

# VI Taller en competencias específicas Jornadas Doctorales en Lenguas y Culturas

Herramientas digitales para la investigación basada en  
corpus

Aroa Orrequia Barea



# TABLA DE CONTENIDOS

- ❑ ¿Qué es un corpus?
- ❑ ¿Por qué utilizar un corpus en nuestras investigaciones?
- ❑ ¿Para qué se utilizan?
- ❑ Tipos de corpus
- ❑ Conceptos básicos
- ❑ Trabajando con corpus
  - ❑ Corpus online
  - ❑ Herramientas digitales
    - ❑ AntConc
    - ❑ Sketch Engine
  - ❑ Algunos ejemplos

**¿Qué es un corpus?**

**¿Por qué utilizar  
corpus en nuestras  
investigaciones?**

# ¿Para qué se utilizan los corpus?

- Cualquier estudio del lenguaje de manera empírica
- Procesamiento del lenguaje natural
- Estudios sincrónicos
- Adquisición de lenguas
- Enseñanza de lenguas
- Traducción
- Sociolingüística
- Psicolingüística
- Lingüística computacional
- Inteligencia artificial

# Tipos de corpus

Generales o de referencia  
vs. especiales,  
especializados o con fines  
específicos

Monolingüe vs. multilingüe

Paralelo vs. comparable

Sincrónico vs. diacrónico

Dinámico vs. estático

Hablado vs. escrito

Hablado vs. escrito

De aprendices o  
aprendientes

Texto plano vs. corpus  
anotado

# Conceptos básicos

- ❑ **Concordancia** → la lista de todos los ejemplos en los que aparece la palabra que hemos buscado.
- ❑ **Lema** → forma canónica de la palabra, e.g., el infinitivo del verbo
- ❑ **Palabras claves** → lista de palabras ordenadas por frecuencia
- ❑ **Palabras tokens** → número de palabras en general
- ❑ **Palabras tipo** → número de palabras que incluye todas las apariciones de una misma palabra en el texto

*Caminante no hay camino, se hace el camino al andar*

Camino = dos palabras tokens, una palabra tipo

## Conceptos básicos (II)

- ❑ **Frecuencia** → número de veces que aparece una palabra en un texto
- ❑ Tipos de frecuencia:
  - ❑ **Absoluta** = total
  - ❑ **Relativa** = número de palabras tipos / número de tokens x 100 (porcentaje)
  - ❑ **Normalizada** = frecuencia por millones de palabras
- ❑ **Colocaciones** → cálculo estadístico de las palabras que más frecuentemente acompañan a la palabra que se ha buscado
- ❑ **Palabras claves** (*keywords*) → lista de palabras que son normalmente frecuentes en el corpus o texto que se está investigando, en comparación con un corpus de referencia
  - ❑ **KWIC** (*key word in context*) → forma de visualización en la que aparece la palabra buscada en una columna central y a los lados la misma cantidad de

# Trabajando con corpus

## Corpus online

- ❑ British National Corpus (BNC)
- ❑ Corpus de Referencia del Español Actual (CREA)

# BNC

- ❑ 100 millones de palabras
- ❑ General o de referencia: diferentes estilos y variedades
- ❑ Monolingüe: inglés británico moderno
- ❑ 90% escrito y 10% oral
- ❑ Sincrónico: finales del siglo XX
- ❑ <http://www.natcorp.ox.ac.uk/>

<b>SPOKEN</b>	<b>9,963,663</b>
S_brdcst_disc	736,229
S_brdcst_doc	40,554
S_brdcst_news	254,370
S_classroom	412,372
S_consult	131,354
S_conv	4,012,457
S_courtroom	125,438
S_demonstratn	30,500
S_interv_oral	798,978
S_interview	119,117
S_lect_arts	49,759
S_lect_com	14,757
S_lect_law	49,774
S_lect_natsci	22,168
S_lect_socsci	154,718
S_meeting	1,334,382
S_parliament	95,025
S_pub_debate	278,458
S_sermon	80,135
S_spch+script	196,615
S_spch-script	448,810
S_sportslive	32,103
S_tutorial	138,888
S_unclass	406,702

<b>FICTION</b>	<b>15,909,312</b>
W_fict_drama	44,975
W_fict_poetry	219,409
W_fict_prose	15,644,928
<b>MAGAZINE</b>	<b>7,261,990</b>
W_pop_lore	7,261,990
<b>NEWSPAPER</b>	<b>10,466,422</b>
W_new_arts1	345,860
W_news_arts2	235,525
W_news_com	416,345
W_news_edit	100,659
W_news_misc	1,019,839
W_news_o_com	407,277
W_news_o_rep	2,681,576
W_news_o_sci	54,327
W_news_o_soc	1,125,324
W_news_o_sprt	1,009,878
W_news_rprt	655,508
W_news_sci	64,634
W_news_script	1,262,351
W_news_soc	80,963
W_news_sprt	292,832
W_news_tabld	713,524

<b>NON-ACADEMIC</b>	<b>16,495,185</b>
W_nonac_arts	3,722,655
W_nonac_engin	1,186,625
W_nonac_law	4,450,696
W_nonac_med	495,734
W_nonac_nat	2,491,219
W_nonac_soc	4,148,256
<b>ACADEMIC</b>	<b>15,331,668</b>
W_ac_engin	678,621
W_ac_hum_arts	3,296,072
W_ac_law_edu	4,615,173
W_ac_medicine	1,412,808
W_ac_nat_sci	1,104,527
W_ac_soc_sci	4,224,467
<b>MISCELLANEOUS</b>	<b>20,835,159</b>
W_admin	218,595
W_advert	549,856
W_biography	3,494,374
W_commerce	3,729,662
W_email	209,815
W_essay_schl	145,041
W_essay_univ	55,477
W_hansard	1,149,732
W_inst_doc	542,553
W_instruction	433,932
W_let_pers	51,840
W_let_prof	65,511
W_misc	9,074,079
W_religion	1,114,692
<b>TOTAL</b>	<b>96,263,399</b>

# **B** RITISH **N** ATIONAL **C** ORPUS

## About

What is the BNC?  
Creating the BNC  
BNC Products  
Copyright  
Contact Us  
Contents A-Z

## Using the BNC

What can I do with  
the BNC?  
FAQ

## Obtaining

How to download

## Search the British National Corpus online

Various online services offer the possibility to search and explore the BNC via different interfaces. Some of the most notable are listed below:

- [BNCWeb at Lancaster University](#) (registration required - [sign up here](#))
- [BYU-BNC \(Brigham Young University\)](#)
- [Intellitext \(University of Leeds\)](#)
- [Phrases in English](#)
- [Audio BNC](#) - access the digital audio files from the spoken corpus

BÚSQUEDAS LIMITADAS

REQUIERE REGISTRARSE  
(GRATIS)

Please note that we cannot answer queries about using any of these services, which are provided by other institutions. If you have a service for querying the BNC online, get in touch and we'll consider adding it to the list.

## About the BNC

The British National Corpus (BNC) is a 100 million word collection of samples of written and spoken language from a wide range of sources, designed to represent a wide cross-section of British English, both spoken and written, from the late twentieth century. [[more](#)]

Here are some of the most popular links to information about the BNC:

- [Download the full BNC \(XML edition\) from the Oxford Text Archive](#)
- [Download the BNC Baby \(4m word sample\) from the Oxford Text Archive](#)
- [Reference Guide for the BNC \(XML edition\)](#)

## Standard Query

Query mode:  ▾  
Number of hits per page:  ▾  
Restriction:  ▾  
Extended audio controls:  ▾

[Simple Query Syntax](#) | [Audio data search](#)

### Main menu

#### Query options

[Standard query](#)

[Written restrictions](#)

[Spoken restrictions](#)

#### User-specific functions

[User settings](#)

[Query history](#)

[Saved queries](#)

[Categorized queries](#)

[Make/edit subcorpora](#)

[Upload external data file](#)

#### Additional functions

[Browse a text](#)

[Scan keywords/titles](#)

[Explore genre labels](#)

[Frequency lists](#)

[Keywords](#)

#### About BNCweb

[BNCweb book](#)

[The BNCweb team](#)

[New features](#)

[Bug reports](#)

[The CLAWS-5 tagset](#)

[Oxford BNC homepage](#)

Your query "apple" returned 2606 hits in 789 different texts (98,313,429 words [4,048 texts]; frequency: 26.51 instances per million words)

< << >> >| Show Page: 1 Show Sentence View Show in random order Show extended audio data controls New Query Go!

