAG = Real Academia galega: *Dicionario da Real Academia Galega*. (<https://academia.gal/dicionario>) [Data de consulta: 2/9/2019]
- EAGLES (1996): *Recommendations for the Morphosyntactic Annotation of Corpora. EAGLES Document EAG-TCWG-MAC/R* (<http://www.ilc.cnr.it/EAGLES96/browse.html>) [Data de consulta: 24/04/2019]
- FERRO RUIBAL, Xesús (2016): “Prólogo” a *Para dar trela. Manual práctico de fraseoloxía galega*. Santiago de Compostela: Centro Ramón Piñeiro para a investigación en humanidades, 9-14.
- LOSADA ALDREY, M. Carmen; CID FERNÁNDEZ, Alba e FERNÁNDEZ CARBALLIDO, Xurxo (2016): *Para dar trela. Manual práctico de fraseoloxía galega*. Santiago de Compostela: Centro Ramón Piñeiro para a investigación en humanidades.
- ROJO, Guillermo (2016): “Los corpus textuales del español”, en GUTIÉRREZ-REXACH, Javier (ed.): *Enciclopedia lingüística hispánica*. Oxon: Routledge, 285-296.
- SEVILLA MUÑOZ, Julia e BARBADILLO DE LA FUENTE, M.<sup>a</sup> Teresa (2004): “Valor didáctico del refrán” en *Pemria* 13, 195-204.
- TILG = SANTAMARINA, Antón (dir.); GONZÁLEZ SEOANE, Ernesto e ÁLVAREZ DE LA GRANJA, María: *Tesouro informatizado da lingua galega* [4.1]. (<http://ilg.usc.es/TILG/>) [Data de consulta: 4/9/2019]
- TOLCHINSKY, Liliana (2014): “El uso de corpus lingüísticos como herramienta pedagógica” en *Textos de didáctica de la Lengua y de la Literatura* 65, 9-17.

## Apéndice

Unidade fraseolóxica	Descrición	Modalidade de consulta / Tipo de resultado	Aspectos computacionais	Dificultade	Aspectos lingüísticos
<i>ser uña e carne</i>	Estar moi unidos por vínculos de amizade (DRAG)	Palabra ortográfica / Concordancias	-Emprego de comodíns para substituír caracteres -Delimitación do segmento de busca -Concordancias	Baixa	-Variación léxica: <i>uña / unlla</i> -Fraxe verbal -Atributo
<i>por mor de</i>	Indica a causa, o motivo de que algo suceda (DRAG)	-Elementos gramaticais / Frecuencia simple / Frecuencia completa -Interconexión frecuencias completas e concordancias	-Sensibilidade a maiúsculas -Frecuencia normalizada -Importancia da distribución temporal para ve-la vitalidade -Importancia da tipoloxía textual en relación co uso	Baixa	-Preposicións e locucións prepositivas -Contraccións con <i>de</i> -Vitalidade dunha unidade -Uso escrito/uso oral
<i>estar ata as orellas</i> (e similares)	Estar moi canso de algo ou de alguén	-Elementos gramaticais / Concordancias	-Lema -Hiperlema -Etiqueta morfolóxica -Criterios de ordenación -Contigüidade	Alta	-Padrón estrutural -Campo semántico -Sinonimia -Eufemismo, vulgarismo -Variedades situacionais -Norma e uso
<i>Aí a vén</i> (e similares)	Clífico acusativo en construcións con verbo intransitivo	-Elementos gramaticais / Frecuencias completas, gráficas -Interconexión gráficas e concordancias	-Emprego de operadores booleanos ( ) -Gráficas -Contexto	Alta	-Focalización -Colocación do pronomes -Problema de análise sintáctica: pronomes de acusativo con verbos intransitivos
<i>poñer / por verde</i>	-Poñer algo de cor verde [unha prenda / un obxecto] -Poñerse verde ~ poñerse en verde [o semáforo] -Poñer verde a alguén (criticar)	Elementos gramaticais próximos	-Linearidade e proximidade -Distancia	Media	-Uso literal / uso figurado -Establecemento de estruturas sintácticas a partir dos resultados
<i>Onde comen dous comen tres</i> (e similares)	Expresión relativa da cantidade	Elementos gramaticais	-Buscas abstractas só con trazos gramaticais	Alta	-Verbos con alternancia vocálica -Paralelismo -Variación nas unidades fraseolóxicas
<i>subir á cabeza levar</i> (a alguén) <i>a cabeza quitar da cabeza, senta-la cabeza meterse en camisas</i> <i>de once varas tal x tal z</i>	<i>Embarrachar levemente Marearse Disuadir Coler xuízo Meterse en leas</i>	Elementos gramaticais próximos	-Recuperación selectiva dun caso de cada modismo -Ruído	Media	-Alomorfos do artigo -Suxeito, CD, Complemento preposicional e CCL
<i>cando o carballo / castiñeiro / piñeiro dea unvas cando ais ras / críe/n pelo</i>	Xustificación das virtudes e defectos dun descendente a respecto dun ascendente Probabilidade nula de que ocorra algo	Elementos gramaticais próximos	Criterios de ordenación	Media	-Polisemia -Complemento preposicional -Estruturas bimembres
		Elementos gramaticais próximos	Afianzamento das destrezas adquiridas	Alta	-Estruturas bimembres
		Elementos gramaticais próximos	Afianzamento das destrezas adquiridas	Alta	-Léxico: as árbores e os seus froitos

