

Corpus para ELE

**Claudia Sánchez-Gutiérrez, Sam Davidson, Sophia
Minnillo**

University of California, Davis
28/3/2023

¿Qué es un *corpus*?


- Una **colección electrónica** de datos **naturales** (o cuasi-naturales) producidos por hablantes *nativos* o *no nativos* de una lengua (o dialecto) y organizados según **criterios de diseño explícitos**

Los corpus de nativos

Corpus de la RAE

Hasta 1975

1975 - 2004

 **REAL ACADEMIA ESPAÑOLA**


Real Academia Española - Corpus Diacrónico del Español (CORDE)

Consulta:

Criterios de selección:

Autor:	<input type="text"/>	Obra:	<input type="text"/>
Cronológico:	<input type="text"/>	Medio:	<div>(Todos) Libros Periódicos Revistas Miscelánea Orales</div>
	<input type="text"/>	Geográfico:	<div>(Todos) Argentina Bolivia Chile Colombia Costa Rica</div>
Tema:	<div>(Todos) 11.- Lírica 12.- Narrativa 121.- Breve 1211.- Relato breve tradicional 1212.- Relato breve culto</div>		

Buscar Limpiar

 **REAL ACADEMIA ESPAÑOLA**

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

Consulta:

Criterios de selección:

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

Buscar Limpiar

¿Cuál es la diferencia entre un lema y una forma?

2001 - 2022

Corpus del Español del Siglo XXI
(CORPES)

Versión beta (0.99)



Concordancias | Coapariciones | Configuración | Ayuda | Estadística | Modo de cita | Sugerencias | Preguntas frecuentes

Lema Forma Clase de palabra

☐ Grafía original

Corpus del Español

Davies, 2006



Corpus del Español: Web/Dialects



SEARCH

FREQUENCY

CONTEXT

OVERVIEW

FIND SAMPLE: [100](#) [200](#) [500](#) [1000](#)

PAGE: << < 1 / 617 > >>



CLICK FOR MORE CONTEXT

HELP



SAVE



TRANSLATE



ANALYZE



1	G AR	1foto1historia.wordpress.com	⌂	🔍	Q	la heroína de los siglos XX y XXI. Muchos cariños desde San Juan, ojalá algún día pueda llegar a conocer su pueblo y su gente y si puedo colaborar
2	G AR	442.perfil.com	⌂	🔍	Q	de la historia de Boca. Carlos (Bianchi) verá cuándo estará disponible y ojalá que pueda aportar lo mismo que hizo en los años anteriores, dijo Orion en
4	G AR	442.perfil.com (1)	⌂	🔍	Q	emociona el cariño que nos brinda la gente y hay que seguir así y ojalá que Newell's lo haga por muchos años porque se van a ver buenos espectáculos
5	G AR	442.perfil.com	⌂	🔍	Q	, falleció a fines del año 95). Les deseo lo mejor, ojalá vuelvan enseguida a Primera y que no lo tomen como un campeonato cualquiera; es
6	G AR	442.perfil.com	⌂	🔍	Q	volveré a su rincón. Siempre le prometí a Sergio terminar la carrera juntos. Ojalá nos dejen... 3 Comentarios LA ingratitud es la peor consejera de una
7	G AR	442.perfil.com	⌂	🔍	Q	la posibilidad de escupir le en la cara o decir le lo que pienso. Ojalá que le vaya lo mejor posible, que saque campeón a Boca, que sea
8	G AR	442.perfil.com	⌂	🔍	Q	, falleció a fines del año 95). Les deseo lo mejor, ojalá vuelvan enseguida a Primera y que no lo tomen como un campeonato cualquiera; es
11	G AR	442.perfil.com (2)	⌂	🔍	Q	en noviembre - 21 en Córdoba y 23 en Buenos Aires -. Y ojalá que antes también para algunas otras cosas, dijo. Sobre la Davis, el
12	G AR	90mas10.com	⌂	🔍	Q	o cinco días, lo que sumado a los días dedicables al Salone (ojalá , mínimo dos), garantiza una semana entera viviendo Milán. Además, cada
13	G AR	acaestalahinchada.blogspot.com	⌂	🔍	Q	gane, en la tribuna no habrá desmadre. Por el bien de Racing, ojalá así sea. La Cámara Penal incrementó una pena contra Chino Fleitas por amenazas y
14	G AR	acaestalahinchada.blogspot.com	⌂	🔍	Q	, los robos, como existe la violencia en la sociedad. Diego Maradona DT Ojalá que algún día Diego Maradona sea el entrenador porque es uno de los gra

Corpus especializados en variedades americanas



Corpus Diacrónico y Diatópico
del Español de América

Corpus orales



Como profesores de español, ¿cómo podéis utilizar los corpus?

