Corpus textuales de nativos y enseñanza/aprendizaje del español como LE/L2

Mar Cruz Piñol, Universitat de Barcelona

http://sites.google.com/site/marcruzpinolub/





Grupo de investigación acqUA aguisidan y aprendizaj e de segundos lengo as y lengoas extranjesas de la Universidad de Alicant

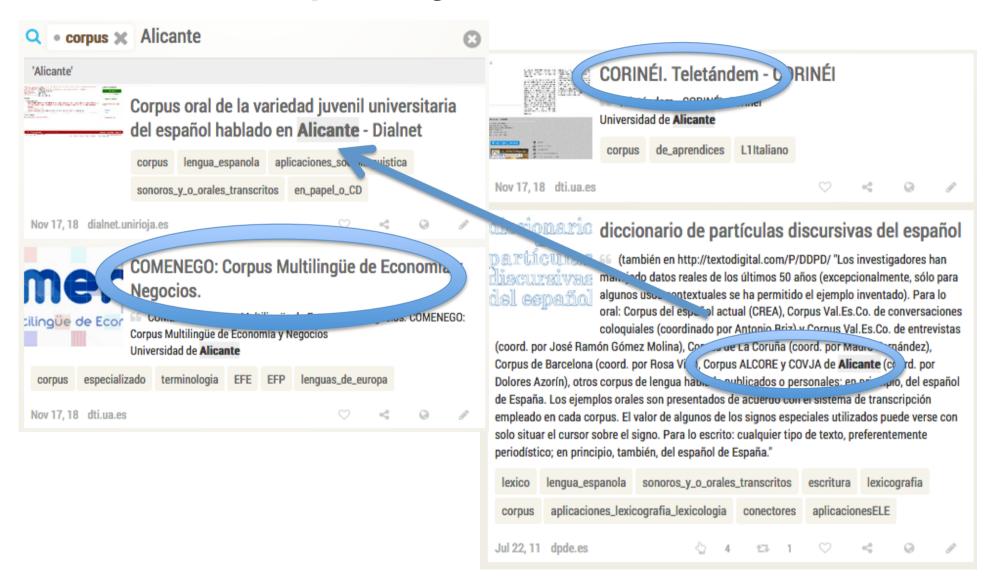
Seminarios de investigación 2018

Veremos...

- Interfaces que muestran ocurrencias, usos y traducción de palabras y expresiones
- Interfaces que muestran relaciones
 sintáctico-semánticas entre palabras
- Materiales y actividades basados en corpus de nativos

... corpus en línea

Aunque hay muchos más:



... para ENSEÑANZA

Corpus y ELE: investigación / enseñanza

Al final...

Un ejercicio o actividad para el aula de ELE con corpus

- Con qué corpus
- Para qué nivel (MCER)
- Qué objetivos
- Qué formación necesita el docente
- Para realizar con los alumnos / Para que el docente prepare materiales
- Dinámica del ejercicio

¿Qué corpus textuales de nativos?

2. 3. 5. 6. 7. 8. 9.

¿Qué ventajas y desventajas?

Muchos corpus textuales de nativos en línea

¿Útiles para el aula de ELE? ¿Difíciles?

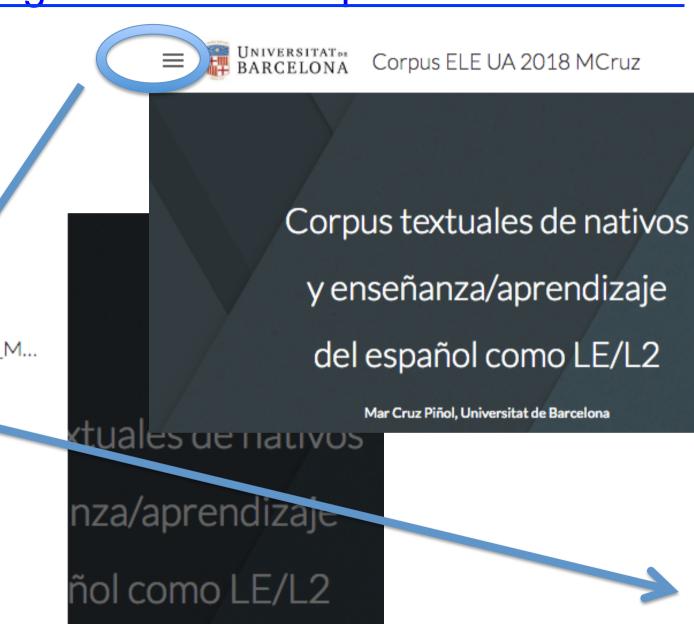
- → Depende
- Conceptos básicos sobre LC
- Tipos de corpus
- Tipos de interfaces
- Proyectos (beta)

Conocer para escoger



http://sites.google.com/view/corpus-ele-ua-mcruz/

ñol, Universitat de Barcelona



X Menú

^ Corpus_ELE_UA_2018_M...

URLs



URLs

1. https://cvc.cervantes.es/ensenanza/biblioteca_ele/asele/pdf/21/21_0277.pdf



¿QUÉ CORPUS EN LÍNEA UTILIZAR PARA QUÉ FINES EN LA CLASE DE ELE?

> Kris Buyse Katholieke Universiteit Leuven / Lessius Hogeschool



La palabra X, ¿ocurre? (¿se dice?)

"enracholar", ¿se dice?

→ URL (2): https://www.google.com/search?q=enracholado

"la dije", ¿se dice?

