

Introducción a CQL para SkE

Corpus Query Language

Qué es CQL

CQL (*Corpus Query Language*) es el lenguaje de consulta avanzada de Sketch Engine. Disponible también en ODE, PS y OSTA (con ligeras variaciones).

Permite formular criterios complejos para localizar uno o varios *tokens* (palabras gráficas) según sus atributos:

forma (atributos: word o form, según los corpus)

lema (atributo: lemma)

etiqueta morfosintáctica (atributos: tag o pos), según los corpus, etc.

Estructura básica de una consulta CQL

Toda consulta en CQL describe un token (palabra) mediante pares atributo-valor

[atributo = "valor"]

atributo

Qué aspecto de la palabra buscamos:
word (forma gráfica), lemma (lema) o tag (etiqueta gramatical)

"valor"

El texto que debe tener ese atributo.
Siempre entre comillas dobles rectas. Admite expresiones regulares para búsquedas más flexibles.

word y lemma: dos formas de buscar en los corpus

word recupera la forma exacta tal como aparece en el texto; lemma recupera todas las variantes flexivas, morfológicas u ortográficas vinculadas a un lema

[word="lánguidos"]

Solo la forma exacta

Encuentra únicamente *lánguidos*.

No recupera *lánguido*, *lánguida* ni *lánguidas*.

Útil cuando la forma gráfica importa (grafías alternativas, mayúsculas, tildes...).

[lemma="lánguido"]

Todas las variantes vinculadas a un lema

Encuentra *lánguido* · *lánguida*

lánguidos · *lánguidas* (languidísimo, etc. según corpus)

→ Ejecuta las dos consultas: ¿cuántos resultados obtienes en cada caso?

Atributos básicos disponibles en SkE

Atributo	Significado
word	la forma tal como aparece en el texto (sensible a mayúsculas)
lc	<i>lowercase</i> : la forma en minúsculas
lemma	el lema (forma (representante) que aparece en los diccionarios)
lemma_lc	el lema en minúsculas
tag	la etiqueta morfosintáctica (a veces el atributo es pos)

Otros atributos pueden estar disponibles según el corpus (por ejemplo, nform, phtags, stags, etags, etc.).

El atributo tag (pos en algunos corpus): etiquetado morfosintáctico

Cuando un corpus está etiquetado, cada token lleva asignada una cadena de caracteres que codifica su categoría gramatical y sus rasgos morfológicos

1

¿Qué es el etiquetado?

Un proceso automático que asigna a cada palabra su categoría gramatical (sustantivo, verbo, adjetivo...) y sus rasgos morfológicos (género, número, tiempo, persona, modo...).

2

¿Qué programa lo hace? Freeling (Neotag en ODE)

Freeling es un analizador de código abierto especializado. Sketch Engine lo aplica al compilar el corpus y seleccionar la opción Spanish, de modo que cada token queda anotado con una etiqueta codificada.

3

¿Para qué sirve el atributo tag?

Permite buscar categorías gramaticales completas, no palabras concretas: todos los verbos en imperfecto, todos los adjetivos calificativos, todos los nombres propios femeninos...

TAGSET

Las etiquetas POS* **varían según los corpus.**

En los corpus de español de Sketch Engine se usa el etiquetador *FreeLing*, cuyas etiquetas empiezan por:

*POS = part of speech

Inicial	Categoría
N	nombre (NC común, NP propio)
V	verbo (VM principal, VA auxiliar, VS ser)
A	adjetivo (AQ calificativo, AO ordinal)
R	adverbio (RG general, RN negativo)
D	determinante (DA artículo, DD demostrativo, DP posesivo)
P	pronombre (PP personal, PR relativo)
S	preposición
C	conjunción (CC coordinante, CS subordinante)
I	interjección
F	puntuación
Z	numerales y fechas

Etiquetas Freeling: los rasgos codificados por posición

Cada posición de la etiqueta codifica un rasgo gramatical. Estos son algunos ejemplos

Etiqueta CQL	Categoría	Rasgos	Ejemplo
NCMS000	Nombre Común	Masculino Singular	<i>corazón, amor, olvido</i>
NCMS00V	Nombre Común	Masc. Sing. eValuativo (dim.)	<i>pisito, cuerpecillo</i>
NPFS000	Nombre Propio	Femenino Singular	<i>Ana, María</i>
AQ0MS00	Adjetivo Qalificativo	Masculino Singular	<i>lánguido, oscuro, pequeño</i>
AQVMS00	Adjetivo Qalificativo	eValuativo (dim.), Masc. Sing.	<i>pequeñito, pobrecillo</i>
DA0MS0	Determinante Artículo	Masculino Singular	<i>el</i>
VMII3S0	Verbo Principal (Main)	Ind. Imperfecto 3.ª Sing.	<i>cantaba, vivía, ardía</i>

En otros corpus, V (evaluativo) se marca como D (diminutivo)

Buscar por categoría gramatical: etiqueta exacta y comodín

Una etiqueta exacta recupera una categoría precisa; el comodín . amplía la búsqueda a toda una clase gramatical*

`[tag="NCMS000"]`

*nombre común
masc. singular (exacto)*

`[tag="AQ0MS00"]`

*adjetivo calificativo
masc. singular (exacto)*

`[tag="VMII3S0"]`

*verbo indicativo imperfecto
3.ª singular (exacto)*

El comodín .* significa «cualquier combinación de caracteres a partir de este punto»

<code>[tag="NCMS000"]</code>	→ solo Nombre Común Masculino Singular
<code>[tag="NC.*"]</code>	→ todos los Nombres Comunes (cualquier género y número)
<code>[tag="VM.I.*"]</code>	→. todos las formas verbales de verbos principales en imperfecto de indicativo y de subjuntivo

Ejercicios básicos: ¿qué recupera cada consulta?

Antes de ejecutar cada consulta en el corpus, anota qué resultados esperas. Luego comprueba.

`[tag="AQ.*"]`

→ ¿Qué recupera?

`[tag="NC.*"]`

→ ¿Qué recupera?

`[tag="VM.*"]`

→ ¿Qué recupera?

`[tag="AQ.*"] [tag="NC.*"]`

→ ¿Qué recupera esta secuencia de dos tokens?

→ *Anota tu respuesta antes de comprobarla en el corpus*

Ejercicios básicos: lo que busca cada consulta

[tag="AQ.*"]

Todos los adjetivos calificativos del corpus (cualquier género y número)

[tag="NC.*"]

Todos los nombres comunes (cualquier género y número)

[tag="VM.*"]

Todos los verbos principales (cualquier tiempo, modo y persona)

[tag="AQ.*"] [tag="NC.*"]

Secuencias de adjetivo calificativo seguido inmediatamente de nombre común

lc / word

¿Qué resultados obtenemos al buscar

lc =
lower case

[lc = "dios"] [lc = "Dios"]

[word = "dios"] [word = "Dios"] [word = DIOS]

información extra: [word = "(?i)dios"] = [lc = "dios"]

a veces: [form = "dios"%c] (ODE)

con lc se recupera la palabra buscada independientemente de que esté escrita con mayúsculas o minúsculas en cualquier posición: "insensible a mayúsculas/minúsculas, no distingue entre mayúsculas y minúsculas": A = a)

lemma / lemma_lc

Qué resultados obtenemos al buscar

[lemma = "Dios"]

[lemma = "dios"]

[lemma_lc = "dios"]

[lemma_lc = "Dios"]

con [lemma_lc = "dios"] [lemma = "Dios"]: busca todas las palabras cuyo lema contenga las letras dios, escritas en may. o min.

