



INSTITUTO DA LINGUA GALEGA (ILG)

**SISTEMA DE BUSCAS E VISUALIZACIÓN DOS TEXTOS**  
**EN *CORTEGAL, CORPUS DE TEXTOS GALEGOS***  
***ESCRITOS POR ESTUDANTES NO ÁMBITO***  
***ACADÉMICO***

**Versión 1 (2022)**



CONSELLERÍA DE CULTURA,  
EDUCACIÓN, FORMACIÓN  
PROFESIONAL E UNIVERSIDADES

PGC2018-096069-B-I00

## Táboa de contidos

1. Acceso .....	3
<b>2. Xerador de consultas</b> .....	4
<b>2.1 Busca de texto</b> .....	4
2.1.1 Dous tipos xerais de busca.....	4
2.1.2 Tipos concretos de busca .....	5
2.1.2.1 Versión final estudante .....	5
2.1.2.2 Estándar ortográfico / morfolóxico / léxico / gramatical / semántico / discursivo.....	5
2.1.2.3 Tipo de desviación do estándar .....	7
2.1.2.4 Fonte da forma non estándar .....	7
2.1.2.5 Clase de palabra estándar e clase de palabra orixinal .....	8
2.1.2.6 Lema estándar e lema orixinal .....	9
2.1.2.7 Conector .....	9
2.1.2.8 Anotacións multipalabra .....	10
<b>2.2 Busca de documentos</b> .....	10
<b>3 Buscas CQL (“Corpus Query Builder”)</b> .....	13
<b>3.1 Presentación xeral</b> .....	13
<b>3.2 Comodíns e operadores lóxicos</b> .....	15
<b>3.3 Busca de palabras e secuencias</b> .....	16
<b>3.4 Busca de clase de palabra estándar e orixinal</b> .....	18
<b>3.5 Cancelación da diferenciación entre maiúsculas e minúsculas</b> .....	18
<b>3.6 Combinación de varios “Tipos de desviacións do estándar” ou “Fontes da forma non estándar”</b> .....	18
<b>4. Algúns sintagmas fusionados</b> .....	19
<b>5. Os resultados das buscas</b> .....	20
<b>5.1 Buscas de texto</b> .....	20
5.1.1 Ordenación dos resultados.....	20
5.1.2 Visualización dos resultados.....	20
5.1.3 Visualización dos textos.....	21
5.1.4 Descarga dos resultados das buscas e dos textos.....	26
<b>6. Consultas sobre frecuencia</b> .....	27
Anexo I. Códigos que identifican o tipo de desviación do estándar.....	29

## 1. Acceso

A través da lapela “Buscar” que figura na páxina web do proxecto (Figura 1), pódese acceder ao buscador de CORTEGAL.

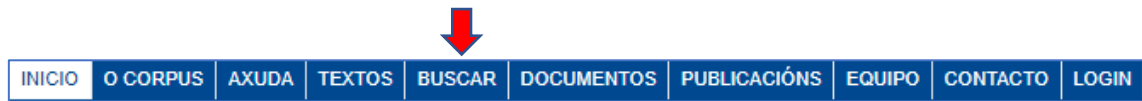


Figura 1. Lapela “Buscar”

A aplicación permite realizar diferentes operacións de busca nos textos incluídos no corpus, ben sexa a través do “xerador de consultas”, ben sexa directamente escribindo na caixa de buscas CQL (Figura 2).

The interface is titled 'Xerador de consultas'. It is divided into two main columns: 'Busca de texto' on the left and 'Busca de documentos' on the right. The 'Busca de texto' column contains a list of criteria with dropdown menus and input fields: Versión final estudante, Estándar ortográfico, Estándar morfolóxico, Estándar léxico, Estándar gramatical, Estándar semántico, Estándar discursivo, Tipo de desviación do estándar, Fonte da forma non estándar, Clase de palabra (estándar), Clase de palabra (orixinal), Lema estándar, Lema orixinal, and Conector. The 'Busca de documentos' column contains criteria with dropdown menus and input fields: Título, Tema, Convocatoria de exame, Comisión Delegada, Número de palabras, Número de lemas, Densidade léxica, Número de enunciados, Palabras por enunciado, Palabras enunciado máis longo, Palabras enunciado máis curto, Número de parágrafos, and Enunciados por parágrafo. At the bottom right, there is a section for 'Anotacións multipalabra' with a 'Código de desviación' dropdown menu. Red arrows point to the 'Xerador de consultas' title and the 'Busca de texto' section.

Figura 2. Xerador de consultas

Se, tras realizar unha busca (de acordo coas indicacións que se darán a seguir) e obter a concordancia cos resultados, se desexa realizar outra busca a través do xerador, débese premer en “xerador de consultas” (sinalado na Figura 3).

Buscas no corpus

Consulta CQL:   [xerador de consultas](#)  
[visualizar](#) | [opcións](#)

112 resultados • ipm: 407.84 • Amosando 0 - 100 • seguinte

Texto:

Anotación:

contexto	un respiro fóra da <b>casa</b> , por   maior comodidade.
contexto	que facelo na <b>casa</b> , por que? Nadie
contexto	e cuxa preparación   na <b>casa</b> é plausible. Teñen verdadeira
contexto	cando chegase a <b>casa</b> de traballar; é unha
contexto	problemas de entendemento   na <b>casa</b> nestas idades son practicamente
contexto	o sofá da   súa <b>casa</b> . Non obstante, tamén
contexto	que se comía na <b>casa</b> ou a comida asociábase
contexto	. Porque cociñar na <b>casa</b> e ensinar isto dunha

Figura 3. Acceso ao xerador de consultas tras unha busca

## 2. Xerador de consultas

O xerador de consultas permite buscar texto e documentos. Comezaremos polas buscas de texto en 2.1 e en 2.2 comentaremos as posibilidades que ofrece a busca por documentos.

### 2.1 Busca de texto

#### 2.1.1 Dous tipos xerais de busca

Nas buscas por “Tipo de desviación do estándar”, “Fonte da forma non estándar”, “Clase de palabra estándar”, “Clase de palabra orixinal” e “Conector” ofrécese un selector con diferentes etiquetas (tal e como se ve na Figura 4, para “Tipo de desviación do estándar”).

[Seleccionar]

[Seleccionar]

D\_con\_ad  
D\_con\_om  
D\_con\_su  
D\_np\_ad  
D\_np\_om  
D\_pm\_ad  
D\_pm\_om  
D\_pm\_su  
D\_pm\_wp  
D\_ref\_ad-su  
D\_ref\_om  
D\_reg\_ad  
D\_reg\_om  
D\_reg\_su  
D\_w\_om

Figura 4. Selector de etiquetas

Cómpre seleccionar a etiqueta que se desexa buscar e premer en “Buscar” na parte inferior esquerda do xerador de consultas (vid. Figura 2).

No resto dos casos débese escribir na caixa de buscas correspondente á consulta que nos interesa a palabra que se desexa buscar. Máis en concreto, é posible localizar, seleccionando a opción correspondente no menú despregable (Figura 5), (a) a palabra completa escrita na caixa de buscas, (b) unha palabra que comeza por unha determinada secuencia de caracteres, (c) unha palabra que remata nunha determinada secuencia de caracteres, ou (d) unha palabra que contén unha secuencia de caracteres.

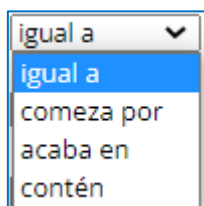


Figura 5. Opcións para as buscas de texto no xerador de consultas

Unha vez máis, unha vez seleccionada a busca que se quere facer, cómpre premer en “Buscar” (abaixo, á esquerda, na Figura 2).

### 2.1.2 Tipos concretos de busca

#### 2.1.2.1 Versión final estudante

Permite localizar unha palabra tal e como figura escrita pola/polo estudante. Se se quere buscar unha secuencia de palabras, pode escribirse unha palabra, a seguir premer en “engada palabra” (que aparece na parte final do xerador de consultas, por riba do botón de “Buscar”, vid. Figura 2), escribir en “palabra” o segundo elemento da secuencia, repetir estes pasos tantas veces como palabras conformen a secuencia que se quere buscar e finalmente premer na parte inferior en “Buscar”. En calquera caso, a busca dunha secuencia de palabras é máis doada a través da caixa de buscas CQL (vid. 3).

Debe terse en conta, por outro lado, que no caso das *contraccións*, dos *verbos con pronomes enclíticos*, das combinacións dos demostrativos co indefinido *outro* e do emprego da *segunda forma do artigo* o sistema divide a palabra nos elementos que a constitúen. Deste xeito, o buscador “Versión final estudante” non vai encontrar, por exemplo, a contracción *das*, nin a forma *trátase*, nin a palabra *estoutro*, nin o complexo *tódolos*. Para atopar estas formas débese facer unha busca CQL (vid. 3) ou ben escribir o primeiro compoñente (por exemplo *trata*) e “engadir palabra” para o seguinte compoñente (por exemplo *se*), tal e como indicamos no parágrafo precedente para as secuencias de palabras. Coherentemente co que se acaba de indicar, unha busca a través da preposición *de* no buscador “Versión final estudante” ofrecerá os exemplos de emprego de *de* tanto en contracción como fóra dela, unha procura pola palabra *trata* mostrará os exemplos de *trata* con pronome enclítico e sen el etc.

#### 2.1.2.2 Estándar ortográfico / morfolóxico / léxico / gramatical / semántico / discursivo

Algunhas palabras do texto son estandarizadas nun ou varios destes seis niveis lingüísticos e as caixas de busca correspondentes permiten localizar as respectivas formas estandarizadas.

Débese ter en conta, en calquera caso, que unha busca por esta vía encontra tanto as formas estandarizadas polo equipo do proxecto como aquelas que xa eran estándar no discurso da/do estudante. Así, por exemplo, quen busque na caixa de buscas “Estándar morfolóxico” unha secuencia que comece por *estea* encontrará tanto os exemplos de *esté*, *estemos*, *esteñan*,

*estén, esten...* que foron estandarizados no proceso de anotación dos textos con *estea, esteamos...*, como os casos en que o/a estudante escribiu directamente a forma estándar.

Se se desexa cinguir os resultados ás formas non estándares é preciso seleccionar na caixa de buscas “Tipo de desviación do estándar” o código identificador da desviación que se agocha tras esas formas (neste caso, M\_v\_su), tal e como se pode ver na Figura 6. Deste xeito, o buscador localiza todas as palabras anotadas con esa etiqueta (que identifica as formas verbais cunha conjugación non estándar) e que ademais foron estandarizadas por unha palabra que comeza por *estea-*.

Busca de texto		
Versión final estudante	igual a ▼	
Estándar ortográfico	igual a ▼	
Estándar morfolóxico	comeza por ▼	estea
Estándar léxico	igual a ▼	
Estándar gramatical	igual a ▼	
Estándar semántico	igual a ▼	
Estándar discursivo	igual a ▼	
Tipo de desviación do estándar	M_v_su ▼	

Figura 6. Busca de formas non estándares (1)

Noutras ocasións tamén poderá ser de interese combinar a busca por algún dos estándares coa busca por “Palabra”, como no caso da Figura 7, en que se pretenden identificar os casos de clítico *te* empregado en vez de *che*:

Busca de texto		
Versión final estudante	igual a ▼	te
Estándar ortográfico	igual a ▼	
Estándar morfolóxico	igual a ▼	
Estándar léxico	igual a ▼	
Estándar gramatical	igual a ▼	che

Figura 7. Busca de formas non estándares (2)

Finalmente, cómpre indicar que nalgúns casos a estandarización supón a supresión da forma escrita polo estudante. Estas formas suprimidas, ademais de seleccionando en “Tipo de desviación do estándar” as etiquetas de adición (\_ad), poden ser recuperadas escribindo dous trazos (--) nas caixas de busca das formas estándares.