- Para enseñar gramática
- Para enseñar vocabulario
- Para enseñar pragmática

¿Qué pueden aportar los corpus a las clases de ELE?

Data-driven learning (DDL) (aprendizaje basado en datos)

- Los **aprendices**

- pueden consultar textos, grabaciones, y videos de interacciones entre hablantes nativos (*modelos auténticos del uso*)
- pueden buscar palabras, expresiones, y elementos gramaticales para aprender sobre el uso de estos elementos (*descubrir tendencias*)



CLICK FOR MORE CONTEXT

HELP



SAVE



TRANSLATE



ANALYZE



1	G AR	1foto1historia.wordpress.com	🔍	🔍	Q	la heroína de los siglos XX y XXI. Muchos cariños desde San Juan, ojalá algún día pueda llegar a conocer su pueblo y su gente y si puedo colaborar
2	G AR	442.perfil.com	🔍	🔍	Q	de la historia de Boca. Carlos (Bianchi) verá cuándo estará disponible y ojalá que pueda aportar lo mismo que hizo en los años anteriores, dijo Orion en
4	G AR	442.perfil.com (1)	🔍	🔍	Q	emociona el cariño que nos brinda la gente y hay que seguir así y ojalá que Newell's lo haga por muchos años porque se van a ver buenos espectáculos
5	G AR	442.perfil.com	🔍	🔍	Q	, falleció a fines del año 95). Les deseo lo mejor, ojalá vuelvan enseguida a Primera y que no lo tomen como un campeonato cualquiera; es

¿Qué pueden aportar los corpus a las clases de ELE?

- Los **profes**
 - pueden basar lecciones y actividades sobre los ejemplos del uso en corpus

[illegible]

Por ejemplo

(Vacas Matos, 2014)

COR.E.M.A.H.

Corpus Español Multimodal de Actos de Habla



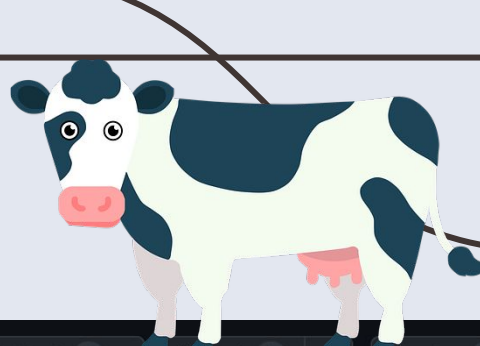
Los corpus de aprendices de ELE

¿Qué es un *corpus de aprendices*?

- Una **colección electrónica** de datos **naturales** (o cuasi-naturales) producidos por hablantes *nativos* o **no nativos** de una lengua (o dialecto) y organizados según **criterios de diseño explícitos**

Tipos de corpus de aprendices de ELE

Mira la descripción de los diferentes corpus en este listado (Rojo et al., 2022) y responde las preguntas de la **actividad 4** de la hoja de ejercicios.



Los corpus de aprendices: COWS-L2H

Yamada et al., 2020

ucdaviscl / cowsl2h Public

Watch 5

Fork 3

Star

<> Code Issues 2 Pull requests Actions Projects Security Insights

master cowsl2h / special / F18 / essays / 102336.txt

Go to file

...



ssdavidson Add 2019 anonymized essays

Latest commit f30e63b on Dec 31, 2019

History

1 contributor

3 lines (2 sloc) 1.36 KB

Raw

Blame



```
1 Me encanta me mama, porque ella es me mejor amiga. Ella es un professora por la universidad de *UNIVERSITY*. Me encanta me mama porque ella es muy simpatica, gr
2
3
```

COWS-L2H: niveles y evolución

Utiliza los textos de cada nivel para organizar tipos de errores y determinar cuáles tardan más en desaparecer.

Ejercicio 5 de la hoja de actividades.

Los corpus de aprendices: CEDEL2

Lozano, 2022




Sept. 2020

CEDEL2: Corpus Escrito del Español L2 (version 2)

 [CEDEL2](#)

 [Search / !\[\]\(a86c7d1c9cb81c81614634a31267440d_img.jpg\) Download](#) ▾

 [User guide](#) ▾

 [Statistics](#)

 [About](#) ▾

 [Contact](#)

Filename: AR_WR_12_26_0.5_14_EH

Placement test score (%): 27.9

Years studying Spanish: 0.5

L1: Arabic

Proficiency: Lower beginner

Stay abroad (months): 0

Age: 26

Proficiency (self-assessment): 1 / 6

Task title: 14. Chaplin

Placement test score (raw): 12 / 43

Age of exposure to Spanish: 25

Medium: Written

Cuando Chaplin estaba caminando, vio a un niño al lado de uno de los contenedores de basura y quiso buscar a su madre. Cuando vio a una mujer caminando en su automóvil con un niño con él, se dio cuenta de que era su hijo, pero lo giró y le pidió que dejara al niño árabe. Para un hombre pobre y huyó, el hombre siguió buscándolo, pero no lo encontró, por lo que lo colocó en su automóvil, la mujer con un niño, y cuando Chaplin se aseguró de que estaba a salvo, apareció de nuevo, pero por su mala suerte, la mujer lo volvió a ver y lo castigó, y la policía le informó sobre él.