→ URL (3): http://corpus.leeds.ac.uk/

cqi-bin/cqp.pl?q=la+cal%C3%B3&c=INTERNET-

ES&searchtype=conc&contextsize=60c&sort1=lemma&sort2=right&terminate=100&llstat=on&cleft=0&cright t=1&cfilter=

la web como corpus

corpus ~ norma

ocurrencias, apariciones, casos...

La palabra X, ¿dónde se dice?

La palabra "altoparlante", ¿dónde se dice?

→ URL (4): http://web.frl.es/CORPES/ (lema)

forma flexionada / lema
corpus lematizados
representatividad
corpus de referencia
marcas sociolingüísticas (zonas)
frecuencias (absolutas, relativas...)



Corpus del Español del Siglo XXI (CORPES)

| Co | Co | apariciones | Configuración Ayuda Modo de cita Sugerencias | | |
|------|----------------------|-------------|--|------------------------|----------------------|
| Ler | ma altoparlante | Forma | Clase de palabra (Todos) | ☐ Grafía original | + Subcorpus |
| | | | Concordancia Estadística Nueva consulta | | |
| REF. | (Clasificación, país | s) | CONCORD | ANCIA | Ordenar |
| 1 | 2001 Col. | | sumido en desesperada soledad. Tal es su ensimismamiento, que la estridencia | de los altoparlantes | en la torre de la iç |
| 2 | 2001 Col. | | Semejante regocijo, desbordado y volcánico, acalla las campanas, ac | alla los altoparlantes | de las torres, aca |
| 3 | 2001 Chile | | una estación el tren se detendrá más de lo necesario. El conductor hablará | | excusándose sin |
| 4 | 2001 Arg. | | empecé a chorrear transpiración. Tuvieron que interrumpir la escena por mí. De | sde un altoparlante | , el flaquito de los |
| 5 | 2001 Arg. | | "Todos los extras de Ta condenada" al estudio 3", dijeror | por el altoparlante | en el momento er |
| 6 | 2001 Arg. | | | El altoparlante | anunció nuestra h |
| 7 | 2001 Ur. | | Los tres toman café en silencio. | Por los altoparlantes | anuncian el emba |
| 8 | 2001 Guat. | | a la gente, y se acordó del entrenamiento en sus años de soldado, cuando | por el altoparlante | ordenaban en tod |
| 9 | 2001 Pan. | | con el fin de evitar el Mitin, ya que el Alcalde mantuvo la prohibición. Desc | le unos altoparlantes | se escucha La Ma |
| 10 | 2001 Col. | | | Por los altoparlantes | anunciaron el reir |
| 11 | 2001 Col. | | silencio en la alcoba contrastaba con el repicar de campanas y villancico | s, en el altoparlante | de la iglesia; con |
| 12 | 2001 Col. | | de la conveniencia de no acudir al llamado, que en ese momento le hacíar | por el altoparlante | . Preferible era pe |
| 13 | 2001 Chile | | a hablar con el "futre" (dueño del basural), de quien sólo aparece la voz | por un altoparlante | . Nada consiguen, |
| 14 | 2001 R.Dom. | | , un puente aéreo, un flujo de gente que viene y va. Allí esperábamos la | voz del altoparlante | , el aviso, la orden |
| 15 | 2001 R.Dom. | | El viaje termina, la | voz del altoparlante | dice que se abroc |
| 16 | 2001 Perú | | de la acera, en su lugar se han instalado ambulantes que vociferan con ay | uda de altoparlantes | o a voz en cuello. |
| 17 | 2001 EE.UU. | | Carlos Kunst era casi como ir a otro planeta. La música rock, proveniente | de los altoparlantes | localizados en cua |
| 18 | 2001 EE.UU. | | rimera vez desde que conocía a los fichas, la casa no retumbaba con la música | de los altoparlantes | . No obstante, la te |



| Concord | ancias I Coa | pariciones | I | Configuración I Ayuda | I M |
|---------|--------------|------------|---|-----------------------|------|
| Lema | altoparlante | Forma | | | Clas |
| + | | | | Concordanci | a |

Frecuencia absoluta: 244

Distribución Zona

| Zona | Freq | Fnorm. |
|------------------------|------|-----------|
| Río de la Plata | 56 | 1,71 |
| Caribe continental | 45 | 1,55 |
| Antillas | 42 | 2,25 |
| Chilena | 42 | 2,94 |
| México y Centroamérica | 29 | 0,63 |
| Andina | 19 | 1,05 |
| Estados Unidos | 8 | 2,51 |
| España | 3 | 0,03 |
| 1 - 8 de 8 | | página: 1 |

"(Las concordancias) no solo garantizan el **uso** y **frecuencia** de una unidad léxica determinada, sino que esta aparece **contextualizada** y clasificada en géneros textuales, períodos históricos, zonas geográficas, etc."

[Elena Alonso Pérez-Ávila, 2007, p. 19]

Noviembre 2007 / Boletín de ASELE

→ URL (5): http://formespa.rediris.es/pdfs/asele37.pdf

Qué palabra se usa más, ¿X o Y?