### 2.1.2.3 Tipo de desviación do estándar

A caixa de buscas “Tipo de desviación do estándar” permite identificar todas aquelas formas que foron etiquetadas cun código identificador de formas non estándares nalgún dos seis niveis lingüísticos arriba indicados. Os códigos que figuran no selector deste tipo de busca e o seu valor figuran como anexo a este texto e tamén poden atoparse no documento [Códigos e etiquetas EAGLES empregados na anotación dos textos en CORTEGAL, Corpus de textos galegos escritos por estudantes no ámbito académico.](#)

Pódese optar entre escoller no selector exclusivamente o tipo de diverxencia que nos interesa, para obter unha concordancia con todas as formas que están etiquetadas con ese tipo de desviación do estándar, ou combinar esta busca con outros criterios de filtraxe. Así, por exemplo, se nos interesa obter unha concordancia cos casos en que se produciu a omisión da preposición *a*, escollemos o código G\_prep\_om en “Tipo de desviación do estándar” e *a* na caixa de buscas “Estándar gramatical”, mentres que se queremos localizar os casos de adición desta preposición, combinaremos a busca da etiqueta G\_prep\_ad en “Tipo de desviación do estándar” e con *a* na caixa de buscas “Versión final estudante”.

### 2.1.2.4 Fonte da forma non estándar

Algunhas das formas etiquetadas cun código de forma non estándar van acompañadas doutra etiqueta que identifica a orixe ou causa da desviación. Esta caixa de buscas permite localizar os códigos que figuran na Táboa 1, que se ofrecen no correspondente selector:

Orixe	Valor	Códigos	Exemplos
<i>anal</i>	Formas que responden a un proceso de analogía con outras palabras da mesma lingua	O_anal M_anal L_anal	O_anal: <i>hippie</i> (sobre o plural <i>hippies</i> )... M_anal: <i>puidendo</i> (sobre o tema do pretérito <i>puid-</i> )... L_anal: <i>nutrinte</i> (a partir de voces como <i>ferinte</i> ou <i>seguinte</i> )...
<i>for</i>	Palabras cuxa desviación se pode explicar pola súa orixe estranxeira e calcos de voces estranxeiras. Exclúense sempre as formas con orixe no español. No ámbito ortográfico, palabras cunha ortografía incorrecta. No ámbito morfolóxico, palabras cunha flexión non estándar. No ámbito léxico, palabras que non presentan a adaptación proposta no estándar. No ámbito semántico, calcos semánticos.	O_for M_for L_for S_for	O_for: <i>graffitti</i> (por <i>graffiti</i> )... M_for: <i>chef</i> (por <i>chef</i> )... L_for: <i>tablet</i> (por <i>tableta</i> )... S_for: <i>pretender</i> (co significado de ‘finxir’)
<i>gal</i>	Palabras, trazos ou usos transferidos desde unha variedade do galego non estándar.	M_gal L_gal	M_gal: <i>conseguen</i> (por <i>conseguín</i> )...

	<p>No ámbito morfolóxico, flexión propia do galego popular.</p> <p>No ámbito léxico, palabra tirada dunha variedade do galego (popular ou culto).</p> <p>No ámbito gramatical, uso propio dalgunha variedade dialectal.</p> <p>No ámbito semántico, uso tirado dunha variedade do galego (galego popular ou culto)</p>	<p>G_gal</p> <p>S_gal</p>	<p>L_gal: <i>atrais</i> (por <i>atrás</i>)...</p> <p>G_gal: uso de <i>lle</i> por <i>lles</i>, cheísmo e teísmo</p> <p>S_gal: <i>mirar</i> (co valor de <i>ver</i>)...</p>
<i>hc</i>	Hipercorreccións xeradas dentro do galego	<p>L_hc</p> <p>G_hc</p>	<p>L_hc: hiperenxebrismos como <i>vacina</i> (por <i>vacina</i>), <i>calqueira</i>...</p> <p>G_hc: colocación enclítica do pronome en vez de proclítica e uso de <i>che</i> por <i>te</i></p>
<i>mix</i>	Mesturas entre dúas palabras ou expresións	L_mix	L_mix: <i>indixesta</i> ( <i>indigestión</i> + <i>inxesta</i> ), <i>a saber Deus</i> ( <i>a saber</i> + <i>sabe Deus</i> ), <i>adedicar</i> ( <i>adicar</i> + <i>dedicar</i> )...
<i>or</i>	Unicamente se emprega no ámbito gramatical para marcar os casos de pronomes enclíticos posibles na lingua oral, pero non recomendados na escrita	G_or	G_or: colocación enclítica do pronome posible na lingua oral, pero non recomendado na escrita
<i>sp</i>	<p>Palabra, expresión, trazo ou uso transferidos desde o español (sen adaptación).</p> <p>No ámbito ortográfico, grafía coincidente co español ou acentuación gráfica explicable por influencia desta lingua.</p> <p>No ámbito morfolóxico, flexión coincidente co español.</p> <p>No ámbito gramatical, colocación proclítica do pronome, así como ausencia ou adición de determinante, adición de preposición ou de pronome e substitución de determinante ou de pronome explicables por influencia do español.</p> <p>No ámbito léxico, unidades léxicas, acentuación de intensidade ou xénero transferidos desde o español.</p> <p>No ámbito semántico, calcos semánticos do español</p>	<p>O_sp</p> <p>M_sp</p> <p>G_sp</p> <p>L_sp</p> <p>S_sp</p>	<p>O_sp: <i>abí</i> (por <i>aí</i>), <i>él</i>...</p> <p>M_sp: <i>anglosaxones</i>...</p> <p>G_sp: <i>levarse o diñeiro</i> (por <i>levar o diñeiro</i>)...</p> <p>L_sp: <i>actitud</i>, <i>élite</i>, (<i>a</i>) <i>leite</i></p> <p>S_sp: <i>pobo</i> (por <i>vila</i> ou <i>aldea</i>)...</p>
<i>spadapt</i>	<p>No ámbito morfolóxico, flexión adaptada desde o español.</p> <p>No ámbito léxico, palabras ou expresións transferidas desde o español e adaptadas ao galego, ou mesturas entre elementos do galego e do español</p>	<p>M_spadapt</p> <p>L_spadapt</p>	<p>M_spadapt: <i>atraxo</i>, <i>supoñe</i>...</p> <p>L_spadapt: <i>semexante</i>, <i>ventaxes</i>, <i>chear</i>...</p>

Táboa 1. Fonte da forma non estándar

Unha vez máis, pode ser de interese combinar diferentes filtros de busca. Por exemplo, se seleccionamos en “Tipo de desviación do estándar” o código L\_w\_su e en “Fonte da forma non estándar” o código L\_sp encontraremos todas as transferencias léxicas do castelán (que



van máis alá, de cambios na acentuación ou no xénero) que non sufriron adaptación (pois nese caso a etiqueta sería L\_spadapt).

#### 2.1.2.5 Clase de palabra estándar e clase de palabra orixinal

Poden facerse buscas polas clases morfosintácticas das palabras do corpus. O buscador ofrece un selector coas principais clases de palabras (substantivo, adxectivo, verbo...), pero se se desexa facer unha busca máis específica (por exemplo, substantivo masculino singular ou 1ª persoa de singular do presente de indicativo) cómpre realizar unha consulta a través da caixa de buscas CQL (vid. 3).

Evidentemente, podemos combinar a busca por clase de palabras con calquera outra co obxectivo de filtrar os resultados. Por exemplo, se queremos encontrar os usos de *a* como preposición, escribiremos *a* en “Versión final estudante” e seleccionaremos Preposición como clase de palabra.

No que respecta á diferenza entre clase de palabra estándar e clase de palabra orixinal, debe terse en conta que algunhas formas son corrixidas no nivel léxico, de xeito que pode haber diferenzas na clase de palabra que escribiu a/o estudante (clase de palabra orixinal) e a clase de palabra da forma normalizada (clase de palabra estándar). Así, por exemplo, a palabra *frente*, masculina, é corrixida por *fronte*, feminina. Só nestes casos haberá diferenzas entre a clase de palabra orixinal (NCMS000) e a estándar (NCFS000). Nos restantes casos, os resultados da busca por unha ou outra vía serán idénticos.

#### 2.1.2.6 Lema estándar e lema orixinal

Estas caixas permiten localizar nos textos as formas correspondentes a un determinado lema (por exemplo, todas as ocorrencias no corpus do verbo *valer* en calquera forma do seu paradigma). Por suposto, é posible tamén realizar procuras combinando esta busca con algún dos elementos anteriores: por exemplo, buscar todas as ocorrencias da “Versión final estudante” *foi* pertencentes ao lema *ir* (e non a *ser*).

Con respecto ás diferenzas entre as buscas por lema estándar e lema orixinal, estas veñen dadas polas estandarizacións que se realizan no nivel léxico. Así, o lema orixinal das palabras *plato* e *platos* será *plato* e o lema estándar *prato*. Do mesmo xeito, o lema orixinal de *artículo*, *artículo* e *artículos* será *artículo* e o lema estándar *artigo*. Así pois, o lema orixinal é a forma canónica representante do paradigma da forma escrita pola/polo estudante, mentres que o lema estándar representa o paradigma da correspondente forma galega estándar. Se non hai estandarizacións no nivel léxico, o lema orixinal coincide co lema estándar.

#### 2.1.2.7 Conector

É posible localizar mediante a caixa de buscas “Conector” os conectores empregados para vincular enunciados no texto. Estes están identificados coas etiquetas que figuran na Táboa 2, onde tamén se indica o valor de cada unha delas.

Código	Valor	Exemplos
ad	Introducen un elemento que se suma ao dito anteriormente na argumentación, mesmo reforzádo.	<i>ademaís, e, é máis, tamén...</i>
caus	Introducen a causa do expresado no enunciado ou enunciados previos	<i>ao fin e ao cabo, pois, porque...</i>

coment	Introducen un comentario sobre o expresado anteriormente	<i>ben, pois, pois ben...</i>
concl	Introducen o resultado ao que se chega a partir do exposto previamente ou ben o seu resumo	<i>en conclusión, en definitiva, en resumo...</i>
confirm	Introducen un argumento, explicación... que serve para confirmar o que se acaba de dicir	<i>efectivamente, en efecto, si</i>
cons	Introducen a consecuencia do expresado no enunciado ou enunciados previos	<i>así, entón, por conseguinte...</i>
contraarg	Introducen un argumento ou explicación que se contrapón dalgún xeito ao que se acaba de indicar	<i>aínda así, agora ben, en cambio...</i>
disjunc	Introducen unha alternativa ao que se acaba de expoñer	<i>ou</i>
exempl	Introducen un exemplo ou concreción do que se acaba de indicar	<i>así, concretamente, por exemplo...</i>
fin	Introducen a finalidade do expresado no enunciado ou enunciados previos	<i>para</i>
ord_sp	Serven para organizar o discurso distribuindo os feitos expostos en diferentes apartados mediante elementos con valor espacial	<i>por outra banda, por unha parte, por un lado...</i>
ord_t	Serven para organizar o discurso indicando a posición que ocupa cada elemento no proceso argumentativo ou expositivo do texto mediante elementos con valor temporal	<i>finalmente, por último, primeiramente...</i>
reform	Serven para explicar ou presentar doutra forma o que se acaba de expoñer	<i>é dicir, explícome, isto é</i>

Táboa 2. Conectores

#### 2.1.2.8 Anotacións multipalabra

Na parte inferior dereita do xerador de consultas pódense facer buscas das Anotacións multipalabra (vid. Figura 2). Algúns tipos de desviacións que poden afectar a secuencias de palabras, son anotados de maneira diferente aos restantes, mediante anotacións stand-off e só son recuperables mediante o selector “Código de desviación” que figura neste apartado do xerador de consultas. Os códigos e o seu significado recóllense na Táboa 3:

Código	Valor
D_ut_com	Enunciado excesivamente longo ou complexo
D_ut_unin	Enunciado inintelixible
D_w_wp	Colocación de palabras discursivamente inadecuada
G_str-ref-top_su	Estrutura gramaticalmente inadecuada por reformulación ou topicalización inadecuada dalgún elemento
G_str_su	Estrutura gramaticalmente inadecuada (agás os casos anotados coa etiqueta anterior e coa seguinte)
G_w_ad	Adición dalgún elemento que fai a estrutura gramaticalmente inadecuada
S_w_ad	Adición dalgún elemento que fai a estrutura semanticamente inadecuada

Táboa 3. Anotacións realizadas en stand-off

## 2.2 Busca de documentos

Ademais das buscas sobre o texto poden facerse buscas de documentos, así como combinar as buscas sobre o texto coa filtraxe por documentos. Estes poden seleccionarse de acordo cos seguintes criterios:

1. *Título* (por exemplo ABAU/2016-2017/CD13/xuño/04). Permite buscar un texto concreto. A primeira parte do título (ABAU/2016-2017) é común para todos os textos. O que varía é o número da comisión delegada (CD13, CD09...) (vid. o punto 4), a convocatoria (xuño ou setembro) e as cifras finais, que diferencian os textos pertencentes á mesma comisión delegada e convocatoria.
2. *Tema*. Ofrece un selector cos catros temas dos textos (“A gastronomía”, “Consumo e produción”, ambos da convocatoria de xuño, “Conflitos familiares” e “Os referentes da mocidade”, ambos os dous da convocatoria de setembro. Pode seleccionarse un deles para facer as buscas.
3. *Convocatoria do exame*. Se se desexa, pódense seleccionar só os exames da convocatoria de xuño ou só os de setembro.
4. *Comisión delegada*. Cada exame da ABAU está vinculado a unha comisión delegada (CD) asignada ao centro da/do estudante e cada CD recolle diversos centros (públicos e privados) dunha determinada área xeográfica. A persoa usuaria pode seleccionar unha comisión delegada (CD) en particular. Os centros asociados a cada CD poden consultarse no documento [Centros de ensino asociados ás comisións delegadas das probas ABAU \(curso 2016-2017\)](#). Na convocatoria de setembro, con moitos menos alumnos ca a de xuño, algunhas CD agrúpanse, o que explica que no buscador aparezan algunhas combinacións como, por exemplo, 01-03-05.
5. *Número de palabras*. Pódense seleccionar todos aqueles textos cuxo número de palabras se sitúe nun rango fornecido pola persoa usuaria. Dado que o cómputo, tal e como indicaremos en breve, se realizou a través da aplicación *DContado*, de acordo co sistema establecido por esta non se contabilizan os signos de puntuación e, por outro lado, as contraccións, os verbos con pronomes enclíticos e as formas acompañadas da segunda forma do artigo compútanse como unidades.
6. *Número de lemas*. Pódense seleccionar todos aqueles textos cuxo número de lemas ou palabras diferentes se sitúe no rango indicado pola persoa usuaria.
7. *Densidade léxica*. A persoa usuaria pode establecer un rango para filtrar os textos por densidade léxica. En principio, esta resulta de dividir o número de lemas entre o número de palabras e sempre se sitúa entre 0 e 1. Con todo, o buscador da aplicación non ten en conta os decimais nas buscas, de tal xeito que optamos por multiplicar a cifra por 100 para poder utilizar este criterio para a filtraxe. Así pois, un texto cunha densidade léxica de 0,54 terá asignado o valor 54. Daquela, o rango do buscador sitúase entre 0 e 100.
8. *Número de enunciados*. Pódense filtrar os textos polo número de enunciados que conteñen e de acordo co rango que se desexe establecer.
9. *Palabras por enunciado*. Outro criterio de selección dos documentos é a media de palabras por enunciado, calculado dividindo o número de palabras do texto entre o número de enunciados que contén. Débese ter en conta que, tal e como se indicou no punto 7, a aplicación só ten en conta a parte enteira dos valores numéricos atribuídos, de tal xeito que se buscamos un texto cunha media de palabras por

enunciado situada entre 13 e 15, encontrará tamén aqueles textos cuxa media é 15 seguida de decimais.

10. *Palabras enunciado máis longo*. Pódense buscar aqueles textos cuxo enunciado máis longo se sitúe no rango de número de palabras establecido pola persoa usuaria.
11. *Palabras enunciado máis curto*. Pódense buscar aqueles textos cuxo enunciado máis curto se sitúe no rango de número de palabras establecido pola persoa usuaria.
12. *Número de parágrafos*. A aplicación permite localizar aqueles textos cun número de parágrafos que se sitúe no rango que se indique.
13. *Enunciados por parágrafo*. A aplicación permite localizar aqueles textos cunha media de enunciados por parágrafo que se sitúe no rango solicitado. Aplícase aquí a mesma consideración sobre os decimais comentada no punto 9.

Debe terse en conta que, como indicamos, os cómputos de palabras, así como os de lemas, enunciados, palabras do enunciado máis longo e palabras do enunciado máis curto se realizan automaticamente mediante a aplicación DContado (<https://ilg.usc.gal/dcontado/index.php?lang=gl>), sen revisión manual posterior.

Tras realizar unha busca exclusivamente a través do Buscador de documentos (sen combinar coa busca de texto) obtense unha listaxe dos documentos atopados, con todas as súas características. Premendo na primeira columna (ID) accédese a cada texto.

ID	Título	Tema	Convocatoria de exame	Comisión Delegada	Número de palabras	Número de lemas	Densidade léxica	Número de enunciados	Palabras por enunciado	Palabras enunciado máis longo	Palabras enunciado máis curto	Número de parágrafos	Enunciados por parágrafo
CORT_0938.xml	ABAU/2016-2017/CD12/setembro/03	Conflictos familiares	Setembro	12	137	81	59	8	17.12	26	8	5	1.60
CORT_0795.xml	ABAU/2016-2017/CD22/xuño/32	Consumo e produción	Xuño	22	149	82	55	6	24.83	35	13	3	2
CORT_0203.xml	ABAU/2016-2017/CD06/xuño/14	A gastronomía	Xuño	06	140	78	56	4	35	55	16	4	1
CORT_0379.xml	ABAU/2016-2017/CD10/xuño/10	A gastronomía	Xuño	10	121	68	56	5	24.20	33	16	4	1.25
CORT_0440.xml	ABAU/2016-2017/CD12/xuño/15	A gastronomía	Xuño	12	132	82	62	5	26.40	58	9	3	1.67
CORT_0656.xml	ABAU/2016-2017/CD19/xuño/40	Consumo e produción	Xuño	19	115	70	61	9	12.78	18	9	3	3
CORT_0455.xml	ABAU/2016-2017/CD12/xuño/30	Consumo e produción	Xuño	12	148	91	61	7	21.14	29	10	4	1.75
CORT_0105.xml	ABAU/2016-2017/CD04/xuño/07	A gastronomía	Xuño	03	135	80	59	5	27	42	16	5	1
CORT_0110.xml	ABAU/2016-2017/CD04/xuño/12	A gastronomía	Xuño	04	101	63	62	3	33.67	42	29	2	1.50
CORT_0931.xml	ABAU/2016-2017/CD08-09/setembro/09	Conflictos familiares	Setembro	08-09	112	66	59	6	18.67	26	11	1	6

Figura 8. Resultado da busca por documentos

Por outra banda, pode combinarse a busca no texto coa busca por documentos para filtrar os resultados. Por exemplo, tal e como se pode ver na Figura 9, podemos traballar só co subcorpus de textos co tema “A gastronomía” seleccionando o “Tema” correspondente no selector e ao tempo buscar no xerador de consultas as desviacións do estándar, lemas... que nos interesen. Desta volta, o resultado será unha concordancia coas liñas de texto que cumpren os criterios da busca, pero só nos textos que tratan o tema da gastronomía.

Busca de texto		Busca de documentos	
Versión final estudiante	igual a <input type="text"/>	Título	<input type="text"/>
Estándar ortográfico	igual a <input type="text"/>	Tema	A gastronomía
Estándar morfológico	igual a <input type="text"/>	Convocatoria de exame	[Seleccionar]
Estándar léxico	igual a <input type="text"/>	Comisión Delegada	[Seleccionar]
Estándar gramatical	igual a <input type="text"/>	Número de palabras	<input type="text"/> - <input type="text"/>
Estándar semántico	igual a <input type="text"/>	Número de lemas	<input type="text"/> - <input type="text"/>
Estándar discursivo	igual a <input type="text"/>	Densidade léxica	<input type="text"/> - <input type="text"/>
Tipo de desviación do estándar	D_pm_su	Número de enunciados	<input type="text"/> - <input type="text"/>
		Palabras por enunciado	<input type="text"/> - <input type="text"/>
		Palabras enunciado	<input type="text"/> - <input type="text"/>

Figura 9. Busca de texto e filtraxe por documentos

### 3 Buscas CQL (“Corpus Query Builder”)

#### 3.1 Presentación xeral

As buscas CQL realízanse na caixa que se mostra a seguir (Figura 10), escribindo nela a busca que se desexa realizar e premendo no botón azul “Buscar” que figura á dereita.

### Buscas no corpus

Consulta CQL:

Buscar

xerador de consultas

visualizar | opcións

Figura 10. Busca dunha palabra no buscador CQL

Este tipo de buscas permite ampliar as posibilidades que ofrece o xerador de consultas e, nalgún caso, facilitalas. Debe terse en conta que calquera das buscas realizadas no xerador se converte automaticamente nunha busca en CQL. Por exemplo, se no xerador de consultas escribimos en “Versión final estudiante” *lle* e en “Tipo de desviación do estándar” seleccionamos a etiqueta *G\_pron\_ad*, a aplicación xera automaticamente unha consulta CQL, tal e como se pode ver na figura 11, onde, con fondo gris, tamén podemos encontrar un cadro explicativo do significado da consulta, cadro que se pode pechar premendo no <x> situado á dereita.

### Buscas no corpus

Consulta CQL:

Buscar

xerador de consultas

visualizar | opcións

Visualización da consulta CQL

1

Versión final estudiante = lle  
e Tipo de problema = (.\*)?G\_pron\_ad[.]\*?

x

Figura 11. Xeración automática da busca en CQL

Con todo, como indicamos, as buscas a través do xerador de consultas non abranguen todas as posibilidades que nos ofrecen as consultas CQL. Nas liñas que seguen mostraremos algunhas potencialidades da consulta que non foron comentadas previamente e que requiren coñecemento de certas expresións regulares. Antes ofreceremos as correspondencias entre os nomes das buscas do xerador de consultas e os nomes das buscas CQL, que son necesarios para as buscas:

Xerador de consultas	CQL
Versión final estudante	form
Estandarización ortográfica	ocform
Estandarización morfolóxica	mcform
Estandarización léxica	lcform
Estandarización gramatical	gcform
Estandarización semántica	scform
Estandarización discursiva	dcform
Tipo de desviación do estándar	problem
Fonte da forma non estándar	psource
Clase de palabra estándar	pos
Clase de palabra orixinal	opos
Lema estándar	lemma
Lema orixinal	olemma
Conector	con

Táboa 4. Equivalencia entre os elementos do xerador de consultas (busca de textos) e das buscas CQL

Pódese realizar calquera busca directamente a través da caixa de consultas escribindo entre corchetes o criterio xeral de busca seguido do signo igual e do elemento que se quere buscar entre comiñas. Por exemplo, para buscar o lema orixinal *casa* podemos escribir esta secuencia:

[olemma="casa"]

Para buscar as formas etiquetadas con determinado código de desviación do estándar, o máis doado é acudir ao xerador de consultas, pero no caso de empregar a consulta CQL debe terse en conta que, dado que “Tipo de desviación do estándar” é un criterio que admite a inclusión de varias etiquetas ao tempo (por exemplo, L\_w\_su e O\_ac\_om), se buscamos unha destas etiquetas (por exemplo, L\_w\_su), é preciso colocar as secuencias *(.\*)?* e *(,\*)?* rodeando este código, ou ben o comodín para cero ou máis caracteres *(.\*)*.