CEDEL2: oral vs. escrito

Utiliza los textos escritos vs. orales de los estudiantes para determinar qué errores son más típicos de los textos orales y cuáles de los escritos. ¿Qué diferencias observáis entre ambas modalidades?

Ejercicio 6 de la hoja de actividades.



En el aula

¿Qué usos les podemos dar a los CAE de cara a mejorar nuestras clases? (actividad 7)

1. Creación de materiales (Granger, 2015)
 - Selección: ¿Qué hay que enseñar?
 - Descripción: ¿Qué explicaciones van a ser más útiles para los estudiantes?
 - Secuenciación: ¿Cuándo hay que enseñar cada contenido gramatical?

Ejemplos de materiales basados en CAE

Diccionarios de aprendices

- Inclusión de palabras que son menos usadas por los estudiantes pero son muy frecuentes.
- Inclusión de notas de uso basadas en errores observados en CAE

! Don't say 'it reminds me her'. Say **it reminds me of her**.

! Don't use **remind** (= make you remember) when you mean **remember**.

Figure 22.1. Help box for the verb *remind* in the *Longman Essential Activator* (1997)

Ejemplos de materiales basados en CAE

Get it right: used to

Don't confuse

- **I am used to doing something**
- **I used to do something**

If you **are used to doing something**, it is familiar to you because you have often done it before. Use the **-ing** form of the verb in this pattern, not the infinitive:

☐ *Looking through the newspapers every day, we ~~are used to read~~ terrifying reports about crimes.*

☒ *Looking through the newspapers every day, we are used to reading terrifying reports about crimes.*

You can also say **get used to doing something**:

*Children soon **get used to spending** much of their free time watching TV.*

If you say that you **used to do something**, you are talking about an activity or habit in the past which has now finished.

***I used to play** squash reasonably well.*

This means that the speaker no longer plays squash well.

Figure 22.2. 'Get it right' box for *used to* in MEDAL (2007)

Ejemplos de materiales basados en CAE

Gramáticas pedagógicas

- Inclusión de notas de uso basadas en errores observados en CAE

I can hear music.

~~(I can hear a music.)~~

The furniture was very old.

~~(The furnitures were very old.)~~

They got good advice from the student counsellor.

~~(They got good advices from the student counsellor.)~~

Figure 22.3: Error warning for non-count nouns in the *CGE* (2006: 339)

Ejemplos de materiales basados en CAE

TREACLE project (O'Donnell, 2012):

- Análisis de errores en diferentes niveles de una misma universidad.
- Análisis del desarrollo gramatical correcto en la misma universidad.
- Los resultados indican:
 - A qué contenidos gramaticales hay que darles prioridad en diferentes niveles.
 - Qué tipos de errores requieren más atención sostenida y feedback a lo largo del tiempo.

¿Qué usos les podemos dar a los CAE de cara a mejorar nuestras clases?

2. Placement tests (Bowles, 2022)

- A través de un corpus de hablantes de herencia y otro de aprendices de L2, se establecen listas de palabras específicas de cada grupo.
- A partir de esas listas, se creó un test para diferenciar rápidamente a hablantes de herencia y de L2 para inscribirlos en los cursos más adaptados a sus necesidades.

¿Qué usos les podemos dar a los CAE de cara a mejorar nuestras clases?

3. Formación de profesores:

- *Spanish Corpus Proficiency Levels Training* (Koike and Witte, 2016)

The screenshot shows the 'Spanish Corpus Proficiency Level Training' (SPT) website. The header is green with the 'SPT' logo and the title. Below the header are navigation tabs: 'Introduction', 'about', 'proficiency levels: practice' (which is selected), 'video activity', and 'glossary and links'. A paragraph of text explains the purpose of the site: 'What level is the Spanish Learner? beginner, beginner-intermediate, intermediate, intermediate-advanced, or advanced? Select a learner and explore each topic they talk about. Answering the questions for each video will help you determine at what level the learner is speaking Spanish.'

Below the text are two columns of lists:

- Learners:** Abigail, Alexi, Allison (highlighted), Amy, Anthony, Brenda, Brian, Cindy, Claire, Debra, Elizabeth, Emily, Grace, Hannah, Heather, Jacob, James, Jenny, Jessica.
- Topics:** Family, Hometown, Comparisons, Major, Other Countries (highlighted), Sports, Apartment, Politics, Past.

