

JOURNÉE D'ÉTUDES

CONSTITUTION DE CORPUS ÉCRITS D'APPRENANTS
EN CONTEXTE UNIVERSITAIRE:
PRATIQUES, ENJEUX ET PERSPECTIVES ACQUISITIONNELLES

LIDILEM, Université Grenoble-Alpes

18 mars 2022

<https://jeca2022.sciencesconf.org/>



Key criteria for the design and compilation of a corpus for L2 acquisition research: Showcasing CEDEL₂ (a multi-L₁ corpus of L2 Spanish)



UNIVERSIDAD
DE GRANADA



A_iNACØREX

PROYECTO I+D+i (PID2020-113818GB-I00)

To know more

- **CEDEL2 corpus (version 2): Design and web interface**
 - Lozano, C. (2021). CEDEL2: Design, compilation and web interface of an online corpus for L2 Spanish acquisition research. *Second Language Research*, first online view. <https://doi.org/10.1177/02676583211050522>
- **Learner corpus design principles:**
 - Lozano, C., & Mendikoetxea, A. (2013). Learner corpora and second language acquisition: The design and collection of CEDEL2. In A. Díaz-Negrillo, N. Ballier, & P. Thompson (Eds.), *Automatic Treatment and Analysis of Learner Corpus Data*. (pp. 65-100). John Benjamins. <https://doi.org/10.1075/scl.59.06loz>
- **Using L2 Spanish corpora:**
 - Lozano, C. (2021). Corpus textuales de aprendices para investigar sobre la adquisición del español LE/L2. En M. Cruz Piñol (Ed.), *E-Research y español LE/L2: Investigar en la era digital* (pp. 138-163). Routledge. <http://doi.org/10.4324/9780429433528-9>
- **Using L2 corpora to test SLA hypotheses:**
 - Lozano, C. (2021). Generative approaches. In N. Tracy-Ventura & M. Paquot (Eds.), *The Routledge Handbook of Second Language Acquisition and Corpora* (pp. 213-227). Routledge. <https://doi.org/10.4324/9781351137904>
- **Using L2 Spanish corpora for vocabulary acquisition:**
 - Lozano, C. (2015). Learner corpora as a research tool for the investigation of lexical competence in L2 Spanish. *Journal of Spanish Language Teaching*, 2(2), 180-193. <http://dx.doi.org/10.1080/23247797.2015.1104035>
- **Triangulation: Combining corpus and experimental approaches:**
 - Lozano, C. (2020). How to do research in SLA by combining corpus and experimental methods: An introduction for language researchers and teachers. In M. Planelles Almeida, A. Foucart, & J. M. Liceras (Eds.), *Perspectivas actuales en la enseñanza y el aprendizaje de lenguas en contextos multiculturales / Current Perspectives in Language Teaching and Learning in Multicultural Contexts* (pp. 389-421). Thomson Reuters Aranzadi. <https://doi.org/10.32029/2605-4655.11.01.2020>