[problem = "(.\*)?L\_w\_su(.\*)?"]

[problem = ".\*L\_w\_su.\*"]

Deste xeito, incluíranse nos resultados tanto as formas que só foron etiquetada con L\_w\_su, como todas aquelas que recibiron este código e algún outro. En caso de non incluír os elementos indicados, na concordancia só aparecerían as formas que recibiron L\_w\_su como única etiqueta. O mesmo é aplicable ás etiquetas de “Fonte da forma non estándar”.

No caso da busca por palabras da versión final da/do estudante, podemos escribir directamente a forma ou secuencia na caixa de buscas (por exemplo, *casa* ou *a casa*) (aínda que, por suposto, podemos tamén escribir [form="casa"] ou [form="a"] [form="casa"]).

Para as buscas dos códigos anotados en stand-off (anotacións multipalabra) (vid. 2.1.2.8), a secuencia que se debe escribir é a que segue, en que exemplificamos cos problemas de estrutura gramatical:

```
<errors> []+ :: match.errors_code = "G_str_su" within errors
```

### 3.2 Comodíns e operadores lóxicos

En calquera busca a través da caixa de buscas CQL poden empregarse elementos comodín: `.*` para 0 ou máis caracteres e `[^q]` para un só carácter. Así, por exemplo, podemos encontrar calquera problema, sexa do nivel ortográfico, morfolóxico ou léxico que afecte á acentuación escribindo `[problem = ".*ac.*"]`.

No caso das buscas por palabras da versión final da/do estudante e por secuencias de palabras podemos escribir directamente o asterico (\*), sen un punto previo, como se indicará en 3.3.

Ademais, pode empregarse o operador lóxico `&` para combinar dous criterios de busca que se deben cumprir ao tempo. Así, por exemplo, coa secuencia

```
[pos = "CS" & lemma = "que"]
```

encontramos todos os exemplos do lema *que* etiquetado como conxunción subordinante.

Así mesmo, utilizaremos o operador lóxico `|` para realizar buscas conxuntas de varias formas. Así, por exemplo, no seguinte caso obtemos unha concordancia coas liñas de texto en que aparecen *meu*, *ten* ou *sen*.

```
[lemma= "meu" | lemma= "ten" | lemma= "sen"]
```

É posible tamén realizar unha filtraxe negativa mediante o elemento `!` colocado inmediatamente despois do criterio de busca e xunto antes do símbolo igual (sen espazos entre eles). Este elemento exclúe a presenza dos elementos situados á dereita do símbolo igual. Así, por exemplo, coa seguinte consulta buscamos todos os casos de *fai* que non foron corrixisos por *hai* no nivel semántico:

```
[form= "fai" & scform!= "hai"]
```

No caso de que estes operadores se apliquen a "Tipo de desviación do estándar" ou "Fonte da forma non estándar", a sintaxe da busca é diferente (vid. 3.6).

Nas buscas de documentos tamén se poden empregar os diferentes operadores. Así, por exemplo, podemos buscar ao tempo os textos das Comisións delegadas 04 e 10 empregando `|`:

```
<text> [] :: match.text_commission = "04" | match.text_commission = "10"
```

Por outra banda, escribindo `<text> []` na caixa da consulta CQL ou simplemente premendo en buscar, pero sen escribir nada nesa caixa de buscas, obtense a listaxe de todos os textos coas súas características.

### 3.3 Busca de palabras e secuencias

En CQL pódese realizar unha busca simple (do mesmo tipo que vimos no buscador “Versión final estudante” do xerador de consultas). Neste caso, simplemente cómpre escribir a palabra que se desexa buscar na caixa de buscas. Ademais, pódese empregar o asterisco (\*) (que o buscador CQL converte en .\*) como comodín equivalente a cero ou máis letras e buscar, por exemplo, *casa\**, busca que devolverá todas as formas que comecen por *casa*, incluída esta mesma forma (*casa*, *casar*, *casas*...), *\*ción* (busca que devolverá todas as formas acabadas nesta secuencia, *a\*mento* (que dará como resultado todas as palabras que comecen por *a* e rematen en *mento*) etc.

Tal e como indicamos no apartado anterior, como comodín equivalente a unha letra debemos empregar [^q]. Deste xeito, se, por exemplo, queremos buscar conxuntamente a terceira persoa de plural do presente de indicativo e de subxuntivo do verbo *ter* unha das opcións de busca sería esta: "teñ[^q]n" (neste caso, cómpre colocar comiñas para que a busca funcione).

Ademais, na procura directa CQL é posible buscar unha secuencia de palabras. Así, por exemplo, podemos buscar a secuencia *De feito* para encontrar todos os exemplos de emprego deste conector. Tal e como se indicou en 2.1.2.1, na consulta CQL poderemos atopar contraccións, verbos seguidos de pronome enclítico ou formas coa segunda forma do artigo, pero sempre descompoñendo a palabra nos seus compoñentes. Así, escribindo *de estes* na caixa de buscas, tal e como se ve na Figura 12,



The image shows a search interface with a text input field containing "de estes". To the left of the field is the label "Busca CQL:". To the right of the field is a blue button labeled "Buscar". Further to the right, there is a link "xerador de consultas |" and below it, the text "visualizar | opcións".

Figura 12. Busca da secuencia *de estes*

xérase unha busca en CQL que ofrece todos os exemplos de *de* seguido de *estes*, figuren ou non en contracción no texto orixinal (Figura 13).



## Buscas no corpus

Consulta:  **Buscar** xerador de consultas  
 CQL: | visualizar | opcións

Visualización da consulta CQL

1	2
Versión final estudante = de	Versión final estudante = estes

67 resultados • ipm: 244.07

Texto: **Transcrición completa** **Versión final estudante** **Estándar ortográfico** **Estándar morfolóxico** **Estándar léxico**  
**Estándar gramatical** **Estándar semántico** **Estándar discursivo**

Anotación: **Lema estándar** **Lema orixinal** **Clase de palabra (estándar)** **Clase de palabra (orixinal)** **Tipo de problema**  
**Orixe do problema** **Corrección derivada** **Conector**

Mostrar contexto	elementos,   e finalmente moitos	destes	produtos son tirados ao
Mostrar contexto	sexa máis fácil o divulgamento	destes	produtos. Todas estas compañías
Mostrar contexto	outra banda, a aparición	destes	programas fai que   moitas persoas
Mostrar contexto	dicir que a fabricación	destes	produtos   prestixiosos se leva a
Mostrar contexto	traballadores, aumentando a desigualdade	destes	cos empresarios.
Mostrar contexto	, así como o consello	destes	, e aínda sendo
Mostrar contexto	normais. A elaboración	destes	pratos é cada vez máis
Mostrar contexto	procedencia. O máis salientable	destes	datos é que estes últimos
Mostrar contexto	e provocan o encarecemento	destes	. Isto non sería un
Mostrar contexto	consumidores.   O problema	destes	consumidores e o motivo   polo
Mostrar contexto	conduce a que os referentes	destes	xóvenes sexan futbolistas e modelos
Mostrar contexto	esta nova moda. Moitos	destes	seguen a servir   comidas clásicas
Mostrar contexto	xa que   as etapas finais	destes	concursos acadan a categoría
Mostrar contexto	a   formación gastronómica	destes	centros excelente.   Eu o
Mostrar contexto	rápido e estas persoas precisan	de estes	obxectos; e a segunda

Figura 13. Primeiros resultados da busca da secuencia *de estes*

Tamén podemos buscar unha secuencia de elementos que cumpran unha serie de características. Así, por exemplo, se queremos encontrar todos os casos do verbo *ir* seguido da preposición *a* e dun infinitivo, escribiremos a seguinte secuencia na caixa de buscas (sobre as etiquetas de clase de palabras, vid. o apartado seguinte).

[lemma = "ir"] [lemma = "a"] [pos = "VMN0000"]

E nesta busca de secuencias, podemos tamén empregar o operador de filtraxe negativa. Así, por exemplo, a seguinte secuencia de busca permite encontrar todos os determinantes posesivos (DP.\*) que non están precedidos dun artigo (DA.\*):

[pos!= "DA.\*"] [pos= "DP.\*"]

### 3.4 Busca de clase de palabra estándar e orixinal

Como indicamos en 2.1.2.5, a través do xerador de consultas podemos seleccionar etiquetas gramaticais básicas, pero se desexamos facer unha busca máis refinada debemos empregar a consulta CQL, colocando na caixa de busca a etiqueta concreta que queremos buscar. A listaxe de etiquetas empregadas (etiquetas EAGLES), figura no documento [Códigos e etiquetas EAGLES empregados na anotación dos textos en CORTEGAL, Corpus de textos galegos escritos por estudantes no ámbito académico](#). Se queremos buscar a clase de palabra estándar, na caixa de consultas CQL deberemos escribir entre corchetes *pos* = seguido da etiqueta entre comiñas, mentres que se desexamos buscar a clase de palabra orixinal (vid. 2.1.2.5 para a diferenza entre ambas) escribiremos *opos* = seguido tamén da etiqueta categorial entre comiñas. Así, por exemplo,

[pos = "NCMP000"] (busca substantivos comúns en masculino e plural na forma léxica estándar)

[opos = "NCMP000"] (busca substantivos comúns en masculino e plural na forma léxica orixinal da/do estudante)

Ademais podemos buscar etiquetas que comecen ou rematen por determinados elementos ou que os conteñan. Por exemplo, escribindo [pos = "NC.\*"] encontraremos todos os substantivos comúns ou escribindo [pos = ".\*FP.\*"] todos os femininos plurais, independentemente da categoría principal

### 3.5 Cancelación da diferenciación entre maiúsculas e minúsculas

Por defecto, o buscador diferencia entre maiúsculas e minúsculas, pero é posible cancelar a diferenciación entre elas mediante *%ci*. Para iso, escribiremos este elemento na caixa de buscas CQL, á dereita da formas buscada: [form="pero" %ci]. No caso de que busquemos unha secuencia de caracteres, a expresión %ci debe integrarse dentro do corchete da palabra ou das palabras en que nos interese a cancelación. Por exemplo, para obter os exemplos de *De feito* e de *de feito*, cómpre ter na caixa de buscas [form="de" %ci] [form="feito"].

### 3.6 Combinación de varios “Tipos de desviación do estándar” ou “Fontes da forma non estándar”

O selector do buscador “Tipo de desviación do estándar” só permite seleccionar unha etiqueta, pero unha mesma forma pode acumular varios tipos de desviacións. Se queremos buscar palabras que acumulen varios delas ao mesmo tempo, podemos acudir a unha busca CQL. Por exemplo, podemos localizar todas as formas que teñan un problema de omisión de acento e de selección léxica inadecuadas e combinalos de calquera dos dous xeitos que se indican a seguir, colocando & entre unha e outra etiqueta. Lémbrese que este elemento supón o cumprimento das dúas condicións a un tempo:

[problem = "(.\*)?L\_w\_su(.\*)?" & problem = "(.\*)?O\_ac\_om(.\*)?" ]

[problem = ".\*L\_w\_su.\*" & problem = ".\*O\_ac\_om.\*" ]

Tamén podemos facer unha busca conxunta de dous criterios mediante |. Así, por exemplo, podemos buscar conxuntamente aqueles elementos que en “Fonte da forma non estándar” teñen ou ben a marca L\_sp ou ben a marca L\_spadapt mediante algunha destas dúas expresións.