Ejemplos

Tarea	Código CQL	Resultado
Localizar la forma <i>fueron</i>	[word="fueron"]	concordancia de <i>fueron</i>
Localizar todas las formas del verbo <i>ir</i>	[lemma="ir"]	<i>voy, vas, fue, iban, ido...</i>
Localizar todas las palabras etiquetadas como nombre propio en algún subcorpus nacional de esTenTen	[tag="NP.*"]	concordancia de varios nombres propios

Buscar varias palabras seguidas

[lemma="dar"] [lemma="el"] [lemma="pésame"]

Los **espacios entre corchetes** son irrelevantes:
están solo para legibilidad.

Pero **dentro de las comillas no debe haber espacios sobrantes**: [lemma=" el"]
no encuentra nada, porque ningún lema empieza por espacio.

actividad extra: probar a ordenar por la primera palabra de la kwic (darle, dio, dar, dieron, etc.)

Atajo: atributo por defecto (default attribute)

Cuando todos los tokens usan el mismo atributo, se puede seleccionar el **default attribute** en el menú desplegable y omitirlo en la consulta. Estos pares son equivalentes:

```
[lemma="dar"] [lemma="e1"] [lemma="pésame"]  
"dar" "e1" "pésame"           ← con default attribute = lemma
```

```
[word="e1"] [tag="A.*"] [tag="N.*"]  
[word="e1"] "A.*" "N.*"       ← con default attribute = tag
```

Empieza por, termina en, contiene: expresiones regulares

Dentro de las comillas se pueden usar expresiones regulares.

Los metacaracteres más usados son:

Metacarácter	Significado		
.	cualquier carácter (uno solo)	?	hace opcional el carácter o grupo anterior
.*	cualquier secuencia, incluida la vacía	[abc]	clase de caracteres: un solo carácter, que puede ser a, b o c
.+	cualquier secuencia de uno o más caracteres	(aver auer)	encuentra aver o auer
		\. \? etc.	escape: el carácter siguiente se interpreta literalmente

Práctica. Uso de & dentro del token. Uso de !=

Tarea	Código CQL
Palabras que empiezan por tl-	[lc = "tl.*"]
adjetivos terminados en -udo en Argentina	[word=".*udo" & tag="A.*"]
palabras terminadas en -udo en Argentina que no sean adjetivos	[word=".*udo" & tag != "A.*"]
Palabras de tres letras que empiezan por <i>p-</i> y terminan en -z	[lemma="p.z"]
Variantes de "covid" en mayúsculas y minúsculas	[lemma_pc = "covid.*"]

Distancia entre tokens y repetición

[] vacío = un token cualquiera.

{ } = repetición del token inmediatamente anterior.

Tarea	Código CQL
<i>dar y consejo</i> con una palabra entre medias	[lemma="dar"] [] [lemma="consejo"]
<i>tener y opinión</i> con 2 a 4 palabras entre medias	[lemma="tener"] []{2,4} [lemma="opinión"]
<i>mi (propia) vida</i> propia es opcional	[word = "mi"] [word = "propia"]? [word = "vida"]

Las llaves admiten: {n} exactamente n; {n,m} entre n y m; {n,} n o más; {0,1} cero o uno (equivalente a ?)

escapar caracteres

¿Qué diferencia hay entre estas dos búsquedas?

[word="\."]

[word="."]

La IA integrada en Sketch Engine: lenguaje natural → CQL

Sketch Engine incorpora un asistente de IA que traduce peticiones en lenguaje natural a expresiones CQL de forma automática

1

Cómo funciona

Se escribe la petición en lenguaje natural directamente en el buscador de Sketch Engine. La IA genera la expresión CQL correspondiente, lista para ejecutar.

2

Para qué sirve

Reduce el tiempo de aprendizaje de CQL y permite construir consultas complejas (expresiones regulares, multitoken, combinaciones de atributos) sin memorizar la sintaxis.

3

Criterio crítico imprescindible

La IA puede cometer errores: etiquetas incorrectas, comodines mal colocados o consultas que no hacen lo que prometen.

IA en Sketch Engine: peticiones simples

Ejemplos de peticiones en lenguaje natural y la expresión CQL que genera automáticamente el sistema

Petición en lenguaje natural	CQL generado automáticamente
encuentra formas = talibán, insensible a mayúsculas en posición inicial	<code>[word=" [Tt] alibán"]</code>
encuentra forma = guion o guión	<code>[word="guion guión"]</code>
encuentra palabras que terminen en -ción	<code>[word=" .*ción"]</code>
encuentra palabras que empiecen por trans-	<code>[word="trans. *"]</code>
encuentra lemas que contengan <i>trans</i> en posición interior	<code>[lemma=" [^tT] .*trans. * [^sS] "]</code>

IA en Sketch Engine: peticiones complejas (no siempre funcionan)

CQL permite encadenar tokens para buscar secuencias de palabras con restricciones gramaticales. El operador & combina condiciones sobre un mismo token

Petición en lenguaje natural	CQL generado automáticamente
Busca tag = verbo indicativo imperfecto	[tag="V.II.*"]
Busca adjetivo calificativo seguido de nombre	[tag="AQ.*"] [tag="N.*"]
verbos en futuro, primera persona del singular	[tag="V..F1S.*"]
el primer token es una forma del verbo haber en tercera persona del plural; el segundo token es un nombre en plural ???	[lemma="haber" & tag="V.*3P.*"] [tag="N.*P.*"]
el primer token es el verbo estar; el segundo token es la forma hasta; el tercer token es un artículo; el cuarto token es un nombre ???	[lemma="estar"] [word="hasta" & tag="S.*"] [tag="DA.*"] [tag="N.*"]

CQL en corpus históricos: ODE, PS y OSTA

- Oralía diacrónica del español (ODE): <http://corpora.ugr.es/ode/>
- Post Scriptum: <http://teitok.clul.ul.pt/postscriptum/>
- Todos los corpus que utilizan TEITOK pueden consultarse en:
- TEITOK projects, Maarten Janssen:
<http://www.teitok.org/index.php?action=projects>
- Old Spanish Textual Archive (OSTA):
<http://osta.oldspanishtextualarchive.org/>

Introducción a CQL en OSTA

Primeras búsquedas

Buscar "putico" en OSTA, en CQL

Buscar "cornudo" en ODE, en CQL

(<http://corpora.ugr.es/ode>)

"putico" aparece en la *Tragicomedia de Calixto y Melibea*.

Pero OSTA incluye muchas ediciones de la Celestina

Buscar solo en las ediciones cuyos codigos HSMS son C01, C02

Ejemplos de obras medievales en OSTA

Código_HSMS	fecha_mas_tardia_testimonio	Título
FRS	1255	Fuero real
GE4	1280	General estoria 4
EE1	1284	Estoria de España, versión primitiva
ULT	1300	Gran conquista de Ultramar
CRD	1327	Cronicón romanceado de Cardeña
ODM	1347	Ordenamiento de Madrid
OA2	1351	Ordenamiento de Alcalá de Henares
PLG	1380	Vida de Santa Pelagia
VIS	1381	Visita y consejo de médicos
BAG	1389	Libro de buen amor
CBO	1466	Corbacho
COL	1493	Carta a Luis de Santángel
ATA	1500	Atalaya de las crónicas

Acotar las búsquedas en varias obras en OSTA

```
[word = "putico"] :: match.text_hsmscod = "C01|C02" within text sort by yearcodice
```

Observa que los metadatos de búsqueda se van sumando con lenguaje CQL y que podemos copiarlos para utilizarlos en búsquedas complejas:

```
[word = ".*miente"] :: match.text_hsmscod = "EE1|TRY" & match.text_lengua1 = "castellano"  
within text sort by yearcodice
```

Por ejemplo, la búsqueda anterior solo encuentra ejemplos en EE1 (Estoria de España) y no en la *Historia Troyana* (de Fernández de Heredia, aragonés).