No	Filename	Hits 1 to 50	Page 1 / 53
1	<a href="#">A06 1192</a>	; you need to convey the spurts of energy vocally. The	<a href="#">Apple</a> Cart by George Bernard Shaw The play is a fable, of
2	<a href="#">A0C 1195</a>	National in Portsmouth won the top dessert prize for her hot Banbury	<a href="#">apple</a> pie. All the winning dishes are to be featured in a
3	<a href="#">A0C 1222</a>	basil butter; roast duckling breast with clover honey and a saffron	<a href="#">apple</a> compote and iced Drambuie parfait with sour cherries. Norbert Stening and
4	<a href="#">A0G 47</a>	picks out plants that positively enjoy feet of clay. Specialist spotlight	<a href="#">Apple</a> appeal Graham Rice takes the graft out of choosing apple rootstocks and
5	<a href="#">A0G 47</a>	Specialist spotlight Apple appeal Graham Rice takes the graft out of choosing	<a href="#">apple</a> rootstocks and varieties as he visits a specialist fruit tree nursery.
6	<a href="#">A0G 99</a>	shrink from a heavy soil — Nigel Colborn on clay, p26	<a href="#">Apple</a> growing made as easy as pie Anne Swithinbank's brilliant bulbs for
7	<a href="#">A0G 132</a>	were brought back from the Horn of Africa for Queen Hatshepsut.	<a href="#">Apple</a> pie order Learn everything you ever wanted to know about apples while
8	<a href="#">A0G 133</a>	to know about apples while tasting cider and trying out some scrumptious	<a href="#">apple</a> recipes at Priorwood Garden, Roxburghshire on September 29. The Apple
9	<a href="#">A0G 134</a>	apple recipes at Priorwood Garden, Roxburghshire on September 29. The	<a href="#">Apple</a> Day runs from 2–6pm and includes an apple question time session,
10	<a href="#">A0G 134</a>	September 29. The Apple Day runs from 2–6pm and includes an	<a href="#">apple</a> question time session, with experts on hand to identify your mystery
11	<a href="#">A0G 647</a>	store is frost-free. • Place a greaseband round the trunk of	<a href="#">apple</a> trees to catch the wingless female winter moth as it crawls up
12	<a href="#">A0G 1088</a>	tree Hedgehogs are the most welcome of garden pest controllers Specialist Spotlight	<a href="#">Apple</a> appeal Growing apples is as easy as apple pie, if you
13	<a href="#">A0G 1089</a>	pest controllers Specialist Spotlight Apple appeal Growing apples is as easy as	<a href="#">apple</a> pie, if you are Joan Greenway of Highfield Nurseries. Graham
14	<a href="#">A0G 1091</a>	the fruit specialists, she described to me her recipe for trouble-free	<a href="#">apple</a> growing: disease-resistant varieties, the right rootstock for your preferred size
15	<a href="#">A0G 1091</a>	of tree, a simple summer and winter prune plus a crab	<a href="#">apple</a> to take care of your pollination needs. She made it all
16	<a href="#">A0G 1093</a>	she and other growers use have certainly taken the anxiety out of	<a href="#">apple</a> growing, a subject wreathed in mystery, according to traditional literature
17	<a href="#">A0G 1094</a>	in mystery, according to traditional literature. Just how small can	<a href="#">apple</a> trees get? Small enough to step over, as Highfield's
18	<a href="#">A0G 1111</a>	are really too big for most gardeners to manage. Choosing rootstocks	<a href="#">Apple</a> trees are made up of two parts, the fruiting variety and
19	<a href="#">A0G 1115</a>	' explains Joan. 'If you simply take cuttings from an	<a href="#">apple</a> tree they will grow vigorously but take years and years to fruit
20	<a href="#">A0G 1147</a>	that group will pollinate most other varieties. The small group of	<a href="#">apple</a> varieties know as triploids present more of a challenge. Although triploids
21	<a href="#">A0G 1152</a>	may not be lost. As Joan points out, bees love	<a href="#">apple</a> blossom and will range far and wide — up to three miles
22	<a href="#">A0G 1153</a>	. So if you're in an area where there are other	<a href="#">apple</a> trees, you should have no problem. However, according to
23	<a href="#">A0G 1157</a>	garden tree, and your pollination worries will be banished forever.	<a href="#">Apple</a> tree care The secret of efficient and easy pruning is to take

Your query "apple" returned 2606 hits in 789 different texts (98,313,429 words [4,048 texts]; frequency: 26.51 instances per million words)

< << >> >| Show Page: 1 Show Sentence View Show in random order Show extended audio data controls **New Query** ▼ Go!

No	Filename	Hits 1 to 50	Page 1 / 53
1	<a href="#">A06 1192</a>	; you need to convey the spurts of energy vocally. The	<a href="#">Apple</a> Cart by George Bernard Shaw The play is a fable, of
2	<a href="#">A0C 1195</a>	National in Portsmouth won the top dessert prize for her hot Banbury	<a href="#">apple</a> pie. All the winning dishes are to be featured in a
3	<a href="#">A0C 1222</a>	basil butter; roast duckling breast with clover honey and a saffron	<a href="#">apple</a> compote and iced Drambuie parfait with sour cherries. Norbert Stening and
4	<a href="#">A0G 47</a>	picks out plants that positively enjoy feet of clay. Specialist spotlight	<a href="#">Apple</a> appeal Graham Rice takes the graft out of choosing apple rootstocks and
5	<a href="#">A0G 47</a>	Specialist spotlight Apple appeal Graham Rice takes the graft out of choosing	<a href="#">apple</a> rootstocks and varieties as he visits a specialist fruit tree nursery.
6	<a href="#">A0G 99</a>	shrink from a heavy soil — Nigel Colborn on clay, p26	<a href="#">Apple</a> growing made as easy as pie Anne Swithinbank's brilliant bulbs for
7	<a href="#">A0G 132</a>	were brought back from the Horn of Africa for Queen Hatshepsut.	<a href="#">Apple</a> pie order Learn everything you ever wanted to know about apples while
8	<a href="#">A0G 133</a>	to know about apples while tasting cider and trying out some scrumptious	<a href="#">apple</a> recipes at Priorwood Garden, Roxburghshire on September 29. The Apple
9	<a href="#">A0G 134</a>	apple recipes at Priorwood Garden, Roxburghshire on September 29. The	<a href="#">Apple</a> Day runs from 2–6pm and includes an apple question time session,
10	<a href="#">A0G 134</a>	September 29. The Apple Day runs from 2–6pm and includes an	<a href="#">apple</a> question time session, with experts on hand to identify your mystery
11	<a href="#">A0G 647</a>	store is frost-free. • Place a greaseband round the trunk of	<a href="#">apple</a> trees to catch the wingless female winter moth as it crawls up
12	<a href="#">A0G 1088</a>	tree Hedgehogs are the most welcome of garden pest controllers Specialist Spotlight	<a href="#">Apple</a> appeal Growing apples is as easy as apple pie, if you
13	<a href="#">A0G 1089</a>	pest controllers Specialist Spotlight Apple appeal Growing apples is as easy as	<a href="#">apple</a> pie, if you are Joan Greenway of Highfield Nurseries. Graham
14	<a href="#">A0G 1091</a>	the fruit specialists, she described to me her recipe for trouble-free	<a href="#">apple</a> growing: disease-resistant varieties, the right rootstock for your preferred size
15	<a href="#">A0G 1091</a>	of tree, a simple summer and winter prune plus a crab	<a href="#">apple</a> to take care of your pollination needs. She made it all
16	<a href="#">A0G 1093</a>	she and other growers use have certainly taken the anxiety out of	<a href="#">apple</a> growing, a subject wreathed in mystery, according to traditional literature
17	<a href="#">A0G 1094</a>	in mystery, according to traditional literature. Just how small can	<a href="#">apple</a> trees get? Small enough to step over, as Highfield's
18	<a href="#">A0G 1111</a>	are really too big for most gardeners to manage. Choosing rootstocks	<a href="#">Apple</a> trees are made up of two parts, the fruiting variety and
19	<a href="#">A0G 1115</a>	' explains Joan. 'If you simply take cuttings from an	<a href="#">apple</a> tree they will grow vigorously but take years and years to fruit
20	<a href="#">A0G 1147</a>	that group will pollinate most other varieties. The small group of	<a href="#">apple</a> varieties know as triploids present more of a challenge. Although triploids
21	<a href="#">A0G 1152</a>	may not be lost. As Joan points out, bees love	<a href="#">apple</a> blossom and will range far and wide — up to three miles
22	<a href="#">A0G 1153</a>	. So if you're in an area where there are other	<a href="#">apple</a> trees, you should have no problem. However, according to
23	<a href="#">A0G 1157</a>	garden tree, and your pollination worries will be banished forever.	<a href="#">Apple</a> tree care The secret of efficient and easy pruning is to take

Your query "apple" returned 2606 hits in 789 different texts (98,313,429 words [4,048 texts]; frequency: 26.51 instances per million words)

< << >> >| Show Page: 1 Show Sentence View Show in random order Show extended audio data controls Go!