[psource = "(.\*)?L\_sp(.\*)?" | psource = "(.\*)?L\_spadapt(.\*)?"]

[psource = ".\*L\_sp.\*" | psource = ".\*L\_spadapt.\*"]

Desta volta obteremos como resultado as formas que cumpren calquera das dúas condicións.

Por outra banda, tamén é posible empregar a expresión .\* para combinar varios subtipos de desviacións (ou de orixes da desviación). Así, por exemplo,

[problem = "O\_ac.\*"]: encontra todos os problemas ortográficos de acentuación, sexan de omisión, adición ou posición errónea.

[problem = ".\*\_ac.\*"]: encontra todos os problemas de acentuación, sexan ortográficos (de omisión, adición ou posición errónea), morfolóxicos ou léxicos.

Finalmente, pode ser tamén empregar a filtraxe negativa. Así, por exemplo, no seguinte caso búscanse todos os casos de formas léxicas non estándares que non son resultado da transferencia do español (nin L\_sp nin L\_spadapt):

[problem= "(.\*)?L\_w\_su(.\*)?" & psource!= "(.\*)?L\_sp.\*(.\*)?"]

#### 4. Algúns sintagmas fusionados

Algunhas expresións pluriverbais, no momento de levar a cabo a estandarización, tiveron que ser fusionadas, posto que a normativización se realizou sobre elas consideradas como unidades<sup>1</sup>. Por exemplo, a expresión *sen embargo* foi estandarizadas no nivel léxico con *porén*, o que determinou a necesidade de fusionar *sen* con *embargo*. Ao levar a cabo esta fusión, a aplicación entende *sen embargo* como unha única palabra, o que ten consecuencias á hora de localizar esta forma. Se realizamos a busca mediante CQL, ao escribir *sen embargo* no buscador, este vai dividir automaticamente a expresión en dúas palabras ([form="sen"] [form="embargo"]) e non localizará a secuencia. Como a aplicación considera *sen embargo* unha única palabra, debemos realizar a seguinte busca [form="sen embargo"]. Por outro lado, se realizamos a busca mediante o xerador de consultas, que unicamente permite buscar unha soa palabra, podemos escribir directamente *sen embargo* na caixa de buscas.

De acordo co indicado, se queremos encontrar todos os usos de *embargo* no corpus (*sin embargo* e *sen embargo*), o que debemos facer é ou ben empregar comodíns en CQL (*\*embargo*) ou ben buscar con “contén” *embargo* no xerador de consultas. En xeral, e dado que as fusións que se realizan non son sempre previsibles, para asegurarse de encontrar todos os exemplos dunha forma, recoméndase empregar “contén” e non “igual”. Así, por exemplo, para encontrar todos os usos do pronome *se* debería realizarse unha busca con “contén” *se* en "Versión final estudante" e con “contén” "PP3CN000" en Clase de palabra estándar. Só así encontraremos,

---

<sup>1</sup> As expresións lexicalizadas só se fusionan cando a estandarización se realiza sobre toda a unidade. No caso de que non haxa estandarización ou esta só afecte a un dos compoñentes, cada elemento mantén a súa propia independencia.

por exemplo, a secuencia “lévanse a cabo”, que nun texto se fusionou para realizar unha estandarización da locución no nivel semántico. Nótese que na clase de palabra tamén cómpre escoller a opción “contén” (posto que a expresión foi etiquetada coa suma das diferentes categorías gramaticais (VMIP3P0+ PP3CN000+SP+NCMS000)). O mesmo é aplicable ao lema e ao lema orixinal.

## 5. Os resultados das buscas

### 5.1 Buscas de texto

#### 5.1.1 Ordenación dos resultados

Por defecto, os resultados das buscas (directamente en CQL ou a través do xerador de consultas) ordénanse alfabeticamente polas distintas formas que resultan da busca (por exemplo, se se busca o lema *men*, primeiro aparecen os exemplos con *men*, despois con *meus*, *miña*, *miñas*). Como segundo criterio ordenador, dentro de cada variante aparecen en primeiro lugar as formas con maiúscula e despois aquelas con minúscula. Como terceiro criterio, a aplicación agrupa as formas buscadas que figuran no mesmo texto e pola orde en que aparecen neste (por exemplo, se buscamos *que*, aparecen xuntos todos os exemplos de *que* que figuran nun mesmo texto). Pero podemos cambiar o criterio de ordenación e escoller como tal a primeira palabra da dereita ou a primeira palabra da esquerda. Para isto cómpre premer en “opcións”, sinalado na Figura 13 e en “Ordenar por” seleccionar “Contexto á dereita” ou “Contexto á esquerda”.

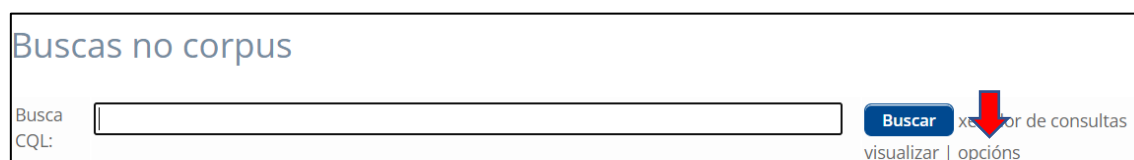


Figura 13. Elemento “opcións” para cambiar a ordenación dos resultados

Así, por exemplo, podemos buscar o lema *grave* e ordenando pola dereita teremos xuntos todos os exemplos da combinación do adxectivo con *problema* ou con *consecuencia*. De igual xeito, podemos buscar na busca CQL a secuencia *de que* e ordenar pola esquerda para termos agrupados todos os exemplos coa mesma palabra previa.

Ademais, premendo na opción “Forma”, o buscador ofrece en primeiro lugar todas as formas con maiúscula inicial (con ordenación alfabética) e a seguir todas as formas con minúscula inicial, tamén con ordenación alfabética.

Por outra banda, seleccionando “Posición no corpus” tamén é posible agrupar os exemplos por textos sen ter en conta como primeiro criterio ordenador a forma concreta que figura neles. Por exemplo, se buscamos o lema *haber* e ordenamos por posición no corpus, o buscador presentará como consecutivas as liñas de concordancia das formas *baxa* e *hai* do texto 827. Se o criterio ordenador fose palabra, tales formas aparecerían separadas.

#### 5.1.2 Visualización dos resultados

En canto ao número de palabras da liña de concordancia, esta pode modificarse. Por defecto figuran cinco elementos (incluídos os signos de puntuación) á dereita e á esquerda da forma buscada, pero esta cifra pode cambiarse en “Opcións” a 3, 4, 6 ou 7.

Por outro lado, por defecto o resultado figura na “Versión final da/do estudante”, pero premendo sobre os botóns que aparecen por riba da concordancia (vid. Figura 14) pódese optar pola visualización na “Transcrición completa” (ou Versión inicial) ou ben nas diferentes capas de estandarización (vid. 5.1.3 para máis información sobre estas). Ao mesmo tempo, poden facerse visibles na propia concordancia as diferentes anotacións (lema estándar e orixinal, categoría gramatical estándar e orixinal, tipo de desviación do estándar...). premendo nos botóns correspondentes (sobre estas anotacións, vid. 5.1.3).

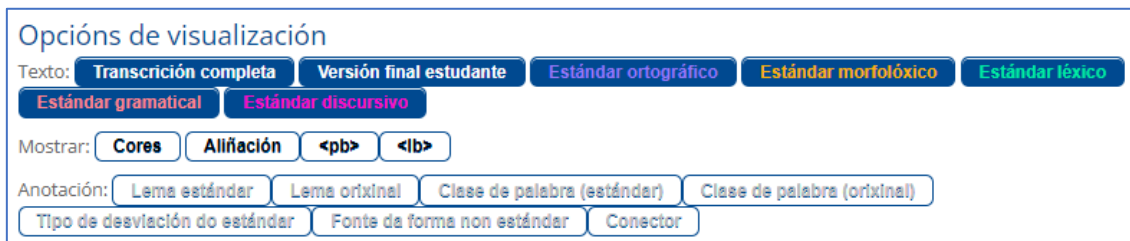


Figura 14. Selección da capa de visualización e das anotacións nas concordancias

As liñas de concordancia amósanse agrupadas en páxinas de 100 liñas. Podemos avanzar ata o seguinte grupo de 100 liñas de concordancia premendo en “seguinte” (e, no seu caso, poderemos volver atrás premendo en “anterior”) (vid. Figura 14).

Finalmente, premendo en Mostrar contexto, á esquerda de cada liña de concordancia, accedemos ao texto completo.

### 5.1.3 Visualización dos textos

Os textos, aos que se accede mediante "Mostrar contexto" despois de cada busca, poden verse en sete diferentes capas, premendo sobre os correspondentes botóns con fondo azul:

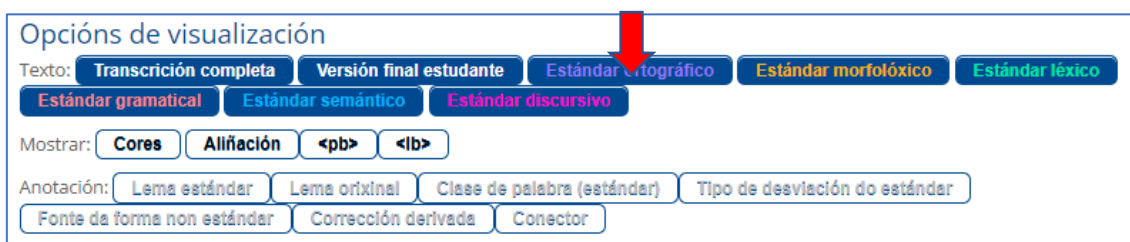


Figura 15. Selección da capa de visualización nos textos

1. *Transcripción completa*. Inclúe as formas riscadas (en gris e cunha liña por riba), as formas engadidas (destacadas en vermello) e as formas de lectura dubidosa (destacadas nun ton gris verdoso).
2. *Versión final estudiante*. Ofrece o texto sen as formas riscadas.
3. *Estándar ortográfico, morfológico, léxico, gramatical, semántico e discursivo*. Son as seis capas progresivas de estandarización. As estandarizacións feitas en cada nivel destácanse mediante diferentes cores (vid. Figura 16), pero para iso a opción "Cores" (vid. figura 15) ten que estar activada. As estandarizacións ortográficas figuran en lila, as morfolóxicas en laranxa, as léxicas en verde, as gramaticais en salmón, as semánticas en azul e as discursivas en fucsia.

A cociña estase a converter nunha verdadeira arte. Mentres que **hai** uns anos os restaurantes caros se caracterizaban por ter unha boa calidade de produto e estar ben **cociñado**, na actualidade para **ser** un restaurante prestixioso requírese moito máis. O anterior xa se presupón, pero **é** que **ademais** debe ser constantemente **innovador**. Se vas agora a un restaurante prestixioso **verás** que cambiou **drasticamente** dende hai uns anos. Agora **probarás** comidas novas, innovadoras **técnicas** de cociñar e inauditas **mesturas** de alimentos nun **prato**.

Figura 16. Fragmento de texto con estandarizacións pertencentes a diferentes niveis

Debe terse en conta que cada capa de herda as estandarizacións das capas previas, de tal modo que a medida que prememos en capas sucesivas, vemos as formas normalizadas previas e as propias da capa en que nos situamos. Claro está, non é posible a acumulación de cores, de xeito que, se, por exemplo, corriximos no nivel ortográfico a forma *este*, por *esté*, e posteriormente levamos a cabo a estandarización morfolóxica, por *estea*, na capa de estandarización morfolóxica (e nas seguintes) só veremos a forma *estea* en cor laranxa.