Búsquedas en OSTA

- Ejemplos de la preposición *pora* 'para'
- Palabras medievales terminadas en -iello, -a, -os, -as en EE1 y GE4
- Ejemplos de *connusco* 'con nos/nosotros', *convusco* 'con vos/vosotros'
- Primeras documentaciones de *alguien*
- Formas verbales terminadas en -ades, -odes, -edes
- Ejemplos de *avién* 'habían' y *avrién* 'habrían'
- Ejemplos de infinitivos con -ee-
- Ejemplos del verbo *fazer* 'hacer' con doble ff- inicial
- Adverbios terminados en -miente o -miente

-iello

RESULTADOS DE LA CONSULTA [DESCARGAR: [TSV](#)]

Consulta => [word = ".*[ijy]ell[oa]s?"] :: match.text_hsmscod = "GE4|EE1" within text sort by yearcodice

Nº total de ejemplos: 924 en 2 textos Frecuencia: 23.72 casos por millón de palabras

Mostrando del 1-500 [Siguientes](#)

Nº	Copiado/impreso	Obra [creación]	Folio	Contexto anterior	Búsqueda
1	1272-1284	Estoria de España, versión primitiva [1270-1274]	2v-a²⁹	la poblaron. E por end nos don alffonso por la gracia de dios. Rey de	C astiella. de T oled C ordoua
2	1272-1284	Estoria de España, versión primitiva [1270-1274]	4r-a⁵¹	Caliz. Tod estas ti-erras sobredichas fueron pobladas assi cuemo uo ^s contamos. e ouo y muchos	cabdiellos que fueron guerras ei
3	1272-1284	Estoria de España, versión primitiva [1270-1274]	5v-a⁴⁸	fasta en duero. e por que auie siete prouincias en su sennorio. fue dicho enlas	fabliellas antiguas c y era giga
4	1272-1284	Estoria de España, versión primitiva [1270-1274]	7r-a¹³	que ouo en Espanna. e tanto la amaua el Rey espan que alli puso su	siella e se coror assi lo fue
5	1272-1284	Estoria de España, versión primitiva [1270-1274]	8r-a²⁸	Rey espan assi cuemo desuso oystes E quando	castiellos muy fuerte

pora

[word = "pora" & pos = "S.*"]

[folio 3r]

- 1 uos dixiemos. E alli la soterro.
- 2 dixo abraam aso sieruo eleazar.
- 3 *que* era mayordomo de toda su ca-sa.
- 4 pon tu mano sobre mi an-ca.
- 5 e jura *con* el sennor dios de lo^s
- 6 cielos. & de la tierra. *que non* prenda^s
- 7 mugier **pora** mio fijo. de fija^s
- 8 de canaan. mas ue amios pa-rientes
- 9 ala tierra do yo naci. &

connusco y convusco

[word = "conn?[uvb]?[uv]sco"]

[word = "conn?[uv]sco | con[bvu][uv]sco"]

alguien

RESULTADOS DE LA CONSULTA [DESCARGAR: [TSV](#)]

Consulta => [word = "alg[uv][iyj]en?"] within text sort by yearcodice

Nº total de ejemplos: **60** en **22** textos Frecuencia: **1.54** casos por millón de palabras

Mostrando del 1-60

Nº	Copiado/impreso	Obra [creación]	Folio	Contexto anterior	Búsqueda	Contexto posterior
1	1330-1330	<i>Fuero general de Navarra, versión C</i> [1330-1330]	47r-b²⁵	auer calonia. Otrosi los clerigos ordenados non deuen labr ^a r pora si nin pora otr ⁱ &	algujen	por la ayno al gunos a[_l]labrar. o Retengar
2	1391-1500	<i>Otas de Roma</i> [1300-1325]	85r-b⁸	uo ^s a escarnir demj / ʒ dixo en su co-rasçon / que com(~)o qu ⁱ er que	alguyen	pesa-se / que ella lo oyd del greton falso e
3	1391-1500	<i>Otas de Roma</i> [1300-1325]	88r-b²⁵	& bien trovada / ferosa qu ⁱ en vos fiso pesar / ca me semeja / que	alguyen	vos meno mal. / se-ñ encu-bierto / yo so

Alguien (búsqueda por lema)

RESULTADOS DE LA CONSULTA [DESCARGAR: TSV]					
Consulta => [lemma = "alguien"] within text sort by yearcodice					
Nº total de ejemplos: 72 en 24 textos Frecuencia: 1.85 casos por millón de palabras					
Mostrando del 1-72					
10	1291-1310	Fuero de Béjar [1284-1295]	126r-a³	mientras estidieren en cor-rai & ie tollier los siruientes ca-tiuos. + ↑ Q vi ouier fijo o siruiet catiuo & ↑ le deuere	algui pennos tollieren + en algun debdo paguelo al pac malfiesto. Si negare
11	1291-1310	Fuero de Béjar [1284-1295]	129r-a⁶	de conceio. D emais qu ^a no coiedores de conçeio ouieren a coger auer de conceio &	algui los de-nostare o los firiere s callonna qu ^a l la fiziere dupl
12	1291-1310	Fuero de Béjar [1284-1295]	136v-a¹⁰	maldi-ga ↑ al iurador a ssu uoluntad fuera que nol diga el nombre uedado que	algui lo foda. Todas otras maldici qu ⁱ siere & el iurador calle fa
13	1291-1310	Fuero de Béjar [1284-1295]	163r-a²	aldea o sse leguen. Q u ^a no el apellido viniere al aldea o non an senna	algui los de fueras al logar de los assi como es dicho.
14	1330-1330	Fuero general de Navarra, versión C [1330-1330]	47r-b²⁵	auer calonia. Otrosi los clerigos ordenados non deuen labr ^a r pora si nin pora otr ⁱ &	algujen por la ayno al gunos dias. d a[.]labrar. o Retengan
15	1391-1500	Otas de Roma [1300-1325]	85r-b⁸	uo ^s a escarnir demj / ¶ dixo en su co-rasçon / que com(~)o qu ⁱ er que	alguyen pesa-se / que ella lo compra del greton falso en

-ades, -edes, odes

```
[word = ".*[aeo]des" & pos = "V.*"]
```

avién, avrién

RESULTADOS DE LA CONSULTA [DESCARGAR: [TSV](#)]

Consulta => [word = "a[uvb]r?[iyj]en?"] within text sort by word

Nº total de ejemplos: **20934** en **213** textos Frecuencia: **537.41** casos por millón de palabras

Mostrando del 1-500 [Siguientes](#)

Nº	Copiado/impreso	Obra [creación]	Folio	Contexto anterior	Búsqueda
1	1301-1400	Leyes del estilo [1310-1310]	192r-b⁹	sus de-mandas sin otro pedimjent ^o ¶ mas Segunt de-recho fue fallado que enla dema[n]da se	abie de fazer testame
2	1326-1350	Estoria de España, versión primitiva [1270-1274]	55r-b³²	mas. & por non auer ende tanta dexauan de comer si non que les non	abⁱe pro ¶ E l & las co
3	1401-1410	Crónica general de España [1387-1387]	67v-a⁸	enel xiiij ^o . Año. que fue enel ayño dela era de .CCclix obo vn pr ^e ste que	abie nombre Religios
4	1455-1455	Lucidario [1293-1293]	51r-a⁷	muchas ademas Ca el nuestro sennor sauje muy vien la obra que fazie & que	a-bie de acab: enten-di
5	1455-1455	Lucidario [1293-1293]	51r-a¹³	de çi-ma del çielo que Son las planetas & las estrellas que a cada vna	abie de dar v mouedo
6	1451-1475	Crónica abreviada [1320-1325]	37v-a²¹	vno con el otro & traye ponpeyo ochen-ta e ocho legiones en cada. vna dellas	abie seys mil cavallo.
7	1451-1500	Libro de los estados [1326-1332]	61v-b²⁵	omne corporal El xxxj ^o capítulo fabla en commo julio dixo al jnfante que ya le	abie dicho la para que
8	1301-1400	Siete partidas [1256-1265]	133v-a¹⁹	feudo (c)[C] ontienda acaesçiendo enel ssenor & enel va-ssallo sobre el fevdo deziendo el señor que	a-bie fecho el dixese q
9	1301-1400	Leyes del estilo [1310-1310]	194v-b³⁶	en casa. del rrey. et non mas Et las costas de venjda. et de tornada.	abien visto del costas d
10	1301-1400	Leyes del estilo [1310-1310]	196v-	se prueua la ferida. tal fe-cho se cuenta por.	abien vista. de