- New Query
- New Query
- Thin...
- Frequency breakdown
- Distribution
- Sort
- Collocations...
- Download...
- Categorize hits...
- Save current set of hits...

No	Filename	Hits 1 to 50		Page 1 / 53
1	<a href="#">A06 1192</a>			
2	<a href="#">A0C 1195</a>			
3	<a href="#">A0C 1222</a>			
4	<a href="#">A0G 47</a>			
5	<a href="#">A0G 47</a>			
6	<a href="#">A0G 99</a>			
7	<a href="#">A0G 132</a>			
8	<a href="#">A0G 133</a>			
9	<a href="#">A0G 134</a>			
10	<a href="#">A0G 134</a>			
11	<a href="#">A0G 647</a>			
12	<a href="#">A0G 1088</a>			
13	<a href="#">A0G 1089</a>			
14	<a href="#">A0G 1091</a>			
15	<a href="#">A0G 1091</a>			
16	<a href="#">A0G 1093</a>			

# Algunas funciones interesantes de esta interfaz

- ❑ Frequency breakdown → proporciona la frecuencia cuando hay más de una posibilidad
- ❑ Ejemplo: *said*, “ → *said* \, “
- ❑ Distribución de la palabra buscada en los tipos de textos del corpus
- ❑ Se puede ordenar alfabéticamente las palabras de la derecha o la izquierda
- ❑ Colocaciones
- ❑ Permite la descarga

# CREA

- ❑ 154 millones de palabras
- ❑ Corpus de referencia
- ❑ Monolingüe: español
- ❑ De 1975 a 2004
- ❑ Recoge muestras de todos los países de habla hispana, incluyendo América
- ❑ 90% escrito y 10% oral
- ❑ <http://corpus.rae.es/creanet.html>

## Real Academia Española - Corpus de Referencia del Español Actual (CREA)

<b>Consulta:</b> <input type="text"/>			
<b>Criterios de selección:</b>			
<b>Autor:</b>	<input type="text"/>	<b>Obra:</b>	<input type="text"/>
<b>Cronológico:</b>	<input type="text"/> <input type="text"/>	<b>Medio:</b>	<b>Geográfico:</b>
		(Todos) ▲ Libros Periódicos Revistas Miscelánea Oral ▼	(Todos) ▲ Argentina Bolivia Chile Colombia Costa Rica ▼
<b>Tema:</b>	(Todos) ▲ 1.- Ciencias y Tecnología. 101.- Biología. 102.- Veterinaria. 103.- Ecología. 104.- Tecnología. ▼		

Buscar

Limpiar

# Algunas funciones interesantes de esta interfaz

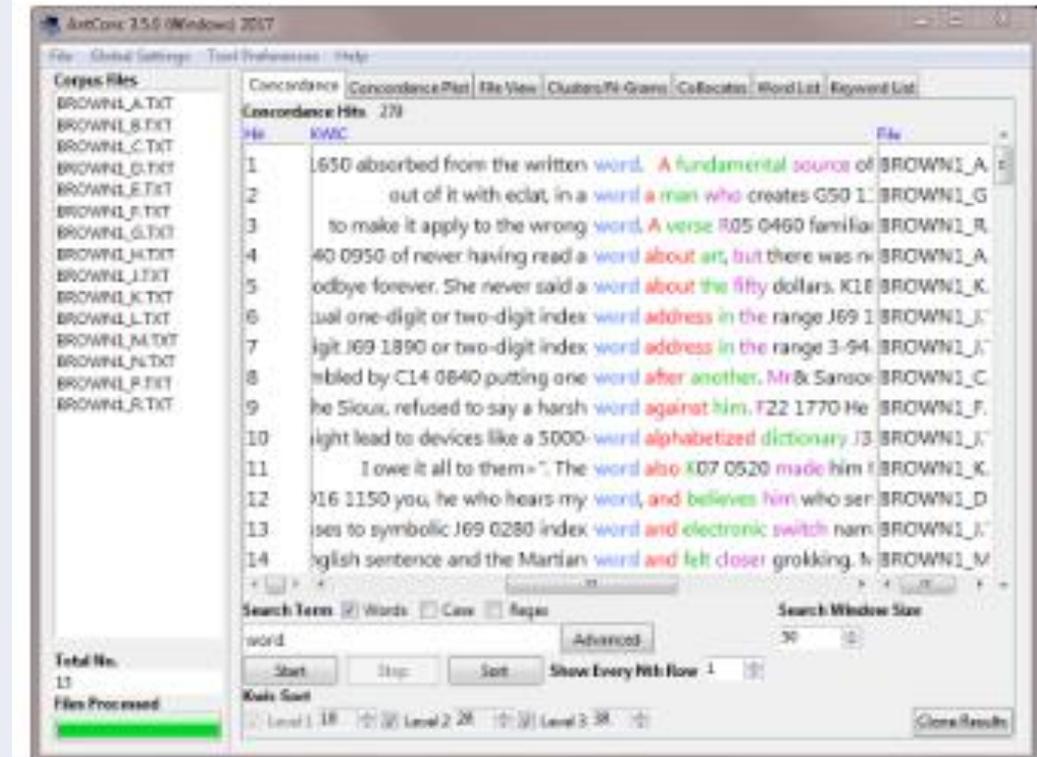
- ❑ Se pueden utilizar operadores lógicos → Y, O, NO
  - ❑ *manzana Y pera O naranja* – se pueden usar paréntesis
- ❑ Existe un operador de distancia → DIST/
  - ❑ *manzana DIST/5 pera* – la palabra manzana tiene que aparecer a una distancia no superior de cinco palabras
- ❑ También se pueden usar comodines
  - ❑ \* = cualquier número de caracteres
    - ❑ *amar\** → amaré, amaría, etc.
  - ❑ ? = sustituye a un carácter en una posición determinada
    - ❑ *c?sa* → cosa y casa

# Herramientas digitales

- ❑ AntConc
- ❑ Sketch Engine

# AntConc

- ❑ Software gratuito de análisis de corpus
- ❑ <http://www.laurenceanthony.net/software.html>



## Otros softwares interesantes en la página de AntConc

- ❑ **AntFileConverter** → para convertir PDF y WORD en texto plano
- ❑ **AntPConc** → para construir corpus paralelos
- ❑ **TagAnt** → un etiquetador de las partes del discurso (POS) construido a partir de TreeTagger
- ❑ **FireAnt** → (Filter, Identify, Report, and Export Analysis Toolkit) análisis de tuits y redes sociales, en general, con herramientas de visualización.

# AntConc

- ❑ Abre AntConc > File > Open file / Open dir

### Corpus Files

01-06-2018.txt  
01-07-2019.txt  
02-03-2018.txt  
02-03-2019.txt  
02-06-2019.txt  
02-07-2018.txt  
02-12-2017.txt  
02-12-2018.txt  
03-02-2018.txt  
03-02-2019.txt  
03-06-2019.txt  
04-05-2018.txt  
04-11-2018.txt  
04-12-2017.txt  
05-01-2018.txt  
05-01-2019.txt  
05-04-2019.txt  
05-05-2019.txt  
05-11-2018.txt  
06-04-2018.txt  
06-05-2019.txt  
06-09-2019.txt  
07-02-2018.txt  
07-04-2019.txt  
07-07-2018.txt  
07-07-2019.txt  
07-10-2018.txt  
08-09-2018.txt  
08-10-2018.txt  
08-12-2017.txt  
08-12-2018.txt  
09-04-2018.txt  
09-06-2019.txt  
10-02-2018.txt  
10-02-2019.txt  
10-03-2018.txt  
10-03-2019.txt

Total No.