Baixo os botóns que permiten seleccionar a capa de visualización figuran outros catro: "cores", "aliñación", "pb" e "lb". Como xa indicamos, o primeiro deles serve para activar as cores das formas corrixidas. Con respecto ao botón "aliñación", cómpre sinalar que existen dúas posibilidades de visualización dos textos: sen respectar os finais de liñas orixinais, co texto ocupando, ao longo, todo o espazo da caixa (Figura 17) ou respectando os finais de liña do orixinal (Figura 18). O botón "aliñación" permite mudar de sistema de presentación.

Este texto mostra un problema familiar moi común a día de hoxe entre pais e fillos adolescentes.

Estes conflitos danse cando os fillos teñen de trece a dezasete anos, nestas idades os adolescentes tenden a desafialas normas moitas veces impostas polos seus pais. Isto pode derivar en casos de rebeldía moi problemática. Aquí, os fillos atópanse en idades moi complexas.

Problemas como este poden levarse ao extremo como pode verse en programas como "Hermano mayor" onde os rapaces causan problemas serios nas casas. Débese en boa parte ao consentimento dos pais e a mala educación que provoca isto nos fillos.

Un dos grandes problemas de España é precisamente este con adolescentes consentidos en moitos fogares.

Como conclusión, queda que para lograr diminuír os conflitos o máis posible, xa que son inevitables, é unha correcta educación e ensino.

Figura 17. Visualización do texto sen respectar os finais de liña orixinais

Este texto mostra un problema familiar moi común a día de hoxe entre pais e fillos adolescentes.

Estes conflitos danse cando os fillos teñen de trece a dezasete anos, nestas idades os adolescentes tenden a desafialas normas moitas veces impostas polos seus pais. Isto pode derivar en casos de rebeldía moi problemática. Aquí, os fillos atópanse en idades moi complexas.

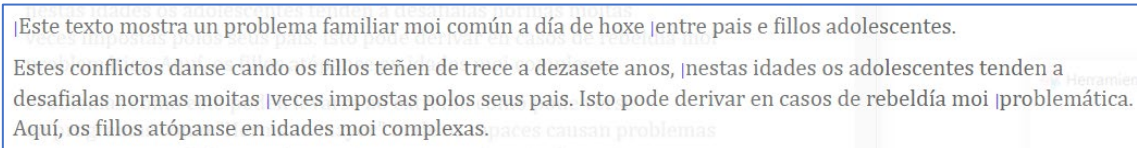
Problemas como este poden levarse ao extremo como pode verse en programas como "Hermano mayor" onde os rapaces causan problemas serios nas casas. Débese en boa parte ao consentimento dos pais e a mala educación que provoca isto nos fillos.

Un dos grandes problemas de España é precisamente este con adolescentes consentidos en moitos fogares.

Como conclusión, queda que para lograr diminuír os conflitos o máis posible, xa que son inevitables, é unha correcta educación e ensino.

Figura 18. Visualización do texto respectando os finais de liña orixinais

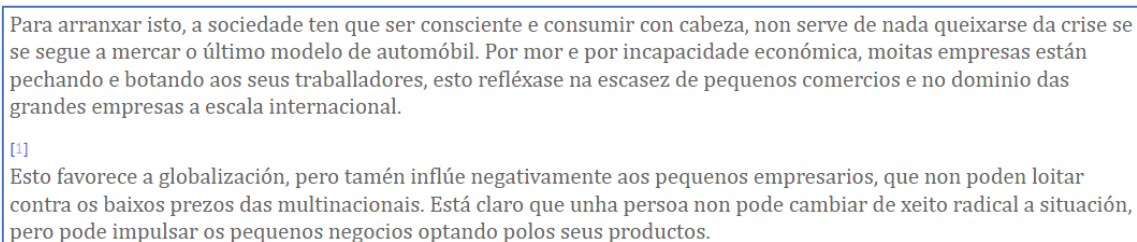
En calquera caso, podemos manter a aliñación corrida e marcar o final de cada liña activando <lb>, tal e como se ve na Figura 19:



Este texto mostra un problema familiar moi común a día de hoxe entre pais e fillos adolescentes. Estes conflitos danse cando os fillos teñen de trece a dezasete anos, nestas idades os adolescentes tenden a desafiar as normas moitas veces impostas polos seus pais. Isto pode derivar en casos de rebeldía moi problemática. Aquí, os fillos atópanse en idades moi complexas.

Figura 19. Visualización do texto despois da activación de <lb>

Por outra banda, o primeiro sistema de visualización non informa sobre os cambios de páxina, mentres que o segundo ofrece unha tenue liña gris naqueles lugares onde a/o estudante pasa a escribir nunha páxina nova. En calquera caso, en calquera dos dous sistemas de visualización pódese premer en <pb> e o sistema incorporará un número entre o final de cada páxina e o inicio da seguinte, tal e como se ve na figura 20:



Para arranxar isto, a sociedade ten que ser consciente e consumir con cabeza, non serve de nada queixarse da crise se se segue a mercar o último modelo de automóbil. Por mor e por incapacidade económica, moitas empresas están pechando e botando aos seus traballadores, esto refléxase na escasez de pequenos comercios e no dominio das grandes empresas a escala internacional.

[1]

Esto favorece a globalización, pero tamén inflúe negativamente aos pequenos empresarios, que non poden loitar contra os baixos prezos das multinacionais. Está claro que unha persoa non pode cambiar de xeito radical a situación, pero pode impulsar os pequenos negocios optando polos seus produtos.

Figura 20. Visualización do texto despois da activación de <pb>

Finalmente, baixo os botóns que acabamos de comentar poden aparecer os seguintes: "Lema estándar", "Lema orixinal", "Clase de palabra estándar", "Clase de palabra orixinal", "Tipo de desviación do estándar", "Fonte da forma non estándar", "Corrección derivada" e "Conector" (unicamente se visibilizan aquelas etiquetas ás que se lles asignou algún valor no proceso de anotación do texto). Premendo nun ou en varios deles, ofrécese baixo cada palabra a anotación correspondente. Así, por exemplo, na Figura 21 vemos un fragmento de texto en que activamos lema estándar e "Tipo de desviación do estándar":





Figura 21. Visualización dun texto despois da activación de lema orixinal e tipo de desviación do estándar

Con respecto a "Corrección derivada" (DC) esta anotación emprégase en dúas situacións. En primeiro lugar, cando se produce unha corrección derivada doutra (por exemplo, a substitución de minúscula inicial por maiúscula inicial despois de corrixir, poñamos por caso, unha coma por un punto). Evidentemente, nun caso como este non ten sentido identificar un problema de omisión de maiúsculas (pois a/o estudante non comete ningún erro deste tipo, dado que escribira unha coma). Por tal motivo, simplemente corriximos e engadimos DC. En segundo lugar, emprégase esta mesma etiqueta naqueles casos en que unha mesma desviación do estándar ten varias manifestacións, como, por exemplo, nos erros de concordancia que afectan a varias palabras: en tales casos, só se asigna a etiqueta de "Tipo de desviación do estándar" a unha das formas afectadas, normalmente a primeira, mentres que as restantes son simplemente corrixidas e anotadas con DC.

As anotacións visibles desaparecen premendo de novo nos botóns correspondentes, pero tamén é posible acceder a elas situándose sobre as palabras do texto. Nese caso desprégase un cadro con todas as anotacións da palabra, tal e como se pode ver na figura 22.



Como sempre o noso Estado caracterízase polas súas modas pasaxeiras. A última moda no sector estudantil e de traballo **ven** sendo o chamado gastronomía. Obviamente, Galiza é unha comunidade que se caracteriza pola súa variedade gastronómica o que fai atraer a moitos extranxeiros o que fixo que este sector económico medrase e poidese optar por máis persoas para os postos de traballo.

Nesta época o que máis saídas ten é un [...] ciclo, como o que se atopa en Santiago de Compostela, ou algunha carreira sobre cociña ou gastronomía, xa que como podemos ver ata o de agora é o sector que máis reduce o paro, xa sexa en Galiza ou en España en xeral, trátase dun sector que atrae a moitos extranxeiros e que se caracteriza por atopar facilmente traballo, xa sexa nun bar de cociñeiro, nun hotel ou nun restaurante.

ven	
Estándar ortográfico	vén
Lema estándar	vir
Clase de palabra (estándar)	Verbo (VMIP3S0) Main; indicative; present; third; singular
Tipo de desviación del estándar	O_ac_om

Figura 22. Información sobre as palabras do texto

Finalmente, podemos encontrar algunhas estandarizacións que non teñen ningún tipo de anotación en "Tipo de desviación do estándar" nin tampouco a etiqueta DC. Trátase neste caso de formas que foron estandarizadas sobre o texto, como as restantes, pero que foron anotadas mediante anotacións stand-off (vid. 2.1.2.7). Para ver as etiqueta correspondentes, cómpre premer en "Anotación multitoken" na parte inferior da pantalla (vid. Figura 23). Esta opción só será visible no caso de que o texto teña algunha anotación deste tipo.

anos Galiza non foi un lugar de atracción turística aínda que nos últimos anos sí, o que fixo que comezasen a facerse e a crear máis estudos de hostelería e se comezasen a crear máis postos de traballo neste ámbito.

Lenda: **Lectura difícil** • Texto riscado • **to engadido**

Descargar vista actual como txt • **Anotación multipalabra**

Figura 23. Acceso ás anotacións e correccións realizadas en stand-off

Premendo sobre "Anotación multipalabra", encontraremos subliñadas as palabras e secuencias estandarizadas, a etiqueta correspondente a cada anotación, así como, na maior parte dos casos, unha proposta de estandarización (vid. Figura 24)

Na actualidade hai moitas técnicas na cociña que nos descoñecemos por completo e soamente empreganse na alta cociña. Tamén utilízanse moitos máis ingredientes e produtos de calidade, cos cales logran unha serie de sabores pouco comúns e os cales despiertan o gusto do consumidor.

Outro aspecto foi a cociña media a cal na actualidade a súa oferta é moi boa comparando a súa calidade-prezo, nisto ten un factor moi importante a xente que traballa e ten un descanso corto entón come fora da casa o que provoca certo aumento no sector e correspondente popularidade.

Eu penso que a popularidade da cociña seguirá incrementando debido ás evolucións e a inmensa cantidade de produtos sen descubrir e ingredientes e produtos cos que xogar para crear un bo produto co cal o coci creador se identifique con el. E por último, penso que será duradeiro xa que nos precisamos comer todos os días e é un sector o cal non pode desaparecer senón que evolucionará.

vista de texto

a cal na actualidade a súa oferta é moi boa	
nivel lingüístico	Grammar (Grammar)
Código	G_str-ref-top_su
Forma corrixida	cuxa oferta na actualidade é moi boa

Figura 24. Visualización das anotacións e estandarizacións realizadas en stand-off

Premendo na parte inferior esquerda da pantalla ("Vista de texto"), volvemos á pantalla de partida.

Por outro lado, na parte superior da pantalla temos información sobre os metadatos do texto.