-ee- (infinitivos)

[word = ".+ee.+" & pos = "V.N.*"]

ffazer

RESULTADOS DE LA CONSULTA [DESCARGAR: [TSV](#)]

Consulta => [word = "ff.*" & lemma = "hacer"] within text sort by yearcodice

Nº total de ejemplos: **7780** en **140** textos Frecuencia: **199.72** casos por millón de palabras

Mostrando del 1-500 [Siguietes](#)

Nº	Copiado/impreso	Obra [creación]	Folio	Contexto anterior	Búsqueda	Contexto posterior
1	1211-1235	<i>Fazienda de Ultramar</i> [1205-1205]	44r-b³²	ydolas e pusieron las por los oteros e solos arbores e fisieron les holocaust encensaronlas	ffisi-eron	cosas malas por ydolas e fisieron l
2	1255-1255	<i>Fuero real</i> [1255-1255]	4v-a²⁰	et buenos por natura. a[.]lucifer. & alos otros qu ⁱ despues por su maldad. son	ffechos	dia-blos & malos. incarnation de nu
3	1255-1255	<i>Fuero real</i> [1255-1255]	7v-b⁵	T odo omne que fuer lla-mado ↑ por mandado del Rey que uenga antel o <i>que</i>	ffaga	otra cosa et desp quisiere uenir. o s
4	1255-1255	<i>Fuero real</i> [1255-1255]	10r-b⁴	senten-cias <i>que</i> los prelados pusieren sobre estas cosas sean bien te-nidas fasta quela emienda sea	ffecha	et qu ^a ndo la emie sea luego tollida.
5	1255-1255	<i>Fuero real</i> [1255-1255]	10r-b⁵	estas cosas sean bien te-nidas fasta quela emienda sea ffecha et qu ^a ndo la emienda fuere	ffecha	la sentencia sea l nin xpistiano nin i
6	1255-1255	<i>Fuero real</i> [1255-1255]	10r-b⁵	estas cosas sean bien te-nidas fasta quela emienda sea ffecha et qu ^a ndo la emienda fuere	ffecha	la sentencia sea l nin xpistiano nin i

-miente o -mientre

[word = ".{3,}m[iyj]entr?e"%c] :: match.text_hsmscod = "FRS|GE4|EE1|ULT" within text sort by yearcodice

en una sola palabra: solamientre

[word = ".{3,}m[iyj]entr?e"%c]

en dos palabras: sola mientre

[pos = "A.*"] [word = "m[yij]entr?e"]

Introducción a CQL en ODE

De cuántas formas diferentes se ha escrito históricamente la forma verbal *recibió*. Hacer la búsqueda en ODE utilizando el atributo *nform*

[nform = "recibió"]

Con *nform* buscamos la forma modernizada (normalizada) y le pedimos al corpus que nos dé todas las variantes ortográficas que se vinculan con esa forma actual. Así obtenemos toda la variación ortográfica.

nform es la búsqueda por defecto en ODE.

construir una búsqueda con *form* que recupere todas las variantes ortográficas de *recibió* (sin usar *nform*)
[form = "rr?es?(ç|z|c|s)i[vbu]i[óo]"%cd]

Intentar obtener todas las variantes ortográficas de *recibió* en el CDH.

Search Query	[nform="recibió"]
Group query	Expanded form
Total	419
Reference size	1801208

Graph:

Table	▼
-------	---

 | Count:

Cou

Group	Count	WPM	Percent
reçibio	93	51.63	22.2
reziuiio	71	39.42	16.95
recibio	50	27.76	11.93
recivio	38	21.1	9.07
reciuio	35	19.43	8.35
rezibio	30	16.66	7.16
reçiuio	14	7.77	3.34
rreciuio	10	5.55	2.39
rreziuiio	8	4.44	1.91
rreçibio	8	4.44	1.91
...

Resultados de

```
[form = "tr.(x|j)o"%c]
```

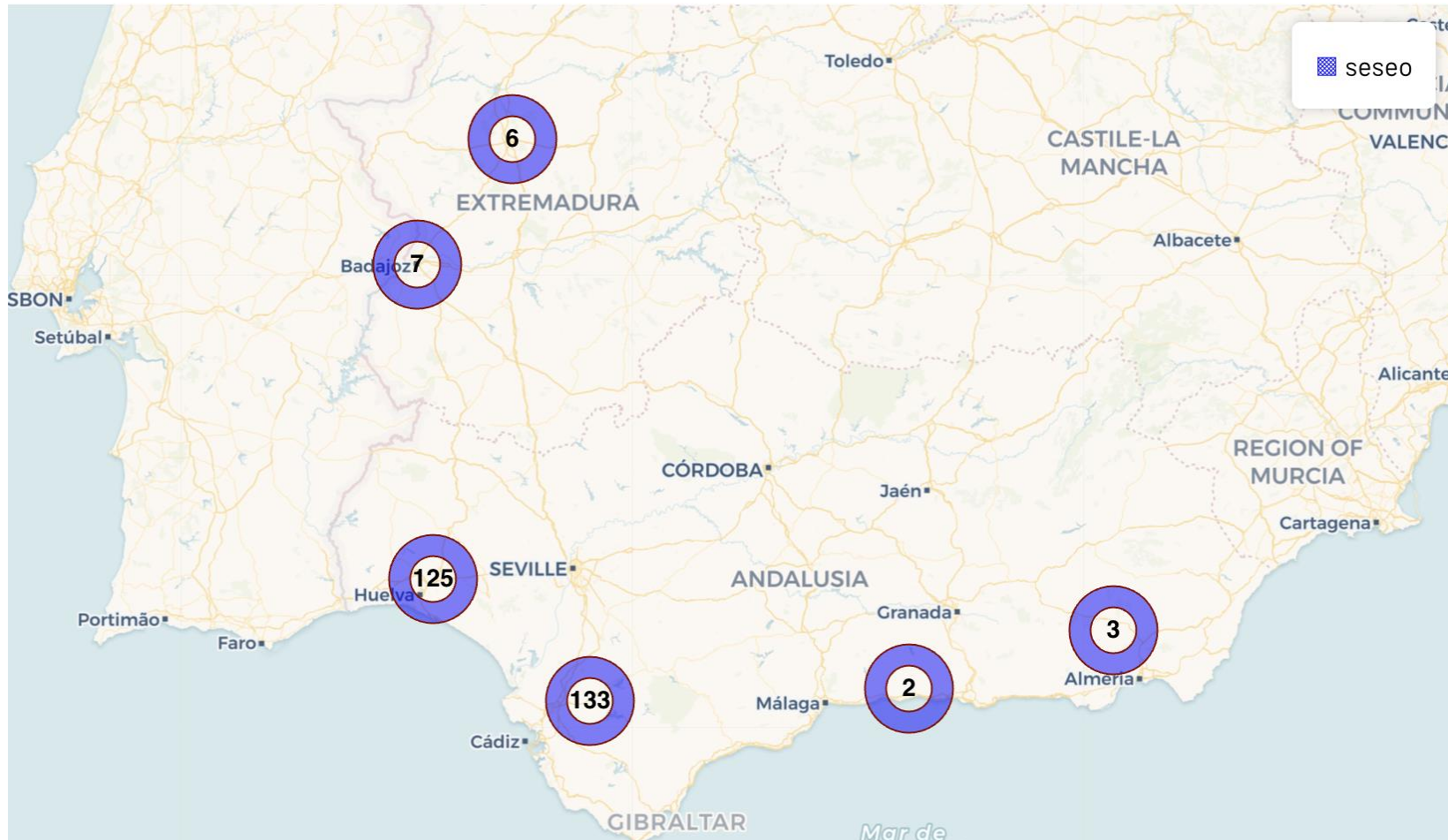
Con esta búsqueda se obtienen todas las formas de pretérito del verbo *traer*
Escribe otro patrón para hallar el mismo resultado

```
[lemma = "traer" & pos = "VMIS3S0" ]
```