125

Files Processed

Concordanc

Concordance Plc

File Vie

Clusters/N-Gram

Collocate

Word Li:

Keyword Li:

Concordance Hits 0

Hit	KWIC	File
-----	------	------

Search Term



Words



Case



Regex

Search Window Size

Advanced

50



Start

Stop

Sort

Show Every Nth Row

1



Kwic Sort



Level 1

1R



Level 2

2R



Level 3

3R



Clone Results

# Algunas funciones interesantes

- Wordlist → palabras ordenadas por frecuencia

### Corpus Files

01-06-2018.txt  
01-07-2019.txt  
02-03-2018.txt  
02-03-2019.txt  
02-06-2019.txt  
02-07-2018.txt  
02-12-2017.txt  
02-12-2018.txt  
03-02-2018.txt  
03-02-2019.txt  
03-06-2019.txt  
04-05-2018.txt  
04-11-2018.txt  
04-12-2017.txt  
05-01-2018.txt  
05-01-2019.txt  
05-04-2019.txt  
05-05-2019.txt  
05-11-2018.txt  
06-04-2018.txt  
06-05-2019.txt  
06-09-2019.txt  
07-02-2018.txt  
07-04-2019.txt  
07-07-2018.txt  
07-07-2019.txt  
07-10-2018.txt  
08-09-2018.txt  
08-10-2018.txt  
08-12-2017.txt  
08-12-2018.txt  
09-04-2018.txt  
09-06-2019.txt  
10-02-2018.txt  
10-02-2019.txt  
10-03-2018.txt  
10-03-2019.txt

Total No.

125

Files Processed

Concordanc Concordance Plc File Vie Clusters/N-Gram Collocate **Word Li:** Keyword Li:

**Word Types: 9707** **Word Tokens: 54534** Search Hits: 0

Rank	Freq	Word	Lemma	Word Form(s)
1	2994	de		
2	2323	que		
3	1983	la		
4	1341	a		
5	1256	y		
6	1208	en		
7	1145	el		
8	890	los		
9	879	no		
10	733	un		
11	690	se		
12	677	las		
13	659	una		
14	527	con		
15	491	su		
16	439	es		

Search Term  Words  Case  Regex

Hit Location Search Only 0

Lemma List  Loaded

Word List  Loaded

Sort by  Invert Order

Sort by Freq

Start Stop Sort

Advanced Clone Results

## Algunas funciones interesantes (II)

- ❑ Lista de lemas
- ❑ Tool preferences > Word list > Lemma list
- ❑ Muy útil en español
  - ❑ Comer 9591 – coma 19, comamos 13, etc.

### Corpus Files

01-06-2018.txt  
01-07-2019.txt  
02-03-2018.txt  
02-03-2019.txt  
02-06-2019.txt  
02-07-2018.txt  
02-12-2017.txt  
02-12-2018.txt  
03-02-2018.txt  
03-02-2019.txt  
03-06-2019.txt  
04-05-2018.txt  
04-11-2018.txt  
04-12-2017.txt  
05-01-2018.txt  
05-01-2019.txt  
05-04-2019.txt  
05-05-2019.txt  
05-11-2018.txt  
06-04-2018.txt  
06-05-2019.txt  
06-09-2019.txt  
07-02-2018.txt  
07-04-2019.txt  
07-07-2018.txt  
07-07-2019.txt  
07-10-2018.txt  
08-09-2018.txt  
08-10-2018.txt  
08-12-2017.txt  
08-12-2018.txt  
09-04-2018.txt  
09-06-2019.txt  
10-02-2018.txt  
10-02-2019.txt  
10-03-2018.txt  
10-03-2019.txt

Total No.

125

Files Processed

Concordanc	Concordance Plc	File Vie	Clusters/N-Gram	Collocate	Word Li:	Keyword Li:
<b>Lemma Types:</b> 6047		<b>Lemma Tokens:</b> 54534		<b>Search Hits:</b> 0		
Rank	Freq	Lemma	Lemma Word Form(s)			
1	2994	de	de 2994			
2	2418	lo	la 1983 lo 435			
3	2323	que	que 2323			
4	1567	los	las 677 los 890			
5	1508	uno	un 733 una 659 unas 29 uno 33 unos 54			
6	1341	a	a 1341			
7	1256	y	y 1256			
8	1208	en	en 1208			
9	1145	el	el 1145			
10	965	no	no 879 nos 86			
11	963	ser	era 75 eran 31 eres 6 es 439 fue 70 fuera			
12	819	haber	ha 247 haber 29 habido 1 habrá 7 habrán			
13	709	su	su 491 sus 218			
14	690	se	se 690			
15	527	con	con 527			
16	478	nr	nr 478			

**Search Term**  Words  Case  Regex

españoles  **Hit Location** Search Only 0

**Lemma List**  Loaded

**Word List**  Loaded

**Sort by**  Invert Order

Sort by Freq

# Algunas funciones interesantes (III)

- ❑ Lista de palabras vacías (*stopwords*)
  - ❑ Tool preferences > Word list > Word List Range > Use a stoplist below
  - ❑ 2 opciones
    - ❑ Manual
    - ❑ Subir una lista

### Corpus Files

01-06-2018.txt  
01-07-2019.txt  
02-03-2018.txt  
02-03-2019.txt  
02-06-2019.txt  
02-07-2018.txt  
02-12-2017.txt  
02-12-2018.txt  
03-02-2018.txt  
03-02-2019.txt  
03-06-2019.txt  
04-05-2018.txt  
04-11-2018.txt  
04-12-2017.txt  
05-01-2018.txt  
05-01-2019.txt  
05-04-2019.txt  
05-05-2019.txt  
05-11-2018.txt  
06-04-2018.txt  
06-05-2019.txt  
06-09-2019.txt  
07-02-2018.txt  
07-04-2019.txt  
07-07-2018.txt  
07-07-2019.txt  
07-10-2018.txt  
08-09-2018.txt  
08-10-2018.txt  
08-12-2017.txt  
08-12-2018.txt  
09-04-2018.txt  
09-06-2019.txt  
10-02-2018.txt  
10-02-2019.txt  
10-03-2018.txt  
10-03-2019.txt

Total No.

125

Files Processed

Concordanc   Concordance Plc   File Vie   Clusters/N-Gram   Collocate   **Word Li:**   Keyword Li:

Word Types: 9210   Word Tokens: 20855   Search Hits: 0

Rank	Freq	Word	Lemma Word Form(s)
1	138	años	
2	78	vida	
3	69	españa	
4	62	gobierno	
5	56	casa	
6	51	madrid	
7	43	mujeres	
8	43	siquiera	
9	42	año	
10	42	madre	
11	41	historia	
12	40	personas	
13	38	hijos	
14	37	país	
15	33	derecha	
16	31	mujer	

Search Term  Words  Case  Regex

Hit Location  Search Only

Advanced

Start Stop Sort

Lemma List  Loaded

Word List  Loaded

Sort by  Invert Order

Sort by Freq

Clone Results

**Corpus Files**

01-06-2018.txt  
 01-07-2019.txt  
 02-03-2018.txt  
 02-03-2019.txt  
 02-06-2019.txt  
 02-07-2018.txt  
 02-12-2017.txt  
 02-12-2018.txt  
 03-02-2018.txt  
 03-02-2019.txt  
 03-06-2019.txt  
 04-05-2018.txt  
 04-11-2018.txt  
 04-12-2017.txt  
 05-01-2018.txt  
 05-01-2019.txt  
 05-04-2019.txt  
 05-05-2019.txt  
 05-11-2018.txt  
 06-04-2018.txt  
 06-05-2019.txt  
 06-09-2019.txt  
 07-02-2018.txt  
 07-04-2019.txt  
 07-07-2018.txt  
 07-07-2019.txt  
 07-10-2018.txt  
 08-09-2018.txt  
 08-10-2018.txt  
 08-12-2017.txt  
 08-12-2018.txt  
 09-04-2018.txt  
 09-06-2019.txt  
 10-02-2018.txt  
 10-02-2019.txt  
 10-03-2018.txt  
 10-03-2019.txt

**Total No.**

125

**Files Processed**

Concordanc   Concordance Plc   File Vie   Clusters/N-Gram   Collocate   **Word Li:**   Keyword Li:

**Lemma Types:** 5871   **Lemma Tokens:** 20855   **Search Hits:** 0

Rank	Freq	Lemma	Lemma Word Form(s)
1	180	año	año 42 años 138
2	90	llegar	llega 9 llegaba 4 llegaban 4 llegada 1 llega
3	86	vida	vida 78 vidas 8
4	82	casar	casa 56 casado 16 casas 9 casó 1
5	77	volver	volvamos 1 volver 24 volverá 1 volverán 1
6	74	mujer	mujer 31 mujeres 43
7	69	españa	españa 69
8	68	hijo	hija 16 hijas 2 hijo 12 hijos 38
9	64	gobierno	gobierno 62 gobiernos 2
10	60	español	española 24 españolas 6 españoles 30
11	60	ir	iba 25 iban 11 ido 12 irá 1 irán 1 iré 1 iría
12	60	llamar	llama 8 llamaba 5 llamaban 1 llamada 3 lla
13	60	seguir	seguido 1 seguimos 2 seguir 10 seguiría 1
14	57	parecer	parecen 12 parecer 4 parecería 1 parecida
15	57	pensar	pensaba 6 pensado 7 pensando 9 pensar
16	57	persona	persona 17 personas 40

**Search Term**  Words    Case    Regex

españoles   **Advanced**   **Hit Location**   Search Only   0

**Lemma List**  Loaded

**Word List**  Loaded

**Sort by**  Invert Order

Sort by Freq

**Clone Results**

## Algunas funciones interesantes (IV)

- ❑ Concordancias
- ❑ Al clicar en la palabra, aparece el contexto
- ❑ En la opción “Concordances” se pueden realizar búsquedas de palabras concretas
- ❑ Concordance plot – visualización de las palabras en el texto

**Corpus Files**

- 01-06-2018.txt
- 01-07-2019.txt
- 02-03-2018.txt
- 02-03-2019.txt
- 02-06-2019.txt
- 02-07-2018.txt
- 02-12-2017.txt
- 02-12-2018.txt
- 03-02-2018.txt
- 03-02-2019.txt
- 03-06-2019.txt
- 04-05-2018.txt
- 04-11-2018.txt
- 04-12-2017.txt
- 05-01-2018.txt
- 05-01-2019.txt
- 05-04-2019.txt
- 05-05-2019.txt
- 05-11-2018.txt
- 06-04-2018.txt
- 06-05-2019.txt
- 06-09-2019.txt
- 07-02-2018.txt
- 07-04-2019.txt
- 07-07-2018.txt
- 07-07-2019.txt
- 07-10-2018.txt
- 08-09-2018.txt
- 08-10-2018.txt
- 08-12-2017.txt
- 08-12-2018.txt
- 09-04-2018.txt
- 09-06-2019.txt
- 10-02-2018.txt
- 10-02-2019.txt
- 10-03-2018.txt
- 10-03-2019.txt

**Total No.**

125

**Files Processed**

Concordanc    Concordance Plc    File Vie    Clusters/N-Gram    Collocate    Word Li:    Keyword Li:

**Concordance Hits**    69

Hit	KWIC	File
1	upto de la historia democrática de	España, y punto. La gran esperanz
2	aís, pero no siempre se puede. En	España, algunos no se han enterac
3	erie sobre lo que está pasando en	España. No la entenderían en ning
4	ra, interesarme por lo que pasa en	España cada día me da más miedc
5	ilidad política. Así de fácil. Ahora,	España es lo que importa, y Españ
6	hora, España es lo que importa, y	España, por lo visto, es cuestión d
7	precisamente ahora, y sólo ahora,	España haya vuelto a ganar un car
8	equidistancia para afirmar que en	España fascistas y demócratas fue
9	siste la simple aplicación de la ley.	España no se rompe solo por Cata
10	investigar lo que había pasado en	España en la primera mitad del XX
11	a de contar la historia reciente de	España, estremecedor el paso del
12	al, peor. En los últimos tiempos, la	España donde creíamos vivir ha ch
13	eza volver a estudiar la historia de	España, puede repasar la de Catal
14	principal virtud. En 1898, cuando	España perdió sus últimas colonia:
15	ezó este proceso, la educación en	España llevaba unos cuarenta año
16	ue pasó el 28-A tiene que ver con	España, que no es una bandera, ni
17	o único, normal que ha pasado en	España últimamente. Los nietos y
--		-----

Search Term     Words     Case     Regex

españa|

Advanced

Search Window Size

50

Start

Stop

Sort

Show Every Nth Row

1

Kwic Sort

Level 1

1R

Level 2

2R

Level 3

3R

Clone Results

### Corpus Files

01-06-2018.txt  
01-07-2019.txt  
02-03-2018.txt  
02-03-2019.txt  
02-06-2019.txt  
02-07-2018.txt  
02-12-2017.txt  
02-12-2018.txt  
03-02-2018.txt  
03-02-2019.txt  
03-06-2019.txt  
04-05-2018.txt  
04-11-2018.txt  
04-12-2017.txt  
05-01-2018.txt  
05-01-2019.txt  
05-04-2019.txt  
05-05-2019.txt  
05-11-2018.txt  
06-04-2018.txt  
06-05-2019.txt  
06-09-2019.txt  
07-02-2018.txt  
07-04-2019.txt  
07-07-2018.txt  
07-07-2019.txt  
07-10-2018.txt  
08-09-2018.txt  
08-10-2018.txt  
08-12-2017.txt  
08-12-2018.txt  
09-04-2018.txt  
09-06-2019.txt  
10-02-2018.txt  
10-02-2019.txt  
10-03-2018.txt  
10-03-2019.txt

Total No.

125

Files Processed

Concordanc	Concordance Plc	File Vie	Clusters/N-Gram	Collocate	Word Li:	Keyword Li:
<b>Concordance Hits 69</b>						
Hit	KWIC	File				
1	s constato que, por ejemplo, toda	España albergaba la esperanza de	27-01-2019.t			
2	aís, pero no siempre se puede. En	España, algunos no se han enterac	02-06-2019.t			
3	pezó con una cita de Camus –"en	España aprendimos que se puede	11-03-2019.t			
4	tólica debería pagar impuestos en	España. Aunque al cabo quizás se	11-11-2018.t			
5	la, club altivo y generoso, eres de	España aureola y del fútbol el colo	19-11-2018.t			
6	ra, interesarme por lo que pasa en	España cada día me da más miedc	02-12-2017.t			
7	ón un nuevo acto de la plataforma	España Ciudadana, que ahora aspi	29-09-2018.t			
8	a, literalmente, rendir y humillar a	España. Como noticia falsa del 28	30-12-2018.t			
9	ento. Ahora, mientras otros gritan ;	España! con ardor guerrero, nosot	13-04-2019.t			
10	nal asociada desde hace siglos en	España con la indefensión y la deb	06-05-2019.t			
11	l atender a lo que ocurre fuera de	España, con la única excepción de	08-12-2017.t			
12	La Historia —de la humanidad, de	España, de la política, de la democ	21-07-2019.t			
13	híbrido perfecto entre la imperial	España del Caudillo y la tenebrosa	25-02-2019.t			
14	os públicos, con aquella pletórica	España del pelotazo donde se prej	16-03-2018.t			
15	cambios que se han producido en	España desde que se aprobó, ni q	17-11-2018.t			
16	al, peor. En los últimos tiempos, la	España donde creíamos vivir ha ch	05-01-2018.t			
17	a flanquea el edificio del Banco de	España, en la calle de Alcalá de M	20-11-2017.t			
--			-----			

Search Term  Words  Case  Regex

españa|

Advanced

Search Window Size

50

Start

Stop

Sort

Show Every Nth Row

1

Kwic Sort

Level 1

1R

Level 2

2R

Level 3

3R

Clone Results

**Corpus Files**

01-06-2018.txt  
 01-07-2019.txt  
 02-03-2018.txt  
 02-03-2019.txt  
 02-06-2019.txt  
 02-07-2018.txt  
 02-12-2017.txt  
 02-12-2018.txt  
 03-02-2018.txt  
 03-02-2019.txt  
 03-06-2019.txt  
 04-05-2018.txt  
 04-11-2018.txt  
 04-12-2017.txt  
 05-01-2018.txt  
 05-01-2019.txt  
 05-04-2019.txt  
 05-05-2019.txt  
 05-11-2018.txt  
 06-04-2018.txt  
 06-05-2019.txt  
 06-09-2019.txt  
 07-02-2018.txt  
 07-04-2019.txt  
 07-07-2018.txt  
 07-07-2019.txt  
 07-10-2018.txt  
 08-09-2018.txt  
 08-10-2018.txt  
 02-07-2018.txt  
 08-12-2018.txt  
 09-04-2018.txt  
 09-06-2019.txt  
 10-02-2018.txt  
 10-02-2019.txt  
 10-03-2018.txt  
 10-03-2019.txt

**Total No.**

125

**Files Processed**

Concordanc

**Concordance Plc**

File Vie

Clusters/N-Gram

Collocate

Word Li:

Keyword Li:

**Concordance Hits** 69**Total Plots (with hits)** 45

Plot: 1 FILE: 01-06-2018.txt

Hits: 1  
Chars: 1747

Plot: 2 FILE: 02-06-2019.txt

Hits: 2  
Chars: 1747

Plot: 3 FILE: 02-12-2017.txt

Hits: 3  
Chars: 1743

Plot: 4 FILE: 02-12-2018.txt

Hits: 1  
Chars: 1754

Plot: 5 FILE: 03-02-2018.txt

Hits: 2  
Chars: 1755**Search Term**

Words



Case



Regex

**Plot Zoom**

españa

Advanced

x1



Start

Stop

**Show Every Nth Row**

1



# Algunas funciones interesantes (V)

- ❑ Search Term
  - ❑ Words
  - ❑ Case
  - ❑ Regex – expresiones regulares - <https://regexr.com/>
    - ❑ \* = repetición del carácter anterior 0 o más veces – be\* → b, be, bee
    - ❑ [ ] = alternancia → dance[sd]
    - ❑ ? = el carácter anterior es opcional → dances?
    - ❑ ^ = niega todo lo que va detrás → [^aeiou]
    - ❑ + = más de una vez
    - ❑ . = cualquier carácter → danc. → dancers, dancing, dances

**Corpus Files**

01-06-2018.txt  
 01-07-2019.txt  
 02-03-2018.txt  
 02-03-2019.txt  
 02-06-2019.txt  
 02-07-2018.txt  
 02-12-2017.txt  
 02-12-2018.txt  
 03-02-2018.txt  
 03-02-2019.txt  
 03-06-2019.txt  
 04-05-2018.txt  
 04-11-2018.txt  
 04-12-2017.txt  
 05-01-2018.txt  
 05-01-2019.txt  
 05-04-2019.txt  
 05-05-2019.txt  
 05-11-2018.txt  
 06-04-2018.txt  
 06-05-2019.txt  
 06-09-2019.txt  
 07-02-2018.txt  
 07-04-2019.txt  
 07-07-2018.txt  
 07-07-2019.txt  
 07-10-2018.txt  
 08-09-2018.txt  
 08-10-2018.txt  
 02-07-2018.txt  
 08-12-2018.txt  
 09-04-2018.txt  
 09-06-2019.txt  
 10-02-2018.txt  
 10-02-2019.txt  
 10-03-2018.txt  
 10-03-2019.txt

**Total No.**

125

**Files Processed**

Concordanc	Concordance Plc	File Vie	Clusters/N-Gram	Collocate	Word Li:	Keyword Li:
<b>Concordance Hits 69</b>						
Hit	KWIC	File				
37	amos responder con el nombre de	España. Pronunciarlo sin levantar l				
38	sangre. Si el 22 de julio no hay en	España un Gobierno de izquierdas				
39	o a un acuerdo con el gobierno de	España. Que ERC lo tumbara a pes				
40	sear envueltos en una bandera de	España estaban pendientes de las				
41	os públicos, con aquella pletórica	España del pelotazo donde se prej				
42	que ha resucitado el fascismo en	España. Y no sé si alguien leerá es				
43	o soy sospechosa. El Gobierno de	España, sin embargo, lo es. A mí, e				
44	rograsa adecuadamente. La "anti-	España" se le resiste, pero ya ha c				
45	urren muchas maneras de romper	España, y casi todas tienen un lug				
46	a cosa es que Napoleón dijera que	España era diferente y otra que la				
47	cambios que se han producido en	España desde que se aprobó, ni q				
48	la, club altivo y generoso, eres de	España aureola y del fútbol el colo				
49	es una novedad lamentable. Si en	España no pasara nada más, no m				
50	ada más, no me quejaría. Pero en	España pasan muchas cosas, apar				
51	e de honor del segundo banco de	España y una organización mafios				
52	a flanquea el edificio del Banco de	España, en la calle de Alcalá de M				
53	erecía que hubiéramos convertido	España en un país mucho mejor qu				

Search Term  Words  Case  Regex

españa

Advanced

Search Window Size

50

Start

Stop

Sort

Show Every Nth Row

1

Kwic Sort

 Level 1

1R

 Level 2

2R

 Level 3

3R

Clone Results

# Algunas funciones interesantes (VI)

- ❑ Analizar más que palabras
  - ❑ Clusters → grupos de palabras alrededor de una *keyword*
  - ❑ N-grams → secuencia contigua de n palabras
  - ❑ Collocaciones → palabras que ocurren juntas con frecuencia

# Sketch Engine

- ❑ Herramienta de corpus para explorar el lenguaje
- ❑ Gestor de corpus
- ❑ <https://www.sketchengine.eu/>
- ❑ Tiene sus propios corpus pero también se pueden añadir propios o compilarlos de Internet
  - ❑ Spanish Web 2011 (esTenTen 11) = 10000 million words
  - ❑ English Web 2015 (enTenTen 15) = 15000 million words
  - ❑ Para crear tu corpus = New corpus – Name – Language – Description
    - ❑ Find texts on the web – palabras claves
    - ❑ I have my own texts – subir archivos en formato .pdf o .txt

# Word Sketch

modifiers of "aprender"	
<b>mucho</b> ...	aprender mucho
<b>rápido</b> ...	aprender rápido
<b>rápidamente</b> ...	aprender rápidamente
<b>pronto</b> ...	pronto aprendió
<b>primero</b> ...	aprender primero
<b>continuamente</b> ...	aprender continuamente

objects of "aprender"	
<b>lección</b> ...	lecciones aprendidas
<b>idioma</b> ...	aprender idiomas
<b>inglés</b> ...	aprender inglés
<b>técnica</b> ...	aprender técnicas
<b>oficio</b> ...	aprender un oficio
<b>lengua</b> ...	aprender la lengua

subjects of "aprender"	
<b>alumno</b> ...	alumnos aprenden
<b>niño</b> ...	niños aprenden
<b>estudiante</b> ...	estudiantes aprenden
<b>lección</b> ...	lección que aprendí
<b>participante</b> ...	participantes aprenderán
<b>hijo</b> ...	hijos aprendan

"aprender" and/or ...	
<b>enseñar</b> ...	enseñar y aprender
<b>practicar</b> ...	aprender y practicar
<b>disfrutar</b> ...	aprender y disfrutar
<b>crecer</b> ...	aprender y crecer
<b>divertir</b> ...	aprender y divertirse
<b>compartir</b> ...	aprender y compartir

prepositional phrases	
"aprender" de ...	...
"aprender" en ...	...
"aprender" sobre ...	...
"aprender" con ...	...
"aprender" a ...	...
"aprender" por ...	...
"aprender" durante	...
...	...
"aprender" desde	...