Qué palabra se usa más:

"ordenador", "computadora" o "computador"

→ URL (6):

https://books.google.com/ngrams/graph?content=ordenador%2Ccomputadora
%2Ccomputador&year_start=1800&year_end=2000&corpus=21&smoothing=3&share=&direct_url=t1%
3B%2Cordenador%3B%2Cc0%3B.t1%3B%2Ccomputadora%3B%2Cc0%3B.t1%3B%2Ccomputador
%3B%2Cc0

\rightarrow URL (7):

https://books.google.com/ngrams/graph?content=esqu%C3%ADs%2Cesqu%C3%ADes&year_start=1800&year_end=2000&corpus=21&smoothing=3&share=&direct_url=t1%3B%2Cesqu%C3%ADs%3B%2Cc0%3B.t1%3B%2Cesqu%C3%ADes%3B%2Cc0

la web como corpus norma - uso lúdico (apuestas) registros (libros)

ojo: representatividad ← frecuencia

Los índices de frecuencias pueden servir para...



[Manuel Alvar Ezquerra, 2004, p. 21]

→ URL (8): http://cvc.cervantes.es/ensenanza/biblioteca_ele/asele/pdf/15/15_0017.pdf

- 1º Conocer la importancia de las palabras de acuerdo con su frecuencia y distribución.
- 2º Decidir qué palabras deben enseñarse o no en cada nivel, en cuáles hay que insistir más y cuáles deben enseñarse para que los alumnos entiendan los materiales que manejan.
- 3º Poseer una información objetiva sobre las palabras y cuándo deben enseñarse.
- 4º Extraer el inventario de palabras del ámbito lingüístico en que se mueven los alumnos para facilitarles la comprensión de su entorno.
- 5º Saber cuáles son las palabras de ámbitos concretos cuando se enseña la lengua con fines específicos.
- 6º Seleccionar los grupos de palabras y formas que deben enseñarse conjuntamente, y en qué orden.
- 7º Determinar el grado de dificultad del léxico de los materiales de lectura y aprendizaje.
- 8° Que quienes confeccionan les materiales pedagógicos puedan seleccionar el léxico que debe aparecer en ellos."

Corpus y manuales → EFL (1/2)

[Michael McCarthy, 2004^a, p. 6] [Touchstone]

how we ask questions, and so on."

"A <u>corpus</u> can be a very rich resource for writers of textbooks and other teaching materials because it <u>gives us a detailed view of **how real people speak and write** in everyday situations. It can give us information about vocabulary, grammar, formality and informality, the differences between spoken and written language, how we perform such basic functions as requesting, greeting, and apologizing, how people open and close conversations, how we change the subject, how we interrupt one another,</u>

STUDINE VIMOK

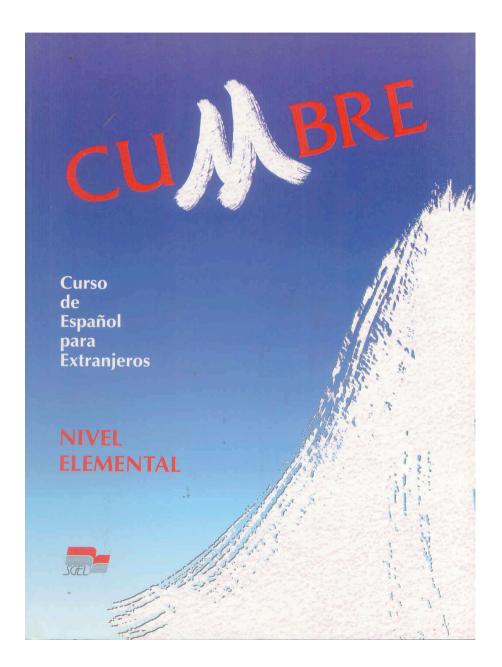
Corpus y manuales → EFL (2/2)

"Many textbooks contain only **invented examples** and their descriptions are based upon intuition or second-hand accounts. Other books, however, are explicitly empirical and use examples and descriptions from corpora or other sources of real life language data. (...) Most studies found that there were considerable differences between what textbooks are teaching and how native speakers actually use language as evidenced in the corpora. Some textbook gloss over important aspects of usage, or foreground less frequent stylistic choices at the expense of more common ones. The general conclusion from these studies is that non-empirically based teaching materials can be misleading and that corpus studies should be used to inform the production of material so that the more common choices of usage are given more attention than those which are less common."

[Tony McEnery & Andrew Wilson, 1996]

→ URL (10): http://ccuc.cbuc.cat/record=b1832534~S23*cat

Corpus y <u>manuales</u> → <u>ELE</u>



Cumbre Aquilino Sánchez (y otros) (1995-2003)

→ URL (11): http://ccuc.cbuc.cat/ record=b1750980~S23*cat

La palabra X, ¿cómo se usa?

La palabra "aunque", ¿cómo se usa?

→ URL (12):

https://es.pons.com/traducci%C3%B3n?q=aunque&l=enes&in=&lf=es

→ URL (13):

https://www.linguee.es/espanol-ingles/search?query=aunque

web como corpus

concordancias (contexto, sintaxis, uso)

corpus: complemento del diccionario

corpus: input (aprendizaje inductivo)

La palabra X, ¿cómo se usa en la lengua oral?

La palabra "mola", ¿cómo se usa en la lengua oral?

→ URL (14): http://valesco.es/?q=palabras

muestras de lengua oral

La expresión X, ¿cómo se dice en otras lenguas?

La expresión "subirse al tren de...", ¿cómo se dice en otras lenguas?

→ URL (15):

https://www.linguee.es/espanol-ingles/traduccion/subirse+al+tren.html

corpus multilingües - diccionarios unidades pluriverbales traducción

La palabra X, ¿cómo se dice en otras lenguas parecidas?