Hacer lo mismo en OSTA

Hacer mapas básicos en ODE

ejemplos de localidades seseantes, a partir de [form = "sinco"]



Construir etiquetas morfosintácticas

Construir de tres maneras diferentes la etiqueta de las siguientes palabras:
había, taburetico, así

había [pos = "VAII3S0"]

taburetico [pos = "NCMS00D"]

así [pos = "R.*"]

Buscar los 10 adjetivos (lemas) más frecuentes en PS y ODE

Search Query	POS Tag = AQ.*
Group query	Lemma
Total	44237
Reference size	971206

Graph: | Count: | Download

ode

Group	Count	WPM	Percent
medio	3,125	3,217.65	7.06
nuevo	2,120	2,182.85	4.79
pequeño	2,080	2,141.67	4.7
viejo	1,949	2,006.78	4.41
grande	1,927	1,984.13	4.36
blanco	1,754	1,806	3.97
negro	1,589	1,636.11	3.59
mismo	1,434	1,476.51	3.24
santo	1,160	1,194.39	2.62
mayor	879	905.06	1.99

Ampliación 1. Corpus novelas

El corpus *novelas* es un corpus privado de novelas en español (unos 17M de palabras), con representación de casi todos los países de habla hispana, alojado en SkE. Para solicitar acceso, debes dirigirte a calderon@ugr.es, indicando la dirección de correo con la que te has dado de alta en SkE

Corpus novelas

Información y gestión del corpus

Crear subcorpus nacionales, si fuera necesario

Argentinismos en el corpus de novelas

Usar esTenTen como corpus de referencia-control; o bien el resto del corpus

Probar con disntintos valores de enfocarse en...

sabés	acordás	vení
nene	guita	bombacha
gaucho	petiso	pija
querés	entendés	cana, etc.
vos	tenés	
boludo	mucama	

Hacer lo mismo con españolismos: vivales, fulano, etc.

En qué país del corpus *novelas* se usa más el pret. perf. (has cantado)

Comparar España, Argentina, Venezuela, Cuba y Bolivia

```
[tag = "VAIP.*"][tag = "VMP.*"]
```

¿Diminutivos en -ito en novelas españolas / mexicanas?

[word = ".+it[oa]s?" & !lemma = ".+it[oa]s?" & tag = "N.*|A.*|R.*"]

[lc=".*it[ao]s?" & lemma_lc!=".*it[ao]s?" & tag="[ANR].*"]

Spain ▼ ✕

CQL [word = ".+it[oa]s?" & !lemma = ".+it[oa]s?" & tag = "N.... • 2324

479,84 por millón tokens • 0.012%



Mexico ▼ ✕

CQL [word = ".+it[oa]s?" & !lemma = ".+it[oa]s?" & tag = "N.... • 1409

770,35 por millón tokens • 0.0072%



⌂ - ...

Diminutivos en -ito en CORPES XXI (adjetivos, España)

Elementos gramaticales ▼

Lema

!*ito!*ita!*itos!*itas



Forma

*ito|*ita|*itos|*itas



Categoría gramatical

A-----



- adjetivo

Sensibilidad

☐

Acentos

☐

Mayúsculas

Filtros

España ▼

☐

América

☒

España

España ▼

Hasta

Bloque ▼

España ▼

Soporte ▼



Resultado diapositiva anterior

Descargar en TSV

Descargar en documento

Descargar en texto

Imprimir

Año ascendente

Sin criterio

Documentos: 3307 / Frecuencia absoluta: 8813 / Frecuencia normalizada: 55,28 por millón

<input type="checkbox"/>		Doc.	País	Año	Contexto izquierdo	Término	Contexto derecho
<input type="checkbox"/>		LE2001_0001	Esp.	2001	... yo tan claro contigo. Eres un poco	blandito	, niño.
<input type="checkbox"/>		LE2001_0001	Esp.	2001	SURCOS.—	Perfumadito	como un novio...
<input type="checkbox"/>		LE2001_0001	Esp.	2001	SURCOS.— No, no aprendes. Eres	cortito	, niñato. No tienes surcos en la cabe...
<input type="checkbox"/>		LE2001_0001	Esp.	2001	... Verás, «ella» se puede poner así,	sentadita	en tus piernas... Y tú te agarras fuer...
<input type="checkbox"/>		LE2001_0001	Esp.	2001	SURCOS.— ¿Eres	tontito	o es que te gusta vacilar?
<input type="checkbox"/>		LE2001_0001	Esp.	2001	...o, saca a tu amigo, que tú eres más	rarito	también... Y no me mires así.
<input type="checkbox"/>		LE2001_0001	Esp.	2001	...ito, tú tranquilo, ¿eh?, que yo soy	jovencita	pero muy profesional.
<input type="checkbox"/>		LE2001_0001	Esp.	2001	...A.— Si lo he cogido del baño. Está	limpita	. (Acercándose a CACHORRO.) No se...

Resultado para México en CORPES XXI (adj., dim. -ito)

Documentos: 1462 / Frecuencia absoluta: 4826 / Frecuencia normalizada: 94,16 por millón						
<input type="checkbox"/>		Doc.	País	Año	Contexto izquierdo	Contexto derecho
<input type="checkbox"/>	ⓘ	LA2001_0002	Méx.	2001	... no le creo". "No tengas miedo,	cabroncito , ven a mi casa aquí cerquita". "N...
<input type="checkbox"/>	ⓘ	LA2001_0002	Méx.	2001	...lí para mirar el mundo, te haces	chiquito , chiquito, como una hormiga.
<input type="checkbox"/>	ⓘ	LA2001_0002	Méx.	2001	...rar el mundo, te haces chiquito,	chiquito , como una hormiga.
<input type="checkbox"/>	ⓘ	LA2001_0002	Méx.	2001	...la cola y hasta se los come. Salió	igualito que su padre.
<input type="checkbox"/>	ⓘ	LA2001_0002	Méx.	2001	...n animal pequeño. Así como yo,	negrito .
<input type="checkbox"/>	ⓘ	LA2001_0002	Méx.	2001	-Primero le damos una	calentadita . Luego le ponemos lumbré en los ...
<input type="checkbox"/>	ⓘ	LA2001_0002	Méx.	2001	-Chócala,	compadrito -El Burro y el guardia se saludan d...
<input type="checkbox"/>	ⓘ	LA2001_0046	Méx.	2001	...isparada una araña con sus crías	pequeñitas .

hubieron accidentes (esTenTen)