Outline

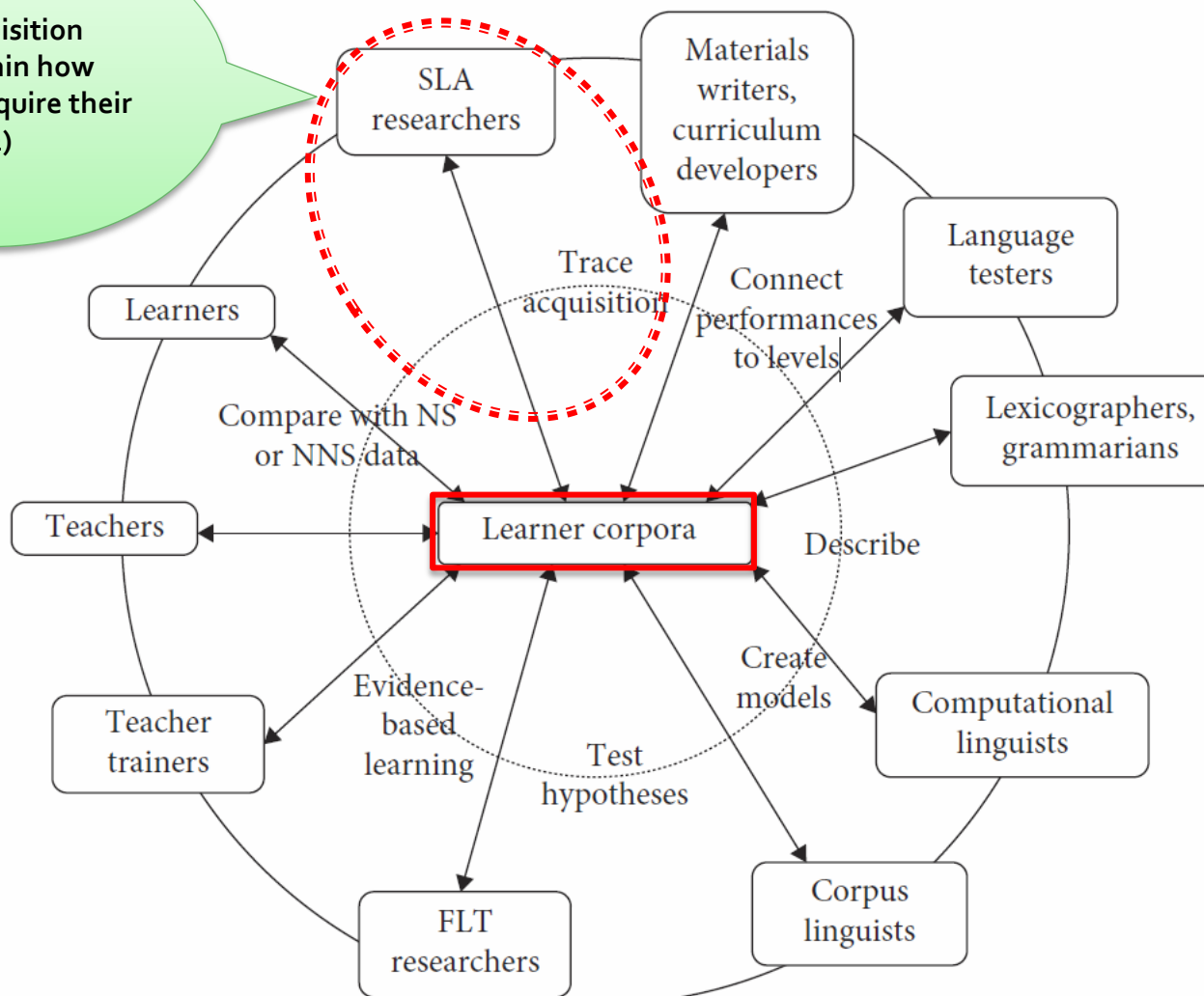
- Aim
 - To discuss the importance of SLA-informed learner corpus design by showcasing CEDEL2 (version 2).

- Outline
 - Contextualising learner corpora in SLA
 - L2 Spanish corpora
 - CEDEL2
 - Intro
 - Design and structure recommendations
 - Data collection recommendations
 - Stats
 - Call for participation
 - CEDEL2: Web interface
 - Final recommendations: Good corpus design + triangulation in SLA

L2 corpora: users & activities

Díaz-Negrillo & Thompson 2013:17)

L2 acquisition
(to explain how
learners acquire their
L2)



Learner corpora: Definition

- **Callies & Paquot (2015:1)**

"**systematic** collections of **authentic, continuous** and **contextualized** language use (spoken or written) by L2 learners stored in **electronic format**"

- Important (but missing) **ingredient** in these definitions:

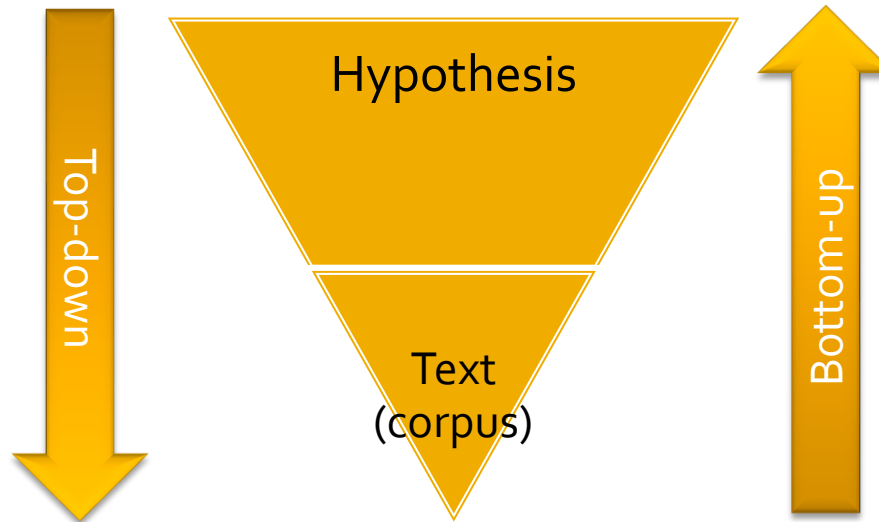
- To be maximally usable by L2 researchers, LC should contain precise and detailed **linguistic-background information** on each learner/native, with **SLA-relevant variables** → We will discuss these today.
- So, a well-designed collection of texts without individual ling-background info **may not be informative enough for SLA purposes.**

Callies, M., & Paquot, M. (2015). Learner Corpus Research: An interdisciplinary field on the move. *International Journal of Learner Corpus Research*, 1(1), 1-6.

2 approaches in LCR

1. Hypothesis-driven

- Testing a specific hypothesis in the corpus → rather **recent** approach.



2. Hypothesis-finding

- Exploring patterns in the corpus to generate hypotheses → **traditional** approach in LCR

But, actually, both approaches are **complementary**.

Limitations of LCR in SLA

- LIMITATION 1: “LCR is not always fully informed by SLA research” (Myles 2015:309)
- LIMITATION 2: the contribution of LCR so far has been much more substantial in **description** than interpretation of SLA data (Granger 2004, 2009 Myles 2005, 2015).
- LIMITATION 3: The majority of LCR studies are **hypothesis-finding** studies “because learner corpus research has been mainly conducted by corpus linguists, rather than SLA specialists”
(Lozano & Mendikoetxea 2013:69. See also Granger 2004: 134-135 Mendikoetxea 2014, Myles 2015).

Some L2 Spa corpora

- **L2 Spanish**: not many corpora available yet DESPITE growing interest in L2 Spanish research, particularly in USA
 - Lafford & Salaberry 2003, Montrul 2004, 2013, Pérez-Leroux & Licerias 2002.
- **SPLLOC** (Spanish Learner Language Oral Corpus) → ORAL (Mitchell et al. 2008).
- **LANGSNAP** (Languages and Social Networks Abroad Project) → ORAL & SPOKEN (Tracy-Ventura et. al. 2015)
- **CORELE** (Corpus Oral del Español como Lengua Extranjera) → ORAL (Campillos-Llanos 2014)
- **CAES** (Corpus de Aprendices de Español) → WRITTEN (Rojo & Palacios 2014)
- **Some limitations**
 - No standardised placement test (SPLLOC)
 - No comparable Spanish native control corpus (SPLLOC, CORELE, CAES)
 - Too many L1s, so small number of words per L1:
 - And then fewer words as per proficiency level (often, just a few thousands).



The CEDEL2 corpus: intro

R&D Projects:

ANACOR (2017-2021): FFI2016-75106-P

ANACOREX (2021-2025): PID2020-113818GB-I00

■ Aim:

- ANACOREX [webpage](#).
- To investigate L2 acquisition
 - L2 Spanish
 - L2 English
 - L2 Japanese (in progress)
- Anaphora Resolution



UNIVERSIDAD
DE GRANADA

ANACOREX
PROYECTO I+D+i (PID2020-113818GB-I00)