La palabra "ajo",

¿cómo se dice en otras lenguas románicas?



- → URL (16): http://opus.lingfil.uu.se/cwb/OpenSubtitles/frames-cqp.html
- → URL (17): http://corpus.consumer.es/corpus/kontsulta (ajo / hijo / helado)

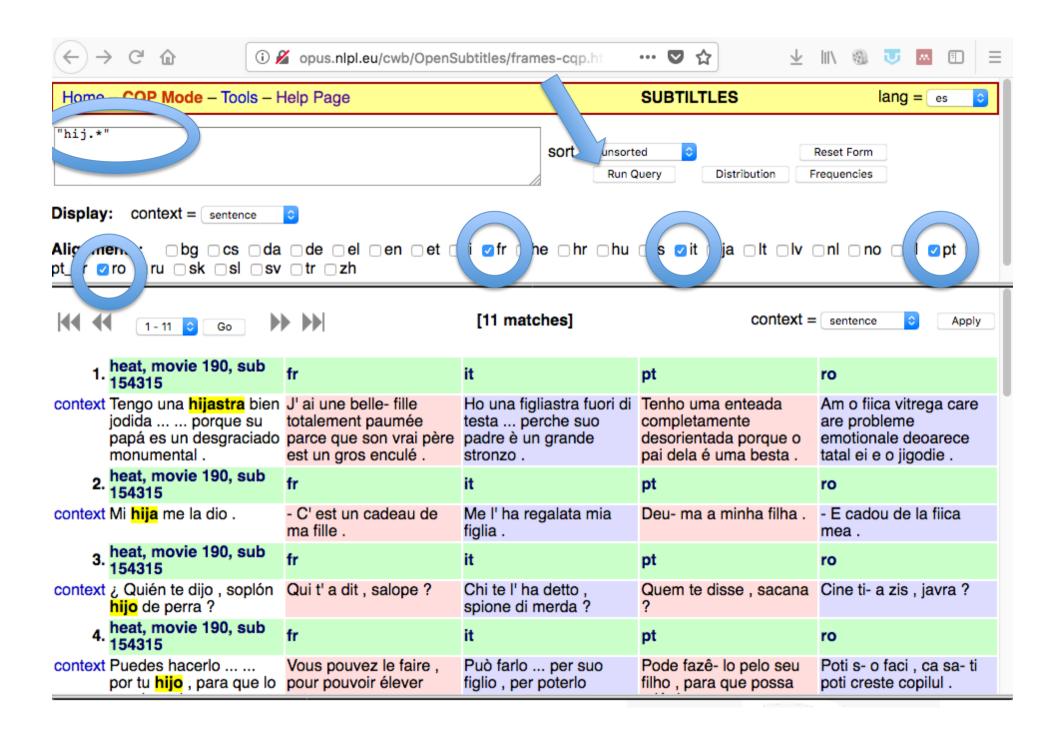
corpus multilingües

corpus lematizados

concordancias

contenidos culturales (lenguas de España)

contenidos lingüísticos (romanística, ha de la lengua, analogías...)



La palabra X aparece en el diccionario, pero, ¿se usa? ¿cómo se usaba?

La palabra "arrufar", ¿se usa? ¿cómo se usaba?

```
→ URL (18): <a href="http://dle.rae.es/srv/fetch?id=3oZagTp">http://dle.rae.es/srv/fetch?id=3oZagTp</a>
```

→ URL (19): http://web.frl.es/CORPES/org/publico/pages/consulta/entradaCompleja.view

→ URL (20): http://corpus.rae.es/cordenet.html

→ URL (21): http://web.frl.es/CNDHE/

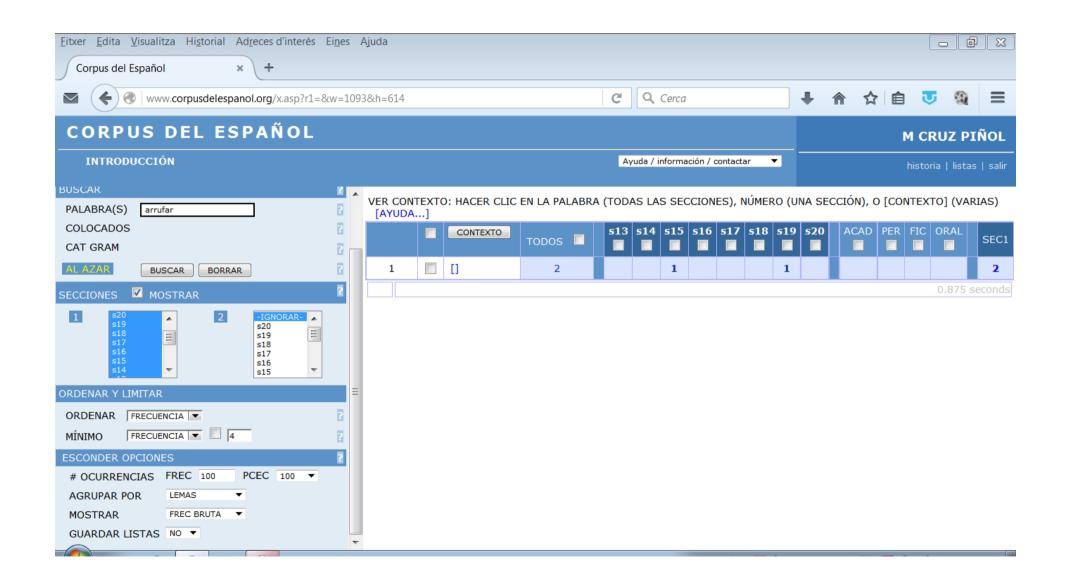
→ URL (22): https://www.corpusdelespanol.org/hist-gen/

corpus "lematizados"

corpus "cronológicos/diacrónicos"

DRAE - "desus" - textos (literatura)

corpus - diccionario



¿Qué palabras se suelen combinar con la palabra X?