CONCORDANCIA		Spanish Web 2023 (esTenTen23)		Obtén más al	
Argentinian domain .ar		CQL [lemma="haber" & tag="V...3P.*"] [word="habido"]? [ta... • 6817 2,63 por millón tokens • 0.000021%			
Detalles		Contexto izquierdo	KWIC	Contexto derecho	
1	tn.com.ar	lista mundial intuyó la infidelidad de Gerard Piqué con Clara Chía porque	habían productos VMII3P0 NCMP000	que mágicamente despar	
2	pts.org.ar	ad a los mercados". "No sea cosa que cuando terminemos de discutir, ya	hayan desaparecidos VMSP3P0 NCMP000	los fondos", defendió gráf	
3	pts.org.ar	ad Sun Yat-sen eran miembros o simpatizantes de la Oposición. También	habían grupos VMII3P0 NCMP000	en las escuelas militares y	
4	aac.org.ar	temario haciendo especial hincapié en la Oncología. Cabe destacar que	hubieron actividades VMIS3P0 NCFP000	que acapararon, extraordi	
5	ayp.org.ar	las ciencias no es sólo un problema de carácter teórico o metodológico.	Habrían razones VMIC3P0 NCFP000	de tipo ideológico que det	
6	sac.org.ar	e 24 (42%) del grupo control (HR 0,11 [95% CI, 0,01-0,85]; P=0,034). No	hubieron muertes VMIS3P0 NCFP000	en el grupo ablación y se	
7	afmb.com.ar	iniciativa tiene cupos limitados. Se le dará prioridad a los equipos que se	hayan inscriptos VMSP3P0 NCMP000	hasta el cierre de la capac	
8	fcdn.org.ar	rencias, controversias, mesas redondas, diálogos con expertos, talleres.	Hubieron actividades VMIS3P0 NCFP000	conjuntas societarias con	
9	diario22.ar	ores les dará salida, y en esa cuerda floja se encuentra el propio Icardi. "	Habrán cosas VMIS3P0 NCMP000	negociables y otras no ne	

Ampliación 2. Corpus de canciones

Generar corpus de canciones (Bad Bunny)

Compilar corpus con otro cantante

Material complementario

Material necesario para este bloque:
archivos incluidos en la carpeta: bad_bunny

PALABRAS CLAVE

Bad_Bunny

🔍

i

- PALABRAS ÚNICAS ✓
- EXPRESIONES MULTIPALABRAS ✓

↔️

corpus de referencia: Spanish Web 2018 (esTenTen18)

(items: 793)

Lema	Lema	Lema	Lema	Lema
1 ey ...	11 hijueputa ...	21 yonaguni ...	31 ozuna ...	41 sorry ...
2 pa'l ...	12 tirarno ...	22 cheese ...	32 dime ...	42 selfie ...
3 perreo ...	13 to'a ...	23 jeva ...	33 say ...	43 hazle ...
4 tití ...	14 botar ...	24 yeh ...	34 ojito ...	44 yeah ...
5 bellaquear ...	15 oh-oh ...	25 jodio ...	35 acariciándolo ...	45 erar ...
6 roncar ...	16 estamo ...	26 nio ...	36 vo'a ...	46 diabla ...
7 vamo ...	17 baby ...	27 déjame ...	37 rumbiar ...	47 casper ...
8 mami ...	18 preguntó-tó-tó-tó-tó-tó-tó ...	28 pa ...	38 totito ...	48 dale ...
9 bellaco ...	19 me ...	29 somo ...	39 chamaco ...	49 vip ...
10 cabrón ...	20 sunblock ...	30 carajo ...	40 darell ...	50 cierro ...

perreo

1. **m.** Baile que se ejecuta generalmente a ritmo de reguetón, con eróticos movimientos de caderas, y en el que, cuando se baila por parejas, el hombre se coloca habitualmente detrás de la mujer con los cuerpos muy juntos.

Estos cabrones me **roncan** de palo Y nosotros, dueños de la ferretería

- me roncan = alardean
- ferretería = todas las armas (de hierro)

Test final

1. Obtener la frecuencia por millón de *jimagua, jimaguas* en el subcorpus cubano de esTenTen23
2. Calcular la frecuencia por millón de la palabra X, que tiene una frecuencia absoluta de 250 en un corpus de 1 500 000 palabras
3. Obtener la frecuencia de *el pus / la pus* en esTenTen23
4. Clasificar la frecuencia normalizada de *vosotros, vosotras* por países, en CORPES XXI
5. Gráfico del word sketch de *obsoleto*
6. Evolución cronológica de *violencia vicaria* en Spanish Trends
7. Obtener el perfil combinatorio de *Trump* y *Miley*
8. Sinónimos de *recalcitrante* y *acérrimo*
9. Nube de sinónimos de *exuberante*
10. cql para encontrar los verbos que aparecen en 1. se/me/te 2. ¿verbo? 3. el/un 4. marrón

1. jimaguas: frec / millón, Cuba

CONCORDANCIA

Spanish Web 2023 (esTenTen23)



Obtén más i

Cuban domain .cu ▾ ✕

forma **jimaguas?** • 282

0,64 por millón tokens • 8.5e-7%



Detalles

Contexto izquierdo

KWIC

Contexto derecho

1	<input type="checkbox"/>	umcc.cu	portal de los vecinos de enfrente. Bajo una tormenta eléctrica, Daysi, la madre de los jimaguas , aborda con uno de sus pequeños el
2	<input type="checkbox"/>	tribuna.cu	rtas ensaladas quieren?" . Mi mente divaga... ¿y si no quiero ensaladas, podré tomar Jimaguas o Tres Gracias? ¿Qué pasó con los s
3	<input type="checkbox"/>	cubadebate.cu	obertura del Triton y el Neptuno, en un verdadero crimen , la inutilizacion del nuestros jimaguas de Miramar. Tremendo terreno, de los
4	<input type="checkbox"/>	cubadebate.cu	0, 2, 2... pues dos hermanos pueden tener la misma edad (en años enteros) y no ser jimaguas , es decir uno ser algunos meses mei
5	<input type="checkbox"/>	cubadebate.cu	ltiplos de 40 que sumados den 14 y son: 2, 2 y 10 y 1, 8 y 5. Como Basilio Jr. no es " jimagua ", las edades de los hijos de Benjamín
6	<input type="checkbox"/>	gob.cu	oca del Negro, de la medida de pie que nos dio el Galleguito. Estábamos el Negro, los jimaguas y yo, pescando clarias en las zanjas c