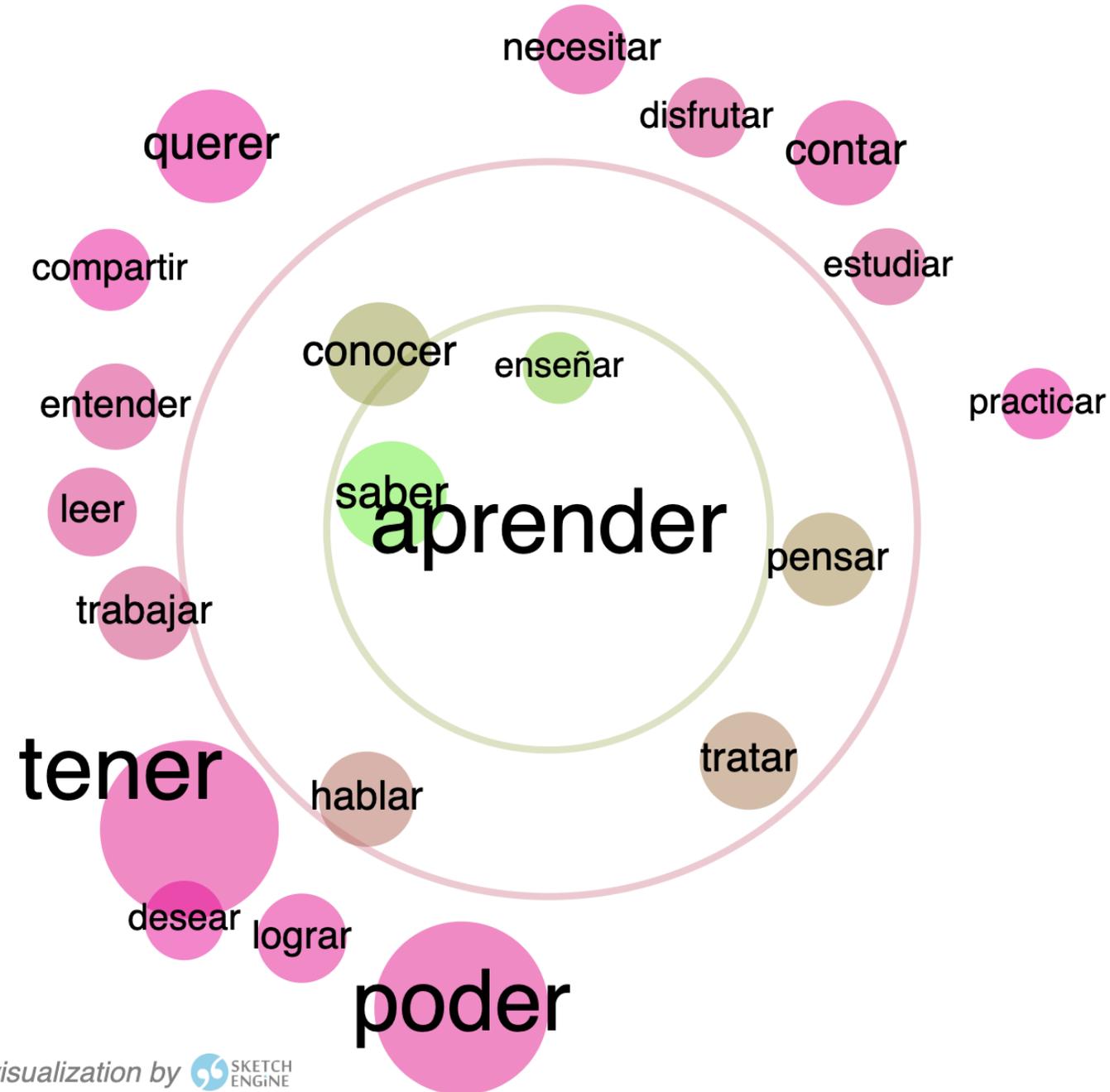
prepositions after "aprender"	
<b>sobre</b> ...	Aprende sobre
<b>a</b> ...	aprender a
<b>de</b> ...	aprender de
<b>durante</b> ...	aprendido durante
<b>desde</b> ...	aprender desde
<b>mediante</b> ...	aprender mediante

"aprender/enseñar" and/or ...			
crecer	6205	63	...
disfrutar	7305	151	...
enseñar	11710	686	...
divertir	4067	298	...
compartir	5207	1062	...
practicar	5149	1374	...
ayudar	1033	1622	...
transmitir	465	985	...
educar	604	1718	...
aprender	2704	11710	...
guiar	79	764	...
predicar	22	1215	...

objects of "aprender/enseñar"			
lección	46525	3827	...
idioma	33954	3493	...
inglés	24011	4119	...
oficio	11974	2400	...
cosa	46711	16005	...
lengua	11059	3451	...
habilidad	9166	2994	...
técnica	27579	10142	...
arte	7042	4383	...
truco	4460	3717	...
doctrina	493	3969	...
diente	0	3181	...

subjects of "aprender/enseñar"			
participante	3546	23	...
alumno	15157	157	...
estudiante	7826	86	...
niño	23568	411	...
hijo	4819	241	...
lección	2902	421	...
profesor	483	2276	...
maestro	346	2613	...
abuelo	158	1840	...
jesús	122	2008	...
biblia	0	1493	...
riek	0	1215	...

# Thesaurus



# Concordance

Details	Left context	KWIC	Right context
1 <a href="#">i</a> wikipedia.org s una película de animación que ofrece por algo menos de 10 dólares		<b>aprender</b>	sobre el sentido de la vida. <a href="#">&lt;/s&gt;&lt;s&gt;</a> ¿Alguna vez te has preguntado ci <a href="#">i</a>
2 <a href="#">i</a> wikipedia.org Worrying and Love Legalized Gambling, llamado \$pringfield (o cómo		<b>aprendí</b>	a amar el juego legalizado) en España y \$pringfield próspero, o el pro <a href="#">i</a>
3 <a href="#">i</a> wikipedia.org ma Mater" (universidad, donde una determinada persona continuado		<b>aprendiendo</b>	) para celebrar el aniversario 650 años de la Universidad Carolina. <a href="#">&lt;/s</a> <a href="#">i</a>
4 <a href="#">i</a> wikipedia.org a la forma correcta de escribir los números, para que la gente lo vea y		<b>aprenda</b>	que así es.--Gustavo Parker (discusión) 03:42 24 jul 2012 (UTC) <a href="#">&lt;/s&gt;</a> <a href="#">i</a>
5 <a href="#">i</a> wikipedia.org rores. <a href="#">&lt;/s&gt;&lt;s&gt;</a> No me di cuenta, los hice sin querer. <a href="#">&lt;/s&gt;&lt;s&gt;</a> Pero ya		<b>aprendí</b>	cómo se editan estas páginas, así que no volveré a cometer esos mis <a href="#">i</a>
6 <a href="#">i</a> wikipedia.org ivo, que altera su forma física y les otorga nuevas habilidades, deben		<b>aprender</b>	a aprovechar sus habilidades y trabajar juntos como un equipo para s <a href="#">i</a>
7 <a href="#">i</a> wikipedia.org a para embarcarse en un viaje no autorizado al Planeta Cero, el cual		<b>aprenden</b>	que es un mundo lleno de sustancias de otro mundo. <a href="#">&lt;/s&gt;&lt;s&gt;</a> Víctor i <a href="#">i</a>
8 <a href="#">i</a> wikipedia.org ipedia --ahianx (discusión) 19:20 14 sep 2011 (UTC) <a href="#">&lt;/s&gt;&lt;s&gt;</a> A ver si		<b>aprendemos</b>	que "shamán" en español no existe y que es "chamán", y lo mismo co <a href="#">i</a>
9 <a href="#">i</a> wikipedia.org no todo 007 es glamour. <a href="#">&lt;/s&gt;&lt;s&gt;</a> Tanto Gil Ro como Seo Won deben		<b>aprender</b>	lo que se necesita para mantener su deber jurado, incluso a costa de <a href="#">i</a>
10 <a href="#">i</a> wikipedia.org ma, "existen 99 nombres que pertenecen sólo a Dios, y aquel que los		<b>aprende</b>	, los comprende y los enumera, entra en el paraíso y alcanza la salva <a href="#">i</a>
11 <a href="#">i</a> wikipedia.org <a href="#">&lt;/s&gt;&lt;s&gt;</a> Plasticidad del aprendizaje: permite a una red neuronal poder		<b>aprender</b>	nuevos patrones. <a href="#">&lt;/s&gt;&lt;s&gt;</a> Estabilidad del aprendizaje: permite a una <a href="#">i</a>

# Frequency

simple **aprender** 3,255,384 (160.31 per million)



## FREQUENCY

BASIC ADVANCED ABOUT

First word to the left

WORD FORMS

PART-OF-SPEECH TAGS

LEMMAS

KWIC

WORD FORMS

PART-OF-SPEECH TAGS

LEMMAS

First word to the right

WORD FORMS

PART-OF-SPEECH TAGS

LEMMAS

More presets

TEXT TYPES

LINE DETAILS

	Word	↓ Frequency	Frequency per million		
1	aprender	1,284,887	63.27		...
2	aprendido	310,159	15.27		...
3	aprende	194,615	9.58		...
4	aprendiendo	185,983	9.16		...
5	Aprende	152,719	7.52		...
6	aprenden	114,753	5.65		...
7	aprendió	94,261	4.64		...

# Colocaciones

simple aprender 3,255,384 (160.31 per million)



	Word	Cooccurrences	Candidates	T-score	MI	↓ LogDice	
1	lección	28,798	182,357	169.53	9.94	8.10	...
2	lecciones	27,033	179,782	164.24	9.87	8.01	...
3	idioma	26,449	637,792	162.00	8.02	7.80	...
4	cómo	62,389	5,953,169	245.96	6.03	7.79	...
5	cosas	65,873	6,546,514	252.57	5.97	7.78	...
6	he	61,796	6,227,785	244.57	5.95	7.74	...

**LogDice** = una medida estadística para identificar los candidatos a una colocación. Se basa en la frecuencia de las palabras  $p_1$  y  $p_2$  y el bigrama  $p_1p_2$ .