To the right of these lists is a video player showing a woman speaking. The video has a progress bar at 0:18 / 2:43 and a 'View Transcript' link below it.

Spanish Corpus Proficiency Levels Training

grammar accuracy

vocabulary

fluency

conversation management

generalizations

evaluate

1. **In what areas did this learner show fairly accurate grammatical control?**

- a. ☐ Subject-verb agreement
- b. ☐ Gender-number agreement
- c. ☐ Subjunctive Use
- d. ☐ Verb conjugations
- e. ☐ None of the above

2. **Generally, what were the main areas in which this learner showed some grammatical inaccuracies?**

- a. ☐ Subject-verb agreement
- b. ☐ Noun-adjective agreement
- c. ☐ Subjunctive Use
- d. ☐ Verb conjugations
- e. ☐ None of the above

3. **Which errors were most glaring in terms of overall understanding of the learner's response?**

- a. ☐ Subject-verb agreement
- b. ☐ Noun-adjective agreement
- c. ☐ Subjunctive Use
- d. ☐ Verb conjugations
- e. ☐ None of the above

4. **How did the learner often correct his/her grammar mistakes, if at all?**

- a. ☐ Self-correction without asking for help
- b. ☐ Self-correction after requesting help
- c. ☐ By repeating what the teacher asked him/her to say.
- d. ☐ No correction

Score

Try Again

Spanish Corpus Proficiency Levels Training

grammar accuracy

vocabulary

fluency

conversation management

generalizations

evaluate

1. Did the learner use any words or expressions that seemed above his/her level?

- a. ☐ Yes
- b. ☐ No

2. What are three words and/or expressions that the learner lacked or that the learner did not understand that interfered with the comprehension, if any?

3. How did the learner usually correct his/her vocabulary mistakes?

- a. ☐ Self-correction without asking for help
- b. ☐ Self-correction after requesting help
- c. ☐ By repeating what the teacher asked him/her to say.
- d. ☐ No correction

4. Did the learner seem to use words consistently and with confidence dealing with:

- a. ☐ Basic objects, places, and kinship terms
- b. ☐ Uncomplicated tasks and social situations like personal history and leisure activities
- c. ☐ A wide range of tasks, including those of school and work-related situations, including elaborating, and hypothesizing

Score

Try Again

Spanish Corpus Proficiency Levels Training

grammar accuracy

vocabulary

fluency

conversation management

generalizations

evaluate

After watching at least 4 videos of varying difficulty, rate the learner according to the guidelines listed below.

Beginner

Communicates minimally using learned material using isolated words, short phrases and frequent long pauses. May repeat interlocutor's words. Communicates simple, elementary needs and basic courtesies.

Beginner-Intermediate

Can partially fulfill basic communicative exchanges, using learned utterances, sometimes expanded through simple recombinations elements, with frequent errors. Can ask questions/make statements with learned material, showing signs of spontaneity.

Intermediate

Can create language by combining learned language, mainly in reactions. Can communicate successfully in a limited number of uncomplicated task-based and social situations: ask/answer questions, initiate/respond to simple statements, and maintain a limited conversation, with frequent errors and long pauses.

Intermediate-Advanced

Can communicate successfully in most uncomplicated tasks and social situations. Can initiate, sustain and close a general conversation with strategies appropriate to a range of circumstances and topics, but with errors. Hesitation due to limited vocabulary but there is emerging evidence of connected discourse.

Advanced

Can converse fairly easily, in a wide variety of everyday communicative tasks. Can handle complicated tasks and social situations, although not with ease. Can narrate and describe with some details, and can communicate facts about topics of current public and personal interest, using general vocabulary.

¿Qué usos les podemos dar a los CAE de cara a mejorar nuestras clases?

4. Automatic error tagging and tracking

- With parallel corrected and corrected essays, you can align sentences and detect student errors on a large scale.
 - Correcting sentences is very time consuming, so few corpora have this type of data.
- Automated tools allow you to tag errors based on part of speech and other grammatical categories.
- This allows you to check if students are implementing what is being taught in the classroom and if they are doing it correctly.

Ejemplo de etiquetado de errores

Original:	Stephen King escribió muchos libros .									
Corrected:	Stephen King ha escrito muchos libros .									
	A	2	2		M:VERB:TENSE		ha		REQUIRED	-NONE- 0
	A	2	3		R:VERB		escrito		REQUIRED	-NONE- 0

¿Qué usos les podemos dar a los CAE de cara a mejorar nuestras clases?

5. Automatic error correction

- You can also use parallel data to train machine learning tools (such as sequence-to-sequence neural networks) to automatically correct student errors.
- These tools are still in development, but they can allow immediate student feedback and reduce instructor time manually correcting student writing.



¡Gracias!