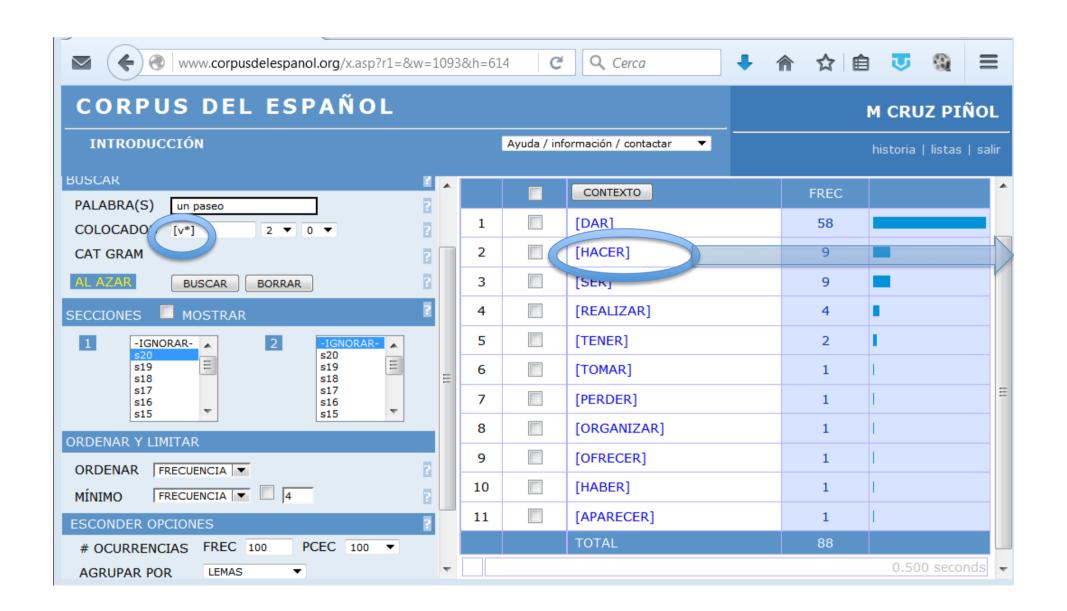
¿Qué palabras se suelen combinar con la palabra "paseo"?

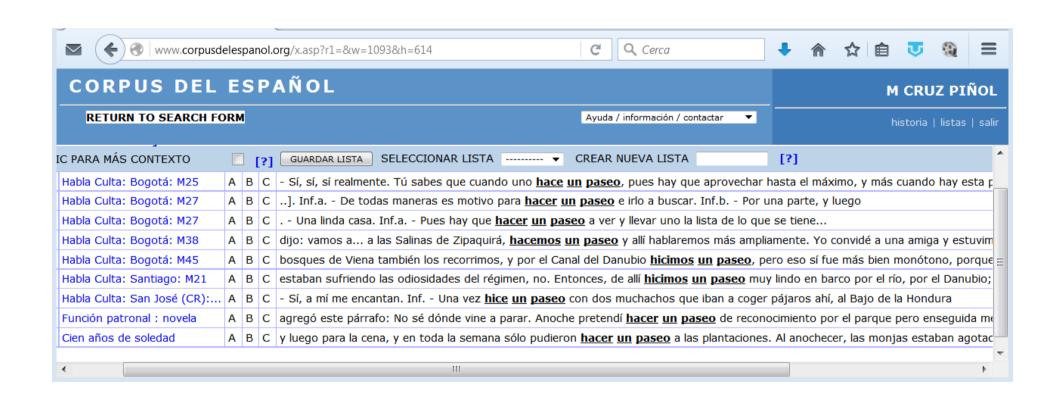
→ URL (23):

http://corpora.informatik.uni-leipzig.de/en/res? word=paseo&corpusId=spa newscrawl 2011

→ URL (24): https://www.corpusdelespanol.org/web-dial/

"agrupaciones", "colocaciones"...
corpus "lematizados"
corpus con etiquetas "POS"





La palabra X, ¿qué relaciones sintáctico-semánticas mantiene?

La palabra "interrogar",

¿qué relaciones sintáctico-semánticas mantiene?