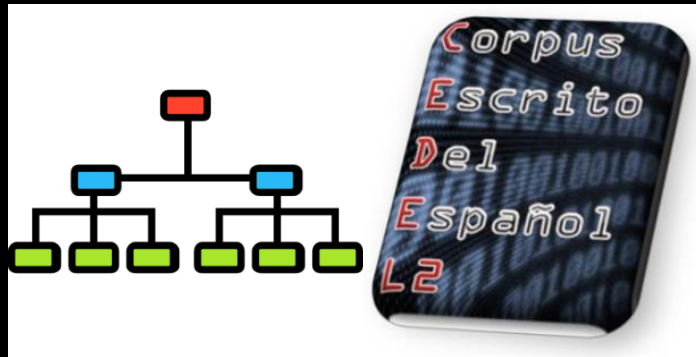
■ Research methods:

■ Learner (L2) corpora:

- CEDEL2 (Corpus Escrito del Español como L2)
- COREFL (Corpus of English as a Foreign Language)
- JFLCorp (Japanese Foreign Language Corpus)

■ Experiments:

- Acceptability Judgement Tasks (AJT)
- Self-Paced Reading Task (SPRT) with reaction times (RT)
- Eye tracking (ET)



CEDEL2: DESIGN & STRUCTURE

- Lozano, C. (2009). [CEDEL2: Corpus Escrito del Español L2](#). In: Bretones Callejas, Carmen M. et al. (eds) *Applied Linguistics Now: Understanding Language and Mind*. Almería: Universidad de Almería, pp. 197-212.
- Lozano, C., & Mendikoetxea, A. (2013). [Learner corpora and second language acquisition: the design and collection of CEDEL2](#). In N. Ballier, A. Díaz-Negrillo, & P. Thompson (Eds.), *Automatic Treatment and Analysis of Learner Corpus Data*. Amsterdam: John Benjamins.
- Lozano, C. (2021). CEDEL2: Design, compilation and web interface of an online corpus for L2 Spanish acquisition research. *Second Language Research*, first online view. <https://doi.org/10.1177/02676583211050522>

CEDEL2 (version 2)

■ CEDEL2: multi-L1 corpus of L2 Spanish

- L1 English – L2 Spanish [Lozano, UGR]
- L1 Japanese – L2 Spanish [López-Sako, UGR]
- L1 Chinese – L2 Spanish [Ciruela, UGR]
- L1 Arabic – L2 Spanish [Haddad, UGR]
- L1 German – L2 Spanish [Torregrossa, Frankfurt]
- L1 Dutch – L2 Spanish [Collewaert, Brussel]
- L1 French – L2 Spanish [Lacroix, Montreal]
- L1 Italian – L2 Spanish [Montserrat, Florence]
- L1 Portuguese – L2 Spanish [Teixeira, N. Lisboa]
- L1 Greek – L2 Spanish [Georgopoulos, Athens]
- L1 Russian – L2 Spanish [Portnova & Barros, UGR]
- L1 Turkish – L2 Spanish (in prep) [Ozäydin, UGR]
- L1 Estonian – L2 Spanish (in prep) [Rapún-Mombiela & Kruse, Tartu]



UNIVERSIDAD
DE GRANADA



■ Native control corpora:

- L1 Spanish
- L1 English
- L1 Japanese
- L1 German
- L1 Portuguese
- L1 Arabic
- L1 Turkish (in prep)

Lozano, C. (2021). CEDEL2: Design, compilation and web interface of an online corpus for L2 Spanish acquisition research. *Second Language Research*, first online view. <https://doi.org/10.1177/02676583211050522>

L2 corpus design principles

- It has long been recognised that LC need to be designed according to some principles / criteria.
- But most LC fail to follow **strict design criteria**, they simply follow 'ad hoc' design criteria.
- In the design of LC, we need to take into account:
 - **Learner** variables
 - **Task** variables
 - Also: corpus design principles (Lozano & Mendikoetxea 2013).

L2 corpus design principles



- According to **Sinclair (2005)**, there are **10 principles** for a well-designed corpus
- These have been **applied to L2 corpora** (Lozano & Mendikoetxea 2013):
 - **content selection, representativeness**, contrast, structural criteria, annotation, **sample size, documentation**, balance, topic, homogeneity.
- **Gilquin (2015)** proposes the CEDEL2 design as a 'good practice' L2 corpus design.