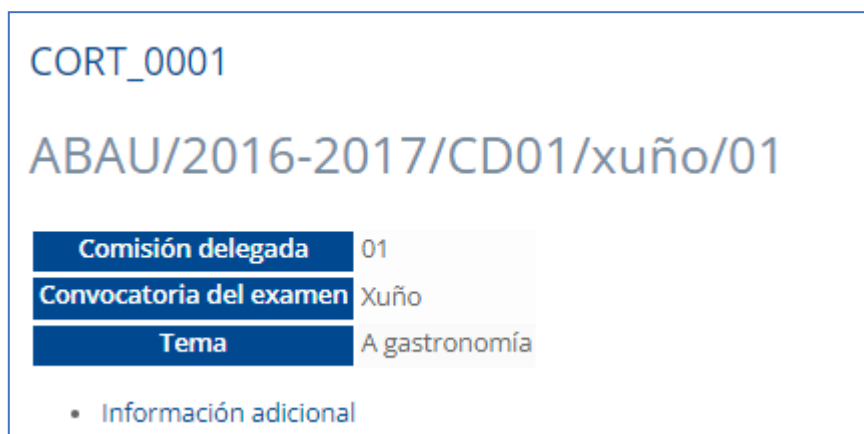


Figura 25. Información básica da cabeceira dos textos

A información visible por defecto é a comisión delegada, a convocatoria do exame e o tema deste, pero premendo en ampliar información (vid. Figura 25) obteremos tamén outros datos descritivos da mostra empregada (o ano dos textos, o nivel, o tipo de texto e o ámbito da proba), así como datos cuantitativos sobre o texto en concreto (número de palabras, número de lemas, densidade léxica, número de enunciados, media de palabras por enunciado, número de palabras do enunciado máis longo, número de palabras do enunciado máis curto, número de parágrafos e media de enunciados por parágrafo) (vid. 2.2). Unha vez despregada esta información, premendo en "Información básica" pode volveuse á cabeceira orixinal.

#### 5.1.4 Descarga dos resultados das buscas e dos textos

Despois de realizar unha busca é posible descargar os resultados en formato .txt premendo na opción "Descargar" que figura ao final da concordancia.

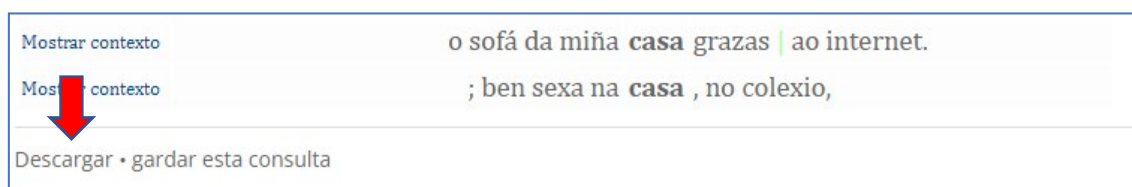


Figura 26. Opción para descargar os resultados en txt.

A opción de "gardar esta consulta" só funciona para as persoas con permisos de edición.

Ademais, os textos tamén poden ser descargados como .txt na vista que en cada momento se estea a visualizar (transcripción completa, aínda que os riscados que identifican as formas borradas non se conservan, versión final da/do estudante ou as diferentes capas de estandarización). Pódense incluír tamén, se se desexa, as anotacións.

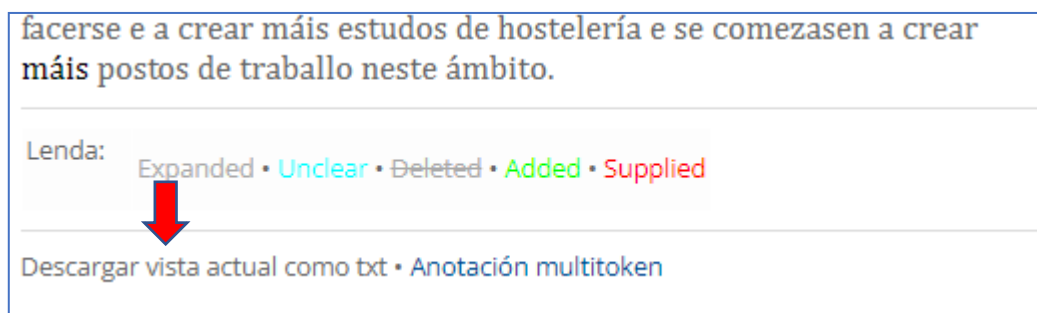


Figura 27. Descarga dos datos en .txt

## 6. Consultas sobre frecuencia

Outra opción que nos ofrece o buscador é a de obter datos de frecuencia e distribución. Para iso debemos acudir á sección “Consultas sobre frecuencia” que encontramos despois dos resultados dunha busca (Figura 28).

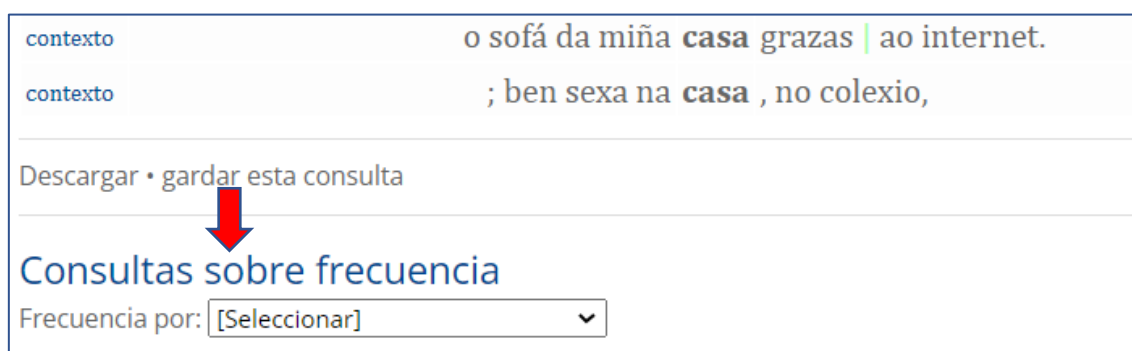


Figura 28. Consultas sobre frecuencia

Así, por exemplo, se queremos obter información rápida sobre a distribución dos casos de adición de preposición entre as diferentes preposicións, deberemos buscar o código G\_prep\_ad en "Tipo de desviación do estándar". Unha vez feita a busca deberemos acudir, en “Consultas sobre frecuencia” a “Frecuencia por” e seleccionar “Lema estándar”. O resultado que obtemos por defecto é unha táboa como a que segue, coa distribución por preposicións e información sobre a frecuencia en números absolutos (segunda columna) e co índice por dez millóns de palabras na terceira columna (así, por exemplo, a cifra 14695,08 que se indica para *en* debe lerse nos seguintes termos: o 0,14695% dos casos de lema *en*, están etiquetados con G\_prep\_ad).

a	339	608,399.14
de	27	21,166.51
en	10	14,695.08
por	4	13,029.32
con	3	18,927.44
para	3	15,337.42
ata	1	50,505.05
ir+a+ser+que	1	10,000.000
sobre	1	35,087.72

Táboa 6. Distribución do código G\_prep\_ad por preposición

Seleccionando o modo de presentación que nos interese en “Gráfica”, os resultados poden ofrecerse tamén mediante gráfica de sectores, de barras, de liñas, de dispersión e mediante un histograma. Así mesmo, na última opción, “Estatística”, atopamos información diversa (media, mediana, desviación estándar...). Para esta información estatística pode seleccionarse en “Contar” a opción WPM (palabras por millón). Ademais, en “descargar” os datos poden descargarse en diferentes formatos.

Do mesmo xeito, se queremos saber, por exemplo, a distribución da omisión dos signos de puntuación entre os diferentes signos, debemos seleccionar D\_pm\_om e a seguir, en “Frecuencia por”, escoller “estándar discursivo” (téñase en conta que as formas que levan unha etiqueta de omisión nos niveis gramatical, semántico ou discursivo non son lematizadas, de xeito que non son recuperables mediante a busca por lema).

Outra opción de busca que nos ofrece a aplicación é aquela que nos permite encontrar directamente a distribución das diferentes formas (“Versión final estudante”) que resultan dunha busca. Deste xeito, se facemos, por exemplo, unha busca por un lema (por exemplo, *facer*) e prememos en “Versión final estudante” encontraremos todas as formas ás que se lle asignou o citado lema estándar, canda a súa frecuencia. De igual maneira, podemos buscar, por exemplo, todas as formas coa clase de palabra SP e a seguir obter a listaxe de preposicións coa súa frecuencia.

## Anexo I

### CÓDIGOS QUE IDENTIFICAN O TIPO DE DESVIACIÓN DO ESTÁNDAR

A seguir preséntanse os diferentes códigos que identifican as desviacións do estándar, co valor de cada código e exemplos. O primeiro elemento, sempre en maiúscula, indica o nivel lingüístico ao que afecta a desviación: O: Ortografía, M: Morfoloxía, L: Léxico, G: Gramática (sintaxe), S: Semántica e D: Discurso. O segundo elemento, o que ofrece maior variación, concreta o aspecto afectado dentro do nivel lingüístico xeral (*ac*: acentuación, *gen* xénero, *ref*: elementos con valor referencial, *prep*: preposicións etc.). Como se pode observar, tales elementos poden ser clases de palabras, trazos lingüísticos, categorías funcionais etc. Cando queremos indicar unha unidade léxica en xeral úsase “w” (“word”), como en L\_w\_su (substitución dunha unidade léxica estándar por unha non estándar). Finalmente, o terceiro elemento explica a diferenza existente entre a forma estándar e a forma proporcionada pola/o estudante, sempre desde esta última perspectiva, ou o trazo que fai que a forma ou secuencia escrita non sexa estándar (*ad*: adición, *om*: omisión, *su*: substitución, *np*: posición incorrecta ou inadecuada, *tr*: transposición, *un*: unión, *spl*: división, *com*: excesiva complexidade, *unin*: inintelixibilidade).

Código	Valor	Exemplo
D_con_ad	Adición dun conector innecesario	Por outra parte, están despois as marcas da terra, como pode ser a lalinesa Toypes, con fábrica en Galicia <b>tamén</b> (non hai referencia previa a outra empresa con fábrica en Galicia)
D_con_om	Omisión dun conector necesario	Na televisión pódense ver programas de cociña, tanto para nenos coma para adultos, un exemplo é "Master Chef". A cociña é algo que che ten que gustar e tes que disfrutar dela para triunfar. Sería preciso un conector para introducir o novo tema (por exemplo, <i>Por outra banda</i> )
D_con_su	Substitución dun conector discursivo por un conector sintáctico (que introduce subordinadas)	As necesidades de hoxe en día tamén implican un avance técnico moito máis superior polo que hoxe en día fanse cousas que fai dez anos eran impensables, como a video conferencia a tempo real, mensaxería instantánea... <b>Polo que</b> hoxe en día a necesidade de profesionais capacitados

		para o deseño o innovación cada vez e maior. (Emprego de <i>Polo que</i> en vez de <i>así pois</i> ou outro conector discursivo)
D_np_ad	Adición innecesaria dun novo parágrafo	[Obs: debe terse en conta que neste caso non é posible levar a cabo a correspondente estandarización, de modo que se asigna o código pero non se modifica a distribución dos parágrafos]. Pódese ver un exemplo en CORT_0157 (entre o segundo e o terceiro parágrafo).
D_np_om	Omisión da creación dun novo parágrafo	[Obs: debe terse en conta que neste caso non é posible levar a cabo a correspondente estandarización, de modo que se asigna o código pero non se modifica a distribución dos parágrafos]. Pódense ver dous exemplos en CORT_0394.
D_pm_ad	Adición dun signo de puntuación innecesario	Pódese ver un exemplo en CORT_0469.
D_pm_om	Omisión dun signo de puntuación necesario	Pódese ver un exemplo en CORT_0469.
D_pm_su	Selección dun signo de puntuación inadecuado	Pódense ver dous exemplos en CORT_0469.
D_pm_wp	Colocación inadecuada dun signo de puntuación	Pero agora e distinto, a xente ten un maior acceso aos alimentos, e ademáis, despois de traballar moito a ninguén lle apetece cociñar (Por [...] aos alimentos e, ademáis, despois de...)
D_ref_ad-su	Adición inadecuada dun elemento con valor referencial ou selección inadecuada deste	Certas persoas cociñan, e despois un gastrónomo probáas elixindo a que lle gustou máis. (No texto non hai un referente para o clítico de <i>probáas</i> ). Por outro lado, tamén é moi bo que aparezan nas nosas cidades restaurantes doutras cidades que cada vez hai máis para aprender un pouco máis <b>desta</b> cultura. (Emprego de <i>esta</i> en vez de <i>a súa</i> )
D_ref_om	Omisión dun elemento con valor referencial que evita a repetición dunha frase ou palabra	A cociña na actualidade ten unha popularidade moi alta provocada pola saturación de programas de televisión relacionados <b>coa cociña</b>