- → URL (25): http://fenix.cnice.mec.es/diccionario/l/l1211.html
- → URL (26): http://www.iula.upf.edu/rec/daele/ ("cuestionar")
- → URL (27): http://multiwordnet.fbk.eu/online/multiwordnet.php
- → URL (28): http://sato.fm.senshu-u.ac.jp/frameSQL/sfn20/notes/index.html
 - → URL (29): http://adimen.si.ehu.es/cgi-bin/wei/public/wei.consult.perl
 - → URL (30): http://adesse.uvigo.es/data/verbos.php?sense=2117



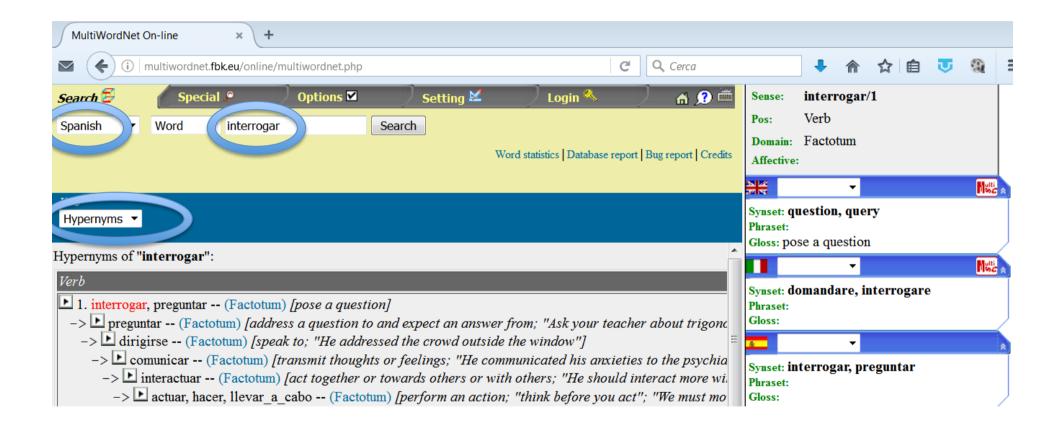
corpus "lematizados"

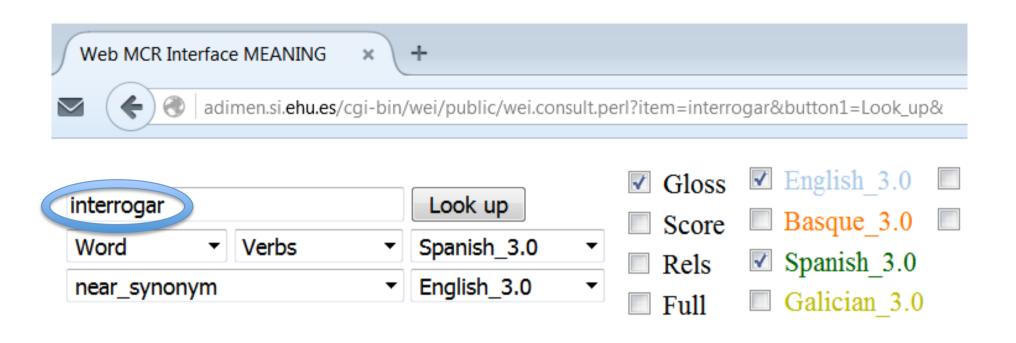
etiquetas semánticas-sintácticas

estructura argumental

ling. cognitiva (frames)

ling. computacional





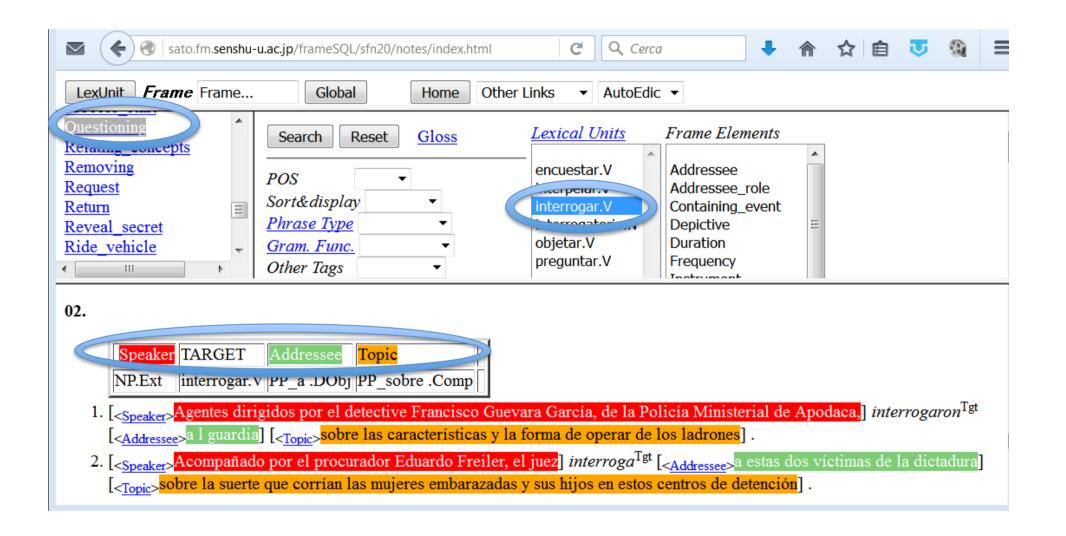
Multilingual Central Repository (ILI 3.0) - WikiMCR

```
ili-30-00729378-v

factotum
cognition
IntentionalPsychologicalProcess

Agentive+
Communication+
Dynamic+
Social+
UnboundedEvent+

Ilitertogar -1 ponderar -1 preguntar -1
```

















[-] Clasificación semántica y potencial valencial

Tipo de proceso: Comunicación

Argumentos:

| | | | | Fre | ecuencia |
|-----------|------|---------------|-------------|-----|----------|
| A1 | COMR | Interrogador | COMUNICADOR | 26 | (96.3 %) |
| A2 | MENS | Interrogación | MENSAJE | 4 | (14.8 %) |
| АЗ | REC | Interrogado | RECEPTOR | 23 | (85.2 %) |
| A4 | ASU | | ASUNTO | 4 | (14.8 %) |
| | | | | | |