2. Calcular frec por millón

$$(250/1\ 500\ 000) * 1\ 000\ 000 = 166,66$$

3. el pus | la pus

frase **el pusla pus** • **4186**

0,13 por millón tokens • 0.000013%



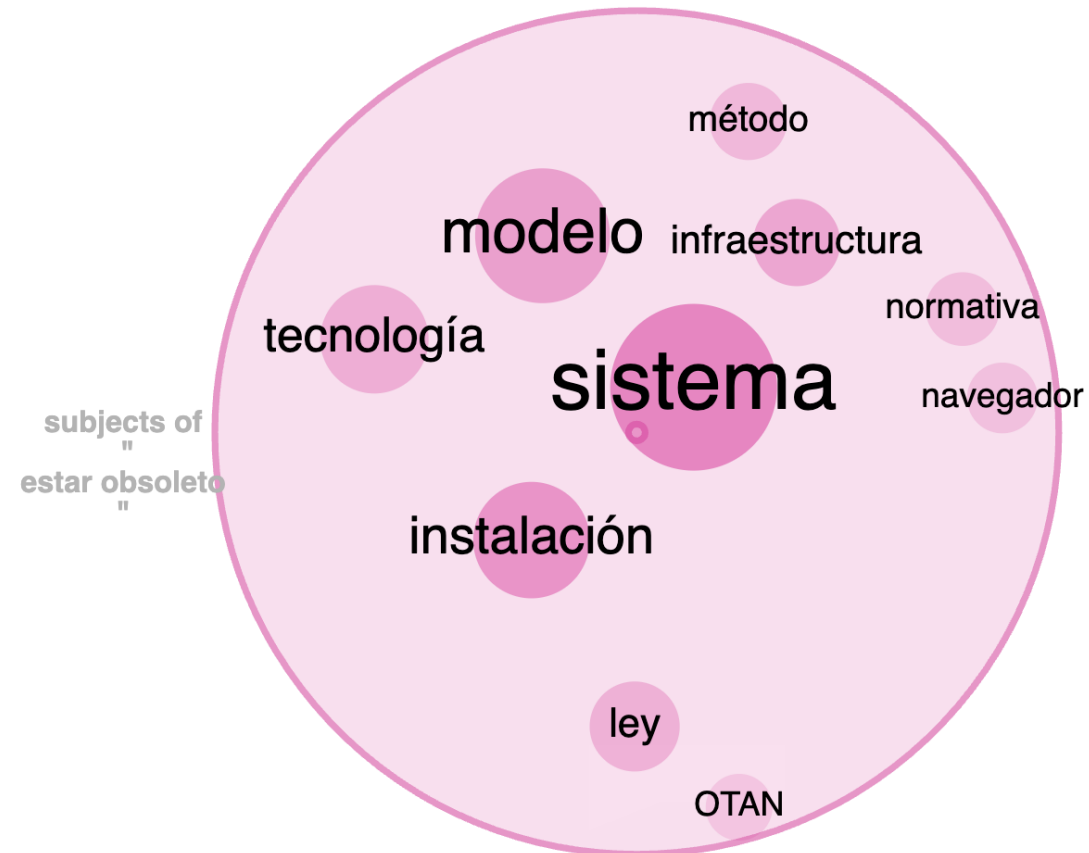
<input type="checkbox"/>	el pus	2722
<input type="checkbox"/>	la pus	1033

4. lema = vosotros, CORPES XXI

	FREC ABS	FREC NORM
España	6000	37,96
Filipinas	5	35,74
Honduras	32	5,76
Uruguay	44	4,37
Cuba	52	3,67
República Dominicana	29	3,66
Venezuela	60	3,28
Paraguay	24	2,85
Argentina	110	2,79

5. Word Sketch de obsoleto

obsoleto



6. Evolución de *violencia vicaria* en Timestamped y Spanish Trends

CONCORDANCIA

Timestamped JSI web corpus 2014-2021 Spanish

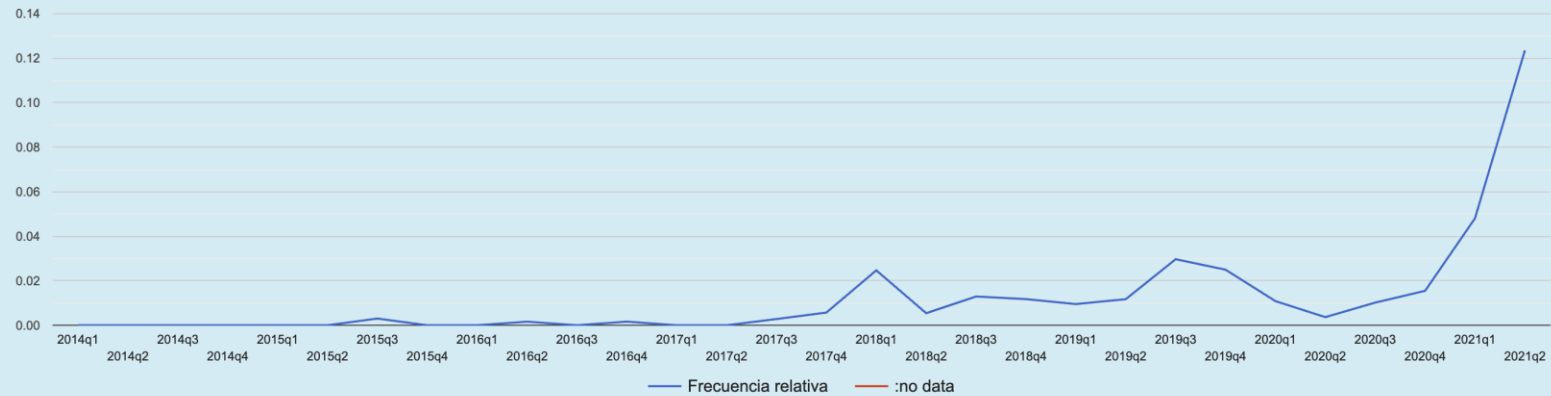
Obtén más almacenamiento

frase **violencia vicaria** • 242
0,01 por millón tokens • 0.0000013%

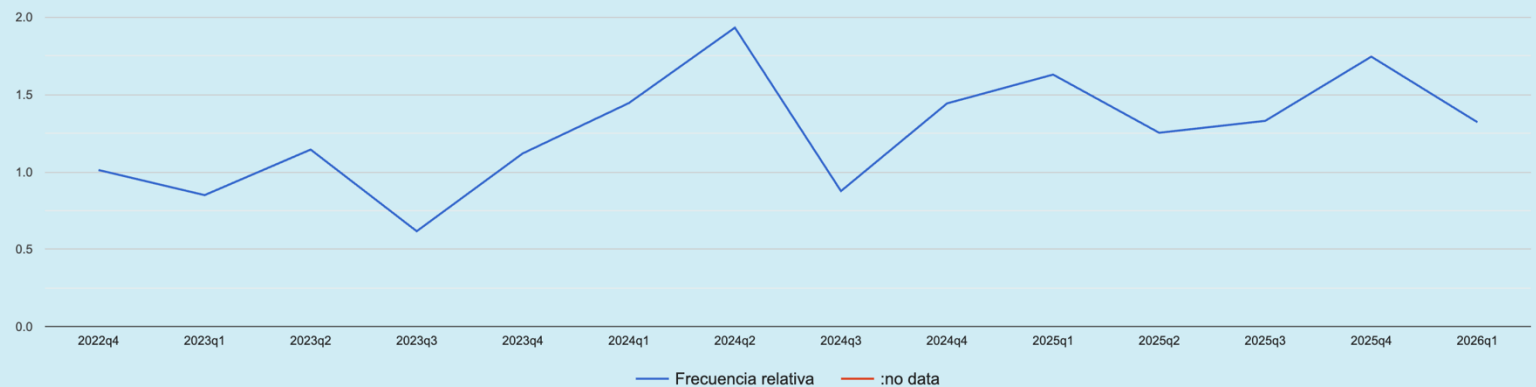
🔍 ⬇️ ☰ 👁️ 🗑️ ✂️ 📏 📄 ⌨️ 🔗 📈 KWIC + ⓘ ☆

TIMELINE

The points represent the [relative frequency](#) (=frequency per million tokens) of [KWIC](#) in each period. Absolute (raw) frequency is rarely a useful value, especially with periods of uneven sizes. Concordance lines from texts without timestamps are not included in the chart.