# Lista de palabras

BASIC

ADVANCED

ABOUT

Lista de *stopwords*

find ?

words

lemmas

nouns

verbs

adjectives

adverbs

all

starting with

ending with

containing

matching regex

from this list:

✓ Exclude these words:

Paste the list here, one word per line

0

1

2

3

4

5

6

7

8

9

-

a

actualmente

Palabras más frecuentes de un corpus  
Sketch Engine proporciona listas de palabras en  
varios idiomas: [https://www.sketchengine.eu/word-  
lists/](https://www.sketchengine.eu/word-lists/)

Se utilizan para comparar corpus

Word	↓ Frequency ?
1 comida	81,054 ...
2 servicio	56,912 ...
3 platos	55,008 ...
4 restaurante	53,061 ...
5 calidad	50,987 ...

Word	↓ Frequency ?
18 cocina	21,630 ...
19 agradable	21,251 ...
20 menú	20,608 ...
21 personal	20,171 ...
22 plato	19,455 ...

Word	↓ Frequency ?
35 carne	13,397 ...
36 mesas	13,016 ...
37 precios	12,853 ...
38 visita	12,694 ...
39 espectacular	12,688 ...

# N-grams

**BASIC**    ADVANCED

N-gram length   2   **3**   4   5   6

Include nonwords

A = a

**GO**

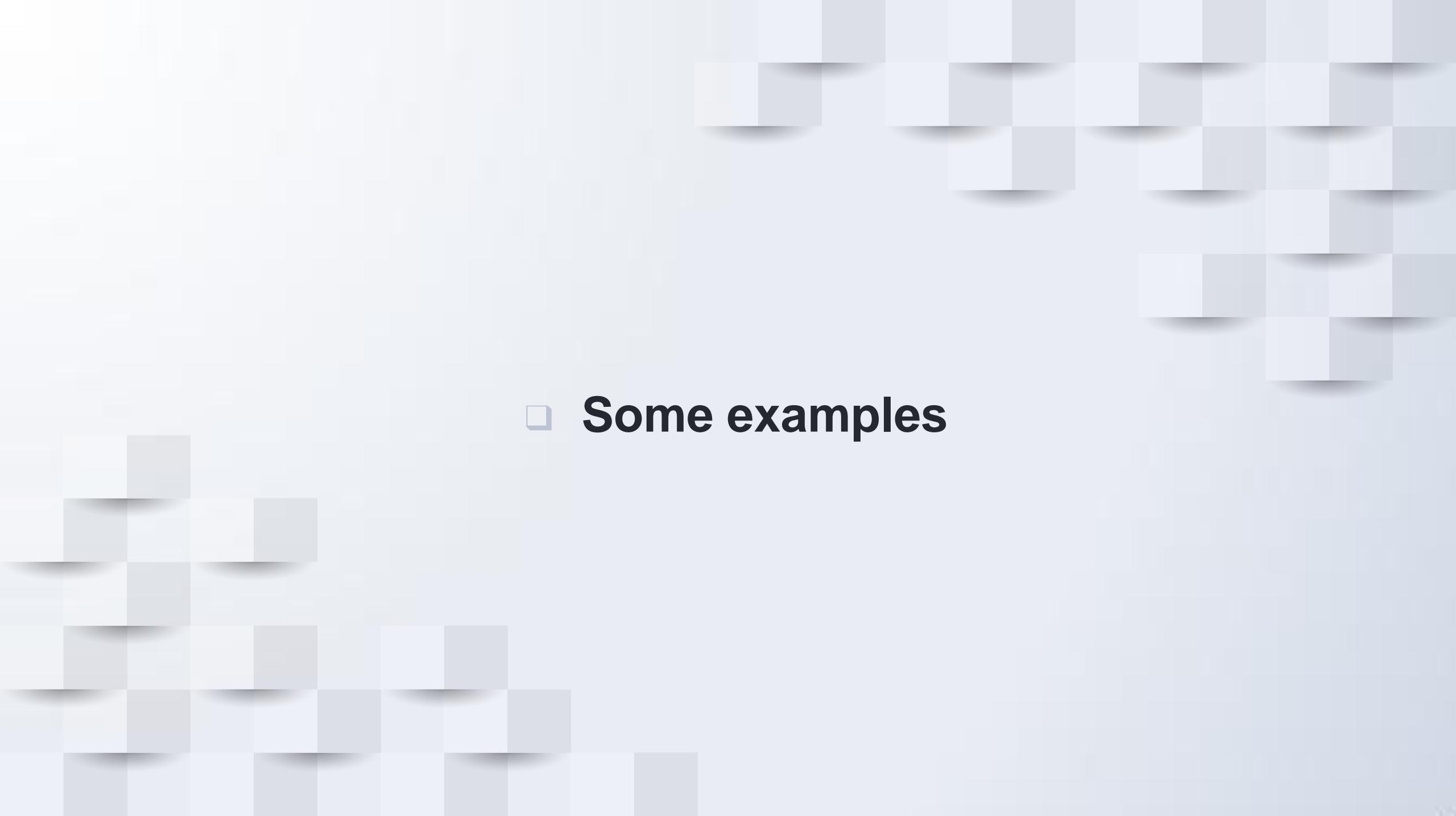
# Keywords

## SINGLE-WORDS

	Word	Focus corpus	Reference corpus	
1	parking	6,000	36,242	...
2	buffet	4,904	38,149	...
3	parador	2,771	26,418	...
4	habitacion	3,108	32,228	...
5	desayuno	16,603	234,566	...

## MULTI-WORDS

	Word	Focus corpus	Reference corpus	
1	habitación amplia	1,990	30	...
2	trato del personal	1,376	7	...
3	personal del hotel	1,323	27	...
4	calidad precio	1,630	114	...
5	personal de recepción	1,095	7	...



- **Some examples**

# Subcorpus para demostrar una teoría semántica

## Agree

1	Hulme	agreed, “	If you were to come from up north and work for certain London councils and come in one morning and say “Hello, chucky ducky lover” it would be instant dismissal.”
2	“I believe,” said Bernard, in and out of the classrooms, cafes, kitchens, street corners, pubs, while Ellen nodded and	agreed, “	I believe.
3	He	agreed, “	I think so.
4	Smiling at her, he	agreed, “	Something like that.”

## Allege

1	Some men were particularly successful in Black Africa, and might return direct to Kufra with slaves, spices, even gold, which they then forwarded north to Alexandria for export, they	alleged, “	to Istanbul”.
2	A few days later while driving him home, she asked him why he’d hit her and, she	alleged, “	He went crazy again, punching me so I nearly lost control of the car.”

## Answer

1	Taff turned away and, staring straight ahead,	answered, “	Yes, piper.
2	You may not recall that, when Edgar Oldroyd asked him how he was enjoying a long spell bowling into the wind, he	answered, “	It’s like bowling up t’bloody cellar steps.”

# Corpus paralelo de onomatopeyas

Search Term  Words  Case Show Nth Entry 1 Search Corpus  Corpus 1  Corpus 2  Corpus 3  
chut Start Context Size 20 words  
Sort 1: CEN Sort 2: 1R Sort 3: 2R Sort Page Size All entries  
Target: Corpus 1 Hits: 1 100% Font Size

Line	KWIC
1	marquise, votre duchesse, votre princesse; elle doit avoir le bras long. -Chut! dit Porthos en mettant un doigt sur ses lèvres, je la crois cardinali

Line	Reference
1	'Why, by your marchioness, your duchess, your princess. She must have a long arm.' 'Hush!' said Porthos, placing a finger on his lips. 'I believe her to be a cardinalist; she must know nothing of the matter.

Line	Reference
1	-Pues por medio de vuestra marquesa, vuestra duquesa, vuestra princesa; debe tener largo el brazo. -¡Chis! -dijo Porthos poniendo un dedo sobre sus labios-. La creo cardenalista y no debe saber nada.

# SentiTur: corpus bilingüe de opiniones del sector turístico

OUT-OF-SCOPE

1 We have just returned from a 4 day break in Mercer Hotel.

Lodging + ↔ [General experience]

2 A wonderful experience in this new hotel in the Gothic Quarter.

Staff [General experience] + ↔ Lodging + ↔ [Design features] Bed + [General experience] ↔ Breakfast + ↔ [General experience]

Lodging + ↑ [General experience]

Staff ★

Lodging ★

Bed ★

Breakfast ★

3 Everything we could have hoped for,

good staff.

stylist decor,

wonderful bed,

lovely breakfast.

4 Thank you