Perfil combinatorio >>

[+] Realizaciones valenciales (Esquemas sintáctico-semánticos):

[-] Ejemplos:

(solo 1000 scogidos aleatoriamente, de un total de 27)

| √erbo | 1 | Texto Texto | Referencia | |
|------------|---------|---|--------------|-------|
| INTERROGAR | | n máximo. Necesito saberEntonces voy a hacerle algunas preguntas. <i>El profesor interroga</i> ticulosamente al viejo sobre sus sensaciones, sus dolores, su reacción a ciertas | [SON:230.13] | +info |
| INTERROGAR | - ti | erá lo mejor. Buenas noches, padre. <i>El viejo, camino de su cuarto,</i> se <u>interroga</u> : «En otros npos me hubiese peleado con mi | [SON:282.36] | +info |
| INTERROGAR | 10 | mirar mejor tanta recia ternura, tan hondo amor viril, misteriosamente encarnado en la dra. <u>Interroga</u> de igual a igual porque, si él hubiera cogido un cincel alguna vez, así se ibiera enfrentado | [SON:094.06] | +info |
| WTERROGAP | | uvo afición por las entradas teatrales, "je sais faire <i>mes entrées" sentenciaba, pero ahora</i> niraba sus pies, <u>interrogando</u> los. La fijeza de su actitud me estremeció. Los | [DIE:150.02] | +info |











[-] Realizaciones valenciales (Esquemas sintáctico-semánticos):

| Voz | Argumentos s | semánticos y | Funcion | es sintácticas | N_ejemplos |
|-------------------------------------|------------------|------------------------|---------------------|--|------------|
| INTERROGAR _{act} | A1:COMR = SUJ | A3:REC =ODIR | | | 15 |
| INTERROGAR _{act} | A1:COMR = SUJ | A2 :MENS = Cita | | | 3 |
| INTERROGAR _{act} | A1:COMR = SUJ | A3:REC =ODIR | | A4 : <i>ASU</i> = sobre OBL | 3 |
| INTERROGARact | A1:COMR = SUJ | | | | 1 |
| INTERROGARSE _{ref} | A1:COMR = SUJ | A2:MENS = Cita | | | 1 |
| INTERROGARSE _{ref} | A1:COMR = SUJ | A3:REC = Refl | | A4:ASU =acerca de OBL | 1 |
| (ser) INTERROGADO _{pasiva} | A3:REC = SUJ | | A1:COMR =por AGT | | 2 |
| (ser) INTERROGADO _{pasiva} | A3:REC = SUJ | | | | 1 |

Materiales y actividades

basados en corpus de nativos

Ejercicios con concordancias

[HONEYFIELD (1989), McCULLOUGH, 2001, p. 136]

- 1. Llenar espacios en blanco en concordancias, i.e., adivinar la palabra clave.
- 2. Completar el contexto más amplio de las líneas de concordancias, i.e., adivinar qué palabras precedían y/o seguían a las que aparecen en el papel. Adivinar la totalidad de la frase que contiene la palabra clave.
- 3. Usar líneas de concordancias como base de datos para diferentes actividades centradas en la gramática, el vocabulario, o el uso de la lengua.
- 4. Ejercicios de análisis del discurso: por ejemplo, el análisis del uso de marcadores del discurso a través de líneas de concordancias.
- 5. Comparación de sentidos o usos de expresiones en diferentes tipos de textos.
- 6. Análisis del tono o estilo de diferentes textos. Por ejemplo, comparar el tono de diferentes palabras o expresiones clave en diferentes géneros del textos.
- 7. Utilización libre del programa de concordancias como ayuda a la escritura, la corrección de redacciones o la comprensión de textos.

- → URL (31): https://eric.ed.gov/?id=EJ468954
- → URL (32): https://dialnet.unirioja.es/servlet/articulo?codigo=2095396

Ejercicios con concordancias (DDL)