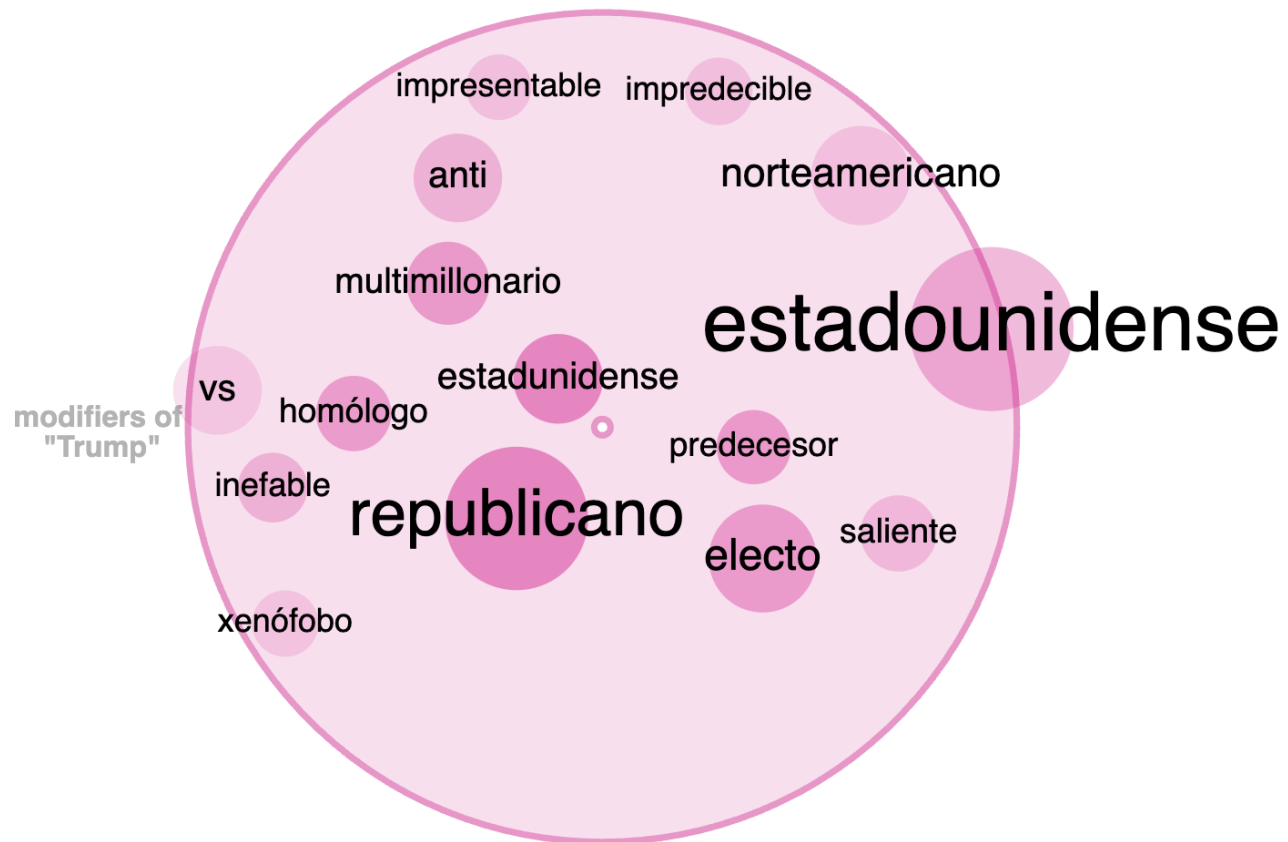
2014-2021 (Timestamped)



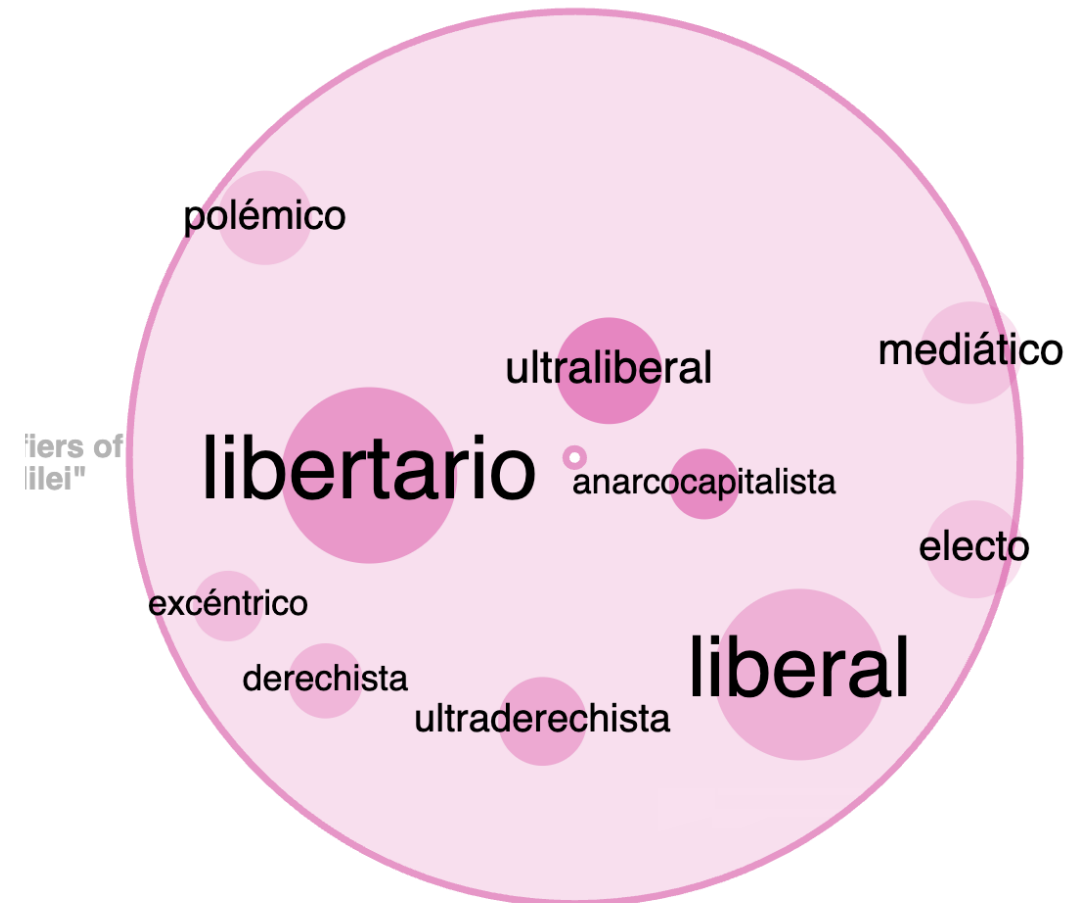
2022-2026 (Spanish Trends)

7. Milei; Trump

Trump



Milei



8. Sinónimos (Tesauro) de *recalcitrante* y *acérrimo*

recalcitrante

	Lempos	Frecuencia ?
1	irredento	1027 ...
2	acérrimo	7070 ...
3	furibundo	2303 ...
4	ferviente	5988 ...
5	intransigente	2958 ...
6	rancio	5440 ...
7	rampante	2882 ...
8	supremacista	2055 ...
9	extremista	13.748 ...
10	confeso	6066 ...

acérrimo

	Lempos	Frecuencia ?
1	ferviente	5988 ...
2	recalcitrante	1432 ...
3	furibundo	2303 ...
4	irredento	1027 ...
5	confeso	6066 ...
6	fervoroso	1943 ...
7	incondicional	30.790 ...
8	purista	1905 ...
9	férreo	23.535 ...
10	conspicuo	826 ...

9. Nube de sinónimos de *exuberante*

exuberante como adjetivo 9024x

...

GRÁFICO DE BURBUJAS

NUBE DE PALABRAS



9. CQL complejo

[word = "[mts]e"][tag = "VM.*"][word = "un|el"][word = "marrón"]

me comiera el marrón

me comiera el marrón

me pasó el marrón

se comerán el marrón

se quita un marrón

se coman el marrón

se comió un marrón

se antojaba el marrón

se destapó el marrón

se comió el marrón

- | | | |
|----|--------------------------|---------------------------|
| 13 | <input type="checkbox"/> | te quite el marrón |
| 14 | <input type="checkbox"/> | se comieron el marrón |
| 15 | <input type="checkbox"/> | te recomendamos el marrón |
| 16 | <input type="checkbox"/> | me comiera el marrón |
| 17 | <input type="checkbox"/> | se quita un marrón |
| 18 | <input type="checkbox"/> | me dejes el marrón |
| 19 | <input type="checkbox"/> | me endosa un marrón |

Gracias