		(en vez de “relacionados con ela”)
<b>D_reg_ad</b>	Adición dunha palabra ou secuencia inadecuadas para o rexistro académico	E eso meus amigos, non se soluciona con inxenieros (adición inadecuada de <i>meus amigos</i> ) é raro que á hora de comer non boten ningún programa relacionado coa cociña na televisión (selección inadecuada de <i>boten</i> en vez de <i>emitan</i> )
<b>D_reg_om</b>	Omisión dunha palabra necesaria para a adecuación ao rexistro académico	Débese a todas as reservas que xa hai, polo que con sorte, mínimo tócache esperar 6 meses para poder ir (emprego de <i>mínimo</i> en vez de <i>como mínimo</i> )
<b>D_reg_su</b>	Selección dunha palabra ou secuencia inadecuadas para o rexistro académico	é raro que á hora de comer non boten ningún programa relacionado coa cociña na televisión (selección inadecuada de <i>boten</i> en vez de <i>emitan</i> )
<b>D_ut_com</b> (stand-off)	Enunciado excesivamente longo ou complexo	Vid. un exemplo en CORT_0962
<b>D_ut_unin</b> (stand-off)	Enunciado inintelixible	Este feito descrito ten moitas ventaxes. e desventaxes se o proceso xorde a inversa.
<b>D_w_om</b>	Elipse inadecuada	No texto, por exemplo, preséntanos a historia de Carmela de Chousa (o suxeito de <i>preséntanos</i> non é recuperable do contexto previo)
<b>G_conj_ad</b>	Adición dunha conxunción que dá lugar a unha secuencia agramatical	Se ben, é certo que non se pode vivir sen consumir, unha simple botella de auga xa custa cartos, <b>mais</b> a maior parte dos produtos que mercamos non son totalmente necesarios
<b>G_conj_om</b>	Omisión dunha conxunción necesaria para ter unha secuencia gramatical	Polo tanto, concursos e programas de televisión nos cales non so se preparen platos complexos e inhabituais, <b>senón tamén</b> se prometa con eles sabores exquisitos e diferentes, supoñen todo un atractivo para a poboación (omisión de <i>que</i> : “senón que tamén”)
<b>G_conj_su</b>	Selección inadecuada dunha conxunción	<b>Coma</b> di o refrán "Somos o que comemos"

G_det_ad	Adición dun determinante innecesario	amas <b>da</b> casa
G_det_om	Omisión dun determinante necesario	debido a que alta cociña era máis cara
G_det_su	Selección inadecuada dun determinante	non se lle pode quitar mérito <b>algún</b> aos cociñeiros
G_gen_su	Problema de concordancia por selección inadecuada de xénero	<b>unha época no</b> que a moda ten
G_num_su	Problema de concordancia por selección inadecuada de número	<b>A relacións</b> que existe
G_per_su	Problema de concordancia por selección inadecuada de persoa	Necesitamos de unha ampla produción para que poida existir consumo e que esta sexa máis rápida que a acción que teñen os consumidores para <b>autoabastecernos</b> .
G_pos_su	Selección dunha categoría gramatical inadecuada	A idea de aprender a cociñar ou ser <b>gastronómico</b> impulsa a creatividade da xente
G_prep_ad	Adición dunha preposición inadecuada desde o punto de vista gramatical. Non se inclúen aquí os casos de construción verbal inadecuada, que van etiquetadas con G_vstr_su	van <b>a</b> seguir abertos
G_prep_om	Omisión dunha preposición necesaria desde o punto de vista gramatical. Non se inclúen aquí os casos de construción verbal inadecuada, que van etiquetadas con G_vstr_su	nos dámonos de conta que os consumidores (por “dámonos de conta de que”)
G_pron-en_wp	Colocación enclítica dun pronome que debería ir proclítico	nos dámon <b>os</b> de conta que os consumidores
G_pron-pro_wp	Colocación proclítica dun pronome que debería ir enclítico	a xente <b>lle</b> gusta comprar moitas cousas
G_pron_ad	Adición dun pronome que resulta agramatical	pasamon <b>os</b> o día consumindo
G_pron_om	Omisión dun pronome, dando lugar a unha secuencia agramatical	Atopamos nun mundo no que nada ten valor por si só
G_pron_su	Selección agramatical dun pronome	Existe moita xente que opina coma eu
G_vmt_su	Selección dun tempo / modo verbal que dá lugar a unha secuencia agramatical. Inclúense aquí os usos inadecuados do xerundio	Por outro lado temos aos consumidores quen se o produto é bo, barato e se ademais aos ollos destes é bonito <b>adquiran</b> con maior frecuencia o produto



<b>G_vo_su</b>	Selección dunha voz gramaticalmente inadecuada (construcións con complemento axente, pero sen voz pasiva)	O enfoque mercantil o recibimos de maneira inconsciente por toda a sociedade
<b>G_str-ref-top_su</b> (stand-off)	Estrutura gramaticalmente inadecuada por reformulación ou topicalización inadecuada dalgún elemento	a relación de consumismo e produtividade están íntegramente ligados
<b>G_str_su</b> (stand-off)	Estrutura gramaticalmente inadecuada (agás os casos anotados coa etiqueta anterior e coa seguinte)	A nosa cultura que temos neste campo dende cada vez que saímos a tomar algo as afamadas tapas ata que como o noso marisco das nosas rías non hai un igual de bó
<b>G_vstr_su</b>	Selección dunha construción verbal inadecuada (verbo transitivo construído intransitivamente, verbo que rexe preposición construído sen preposición e ao revés...)	adoitamos a querer sempre o último
<b>G_w_ad</b> (stand-off)	Adición dalgún elemento que fai a estrutura gramaticalmente inadecuada	Case todas ou todos as persoas
<b>G_w_om</b>	Omisión dunha palabra ou secuencia (excluíndo preposicións, determinantes e conxuncións) necesaria para a gramaticalidade do texto	En conclusión a cociña avanzou moito e agora xa non é simplemente cociñar senón que <b>é unha nova de</b> expresión artística
<b>L_ac_su</b>	Selección dunha palabra non estándar por diferenzas na acentuación	élite
<b>L_gen_su</b>	Selección dunha palabra non estándar por diferenzas no xénero	leite (masculino)
<b>L_num_su</b>	Selección en singular dunha palabra ou expresión lexicalizada en plural	vacación
<b>L_w_su</b>	Selección dunha palabra non estándar por diferenzas que van máis alá da acentuación, do xénero ou do número	plato
<b>M_ac_su</b>	Selección dunha forma verbal non estándar por diferenzas na acentuación	cantábamos
<b>M_cv_su</b>	Selección dunha forma verbal composta	ha chegado
<b>M_eval_su</b>	Selección dun diminutivo ou aumentativo non estándar	graniño
<b>M_gen_su</b>	Selección dun feminino, masculino ou neutro non estándar de substantivos, adxectivos ou palabras gramaticais	catalana, miño (por <i>men</i> ), esto

<b>M_num_su</b>	Selección dun plural non estándar de substantivos, adxectivos ou palabras gramaticais	sociales, estos
<b>M_v_su</b>	Selección dunha forma verbal non estándar por diferenzas que van máis alá da acentuación	aparezcan
<b>O_ab_su</b>	Selección dunha abreviatura, sigla ou símbolo non estándar	E.R:E (por ERE)
<b>O_ac_ad</b>	Adición de acento gráfico	ademáis
<b>O_ac_om</b>	Omisión de acento gráfico	aínda
<b>O_ac_wp</b>	Colocación do acento gráfico nunha vogal átona	maís
<b>O_as_ad</b>	Adición de asimilación que non debe representarse graficamente no estándar	permítenos (por <i>permítennos</i> )
<b>O_as_om</b>	Omisión de asimilación que debe representarse graficamente no estándar	cociñarlas
<b>O_cons_ad</b>	Adición de consoante que non altera ou pode non alterar a pronuncia ou que representa unha pronuncia non estándar pero posible da palabra	producción
<b>O_cons_om</b>	Omisión de consoante que non altera ou pode non alterar a pronuncia ou que representa unha pronuncia non estándar pero posible da palabra	satisfacción
<b>O_cons_su</b>	Substitución dunha consoante ou dun dígrafo por outra letra ou dígrafo que representa a mesma pronuncia ou unha pronuncia próxima	avarca
<b>O_cont_ad</b>	Adición de contracción	nalgo
<b>O_cont_om</b>	Omisión de contracción	de esta
<b>O_cont_su</b>	Selección dunha forma contracta non estándar	ca (por <i>coa</i> )
<b>O_die_om</b>	Omisión de diérese	antigüedade
<b>O_fw_su</b>	Selección dunha forma gráfica non estándar dunha forma estranxeira	graffitti
<b>O_hy_ad</b>	Grafía con trazo en vez de conxunta ou separada	tóda-las
<b>O_let_ad</b>	Adición, probablemente por lapsus, dunha ou varias vogais ou consoantes que dá lugar a unha pronuncia diferente	coincidencia
<b>O_let_om</b>	Omisión, probablemente por lapsus, dunha ou varias vogais ou consoantes que dá lugar a unha pronuncia diferente	apredizaxe

<b>O_let_su</b>	Substitución, probablemente por lapsus, dunha ou varias vogais ou consoantes por outra(s) letra(s) que non representan a mesma pronuncia ou unha pronuncia posible	candidade
<b>O_let_tr</b>	Transposición dunha ou varias vogais ou consoantes	borbando
<b>O_uc_ad</b>	Adición inadecuada de maiúscula	efecto “Boomerang”
<b>O_uc_om</b>	Omisión inadecuada de maiúscula	ferrari
<b>O_vow_om</b>	Omisión de vogal que non altera ou pode non alterar a pronuncia	sobrexplotación
<b>O_vow_su</b>	Substitución dunha vogal por outra letra ou dígrafo que representa unha pronuncia non estándar pero posible da palabra ou en todo caso unha pronuncia próxima	saerán
<b>O_wb_spl</b>	Escritura separada en vez de conxunta ou con trazo	tal vez
<b>O_wb_un</b>	Escritura conxunta en vez de separada ou con trazo	apiques
<b>S_w_ad</b> (stand-off)	Adición dalgún elemento redundante ou que fai a estrutura semanticamente inadecuada	A gastronomía todavía segue crescendo aínda hoxe
<b>S_w_wf_su</b>	Uso/significado non estándar dunha palabra ou expresión que se utiliza en vez doutra palabra ou expresión da súa mesma familia léxica	calidade <b>alimenticia</b> por “calidade alimentaria”
<b>S_w_om</b>	Omisión de palabras necesarias para expresar o sentido buscado ou cando menos esperable	lamentablemente estamos nunha economía na que temos que consumir en exceso para que a produción non se deteña e haxa despidos (falta <i>non</i> : “non haxa despidos”)
<b>S_w_su</b>	Selección dunha palabra á que se lle atribúe un uso ou significado non estándar	O mundo da hostelería volverá as súas <b>facultades</b> habituais