4.2 Activities and worksheets

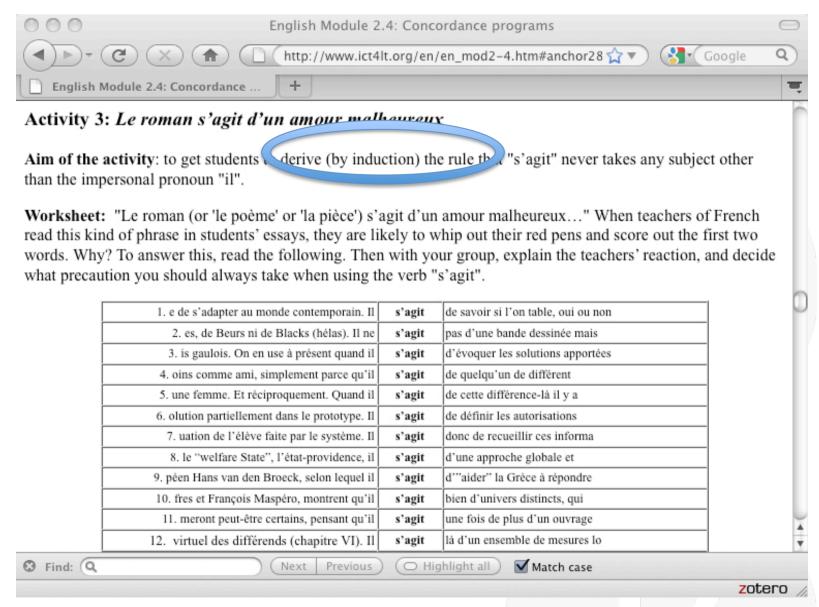
| ACTIVITY NAME AND LANGUAGE | PURPOSE OF ACTIVITY | PRESENTATION |
|--|--|-----------------------|
| | | |
| 1. Guess the mystery word (F) | beginners, lexis | paper,online (LAN) |
| 2. Donc on peut dire que (F) | style, usage | paper, online (LAN) |
| 3. S'agit (F) | derive a rule, grammar | paper, online (LAN) |
| 4. Beware false friends (F) | lexis | paper, online (LAN) |
| 5. Coffee or tea? (F) | cultural differences | paper, online (LAN) |
| 6. Les Anglais et les Britanniques (F) | political correct, usage, lexis | paper, online (LAN) |
| 7. Changing lifestyles (F) | cultural differences, usage | paper, online (LAN) |
| 8. dürfen and müssen (G) | lexis, usage | paper |
| 9. Preposition am (G) | usage | paper |
| 10. Syntax of adverbs (E) | driving a rule, grammar, syntax, statistics, usage | online (WWW) |
| 11. reason + because (E) | correctness, usage | online (WWW) |
| 12. Students' own writing (E) | variety, usage | online, standalone PC |
| 13. L'aigle noir (F) | literary analysis | online (LAN) |
| 14. Mariana (E) | literary analysis | online (LAN or WWW) |

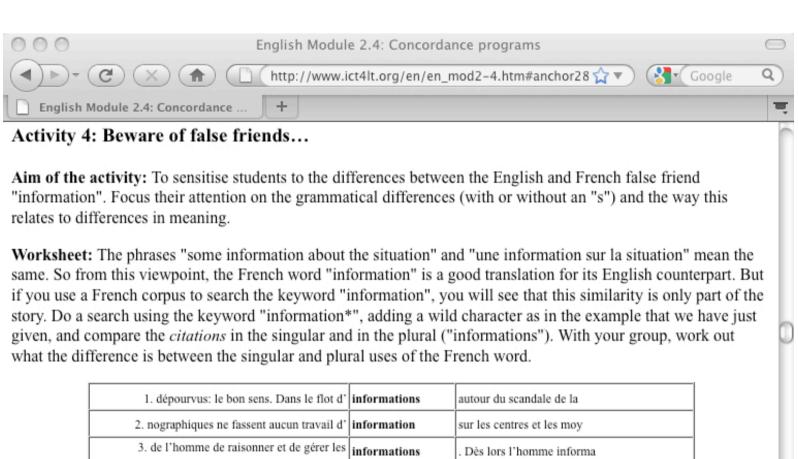
LAMY M-N. y KLARSKOV MORTENSEN H. J. (2009): «Activities and worksheets»

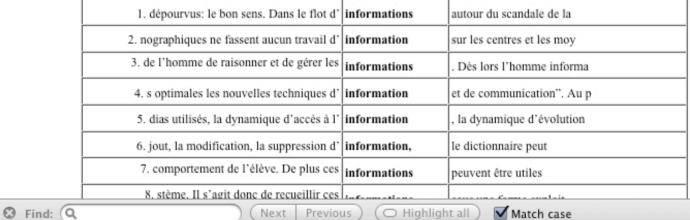
URL (33): http://www.ict4lt.org/en/en_mod2-4.htm#sec4.2



Ejercicios con concordancias (DDL)

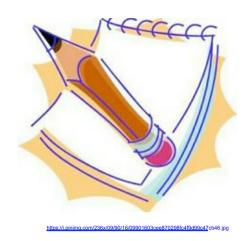






zotero





Un ejercicio o actividad para el aula de ELE con corpus

- Con qué corpus
- Para qué nivel (MCER)
- Qué objetivos
- Qué formación necesita el docente
- Para realizar con los alumnos / Para que el docente prepare materiales
- Dinámica del ejercicio

109

72

54

50

41

36

34

33

33

31

30

28

27

24

23

→ URL (34): http://tagpacker.com/user/mar.cruz.pinol?t=corpus

U tagpacker

TAGS

corpus

lengua_espanola

concordancias

sonoros_y_o_orales_transc 62;

de_aprendices

aplicacionesELE

textos_literarios 45

ELE_EL2

paralelos_y_multilingues 41

frecuencias lexicas 37

aplicaciones_sociolinquisti 36

web_como_corpus

lematizado

software

tecnologias_linguisticas

lexico

especializado

cronologicos_o_diacronico: 28

terminologia

dialectologia

aplicaciones_lexicografia_l 25 o

literatura

colocaciones

aplicaciones_traduccion

21

21

lenguas_de_europa

marcaje_cat_gramatical 19

medios_comunicacion 19

19 recursos

bibliografia 18

diccionarios 16

flexionador 15

marcaje_semantica 15

en_papel_o_CD 14

13 escritura

ingles_comoL2 12

12 morfologia

Corpus textuales de nativos y enseñanza/aprendizaje del español como LE/L2

Mar Cruz Piñol, Universitat de Barcelona

http://sites.google.com/site/marcruzpinolub/





Grupo de investigación acqUA aguisidan y aprendizaj e de segundos lengo as y lengoas extranjesas de la Universidad de Alicant

Seminarios de investigación 2018