- Gilquin, G. (2015). From design to collection of learner corpora. En S. Granger, G. Gilquin, & F. Meunier (Eds.), *The Cambridge Handbook of Learner Corpus Research* (pp. 9-34). Cambridge University Press. <https://doi.org/10.1017/CBO9781139649414.002>
- Lozano, C., & Mendikoetxea, A. (2013). Learner corpora and second language acquisition: The design and collection of CEDEL2. En A. Díaz-Negrillo, N. Ballier, & P. Thompson (Eds.), *Automatic Treatment and Analysis of Learner Corpus Data*. (pp. 65-100). John Benjamins. <https://doi.org/10.1075/scl.59.06loz>
- Sinclair, J. (2005). How to build a corpus. En M. Wynne (Ed.), *Developing Linguistic Corpora: A Guide to Good Practice* (pp. 79-83). Oxbow books.

Additional recommendations for L2 corpora applied to CEDEL2

- **Tracy-Ventura & Paquot (2021b) recommend:**
 1. focus on **L2s other than English**;
 2. include learners at **all proficiency levels**, with **varied L1s**, from **different ages**, and from **different learning backgrounds** and settings;
 3. promote **cross-linguistic comparisons**;
 4. include **more learner and task variables** (metadata);
 5. include **varied tasks**, some of which promote hypothesis-testing research;
 6. consider different perspectives on what a '**control**' corpus is;
 7. be **freely available** to the research community (Open Science).

Tracy-Ventura, N., & Paquot, M. (2021). The future of corpora in SLA. En N. Tracy-Ventura & M. Paquot (Eds.), *The Routledge Handbook of Second Language Acquisition and Corpora*. Routledge. <https://doi.org/10.4324/9781351137904>

PRINCIPLE: Content selection



The contents of a corpus should be selected without regard for the language they contain, but according to their **communicative function**



(Sinclair 2005:1)

- Sinclair distinguishes 2 content criteria:
 - **Internal criteria:** language of texts.
 - **External criteria:** communicative function of texts.
- **CEDEL2** design follows **external criteria** only.
 - So, CEDEL2 is not designed to elicit any specific linguistic structures/items, but potentially a wide range of them.
 - Cf next principle.

PRINCIPLE: Representativeness

“ Corpus builders should strive to make their corpus as **representative** as possible of the language from which it is chosen. ”

(Sinclair 2005:2)

- Different types of tasks:
 - **14 topics** (compositions): argumentative, descriptive, narrative...
 - **Graded topics** according to difficulty
 - Designed to **elicit many kinds** of linguistic phenomena.

- Also representative in terms of:
 - Different chronological **age ranges**
 - Different **settings** (naturalistic, secondary school, university)
 - Different **ages of exposure** (to study the age factor).

Title of composition about to be written:

1. ¿Cómo es la región donde vives?

PLEASE CHOOSE ONLY ONE COMPOSITION TITLE:

1. ¿Cómo es la región donde vives?

2. Habla de una persona famosa.

3. Resume una película que has visto recientemente.

4. ¿Qué hiciste el año pasado durante las vacaciones?

5. ¿Cuáles son tus planes para el futuro?

6. Describe un viaje que has hecho recientemente.

7. Cuenta una experiencia que hayas vivido.

8. Habla del problema del terrorismo en el mundo.

9. ¿Qué opinas de la nueva ley anti-tabaco?

10. ¿Crees que las parejas gay tienen el derecho de casarse y adoptar niños?

11. ¿Crees que la marihuana se debería legalizar?

12. Analiza los principales aspectos de la inmigración.

PRINCIPLE: Contrast

“ Only those components of corpora which have been designed to be **independently contrastive** should be contrasted ”

(Sinclair 2005:3)

- CEDEL2 allows for **full CIA** (Granger 1996, 2015)
- Importantly, all subcorpora are **fully contrastable**:
 - they have been collected under the **same conditions** (same tasks, etc) and follow the **same design principles**.

CEDEL2 & COREFL: Between-corpora contrasts

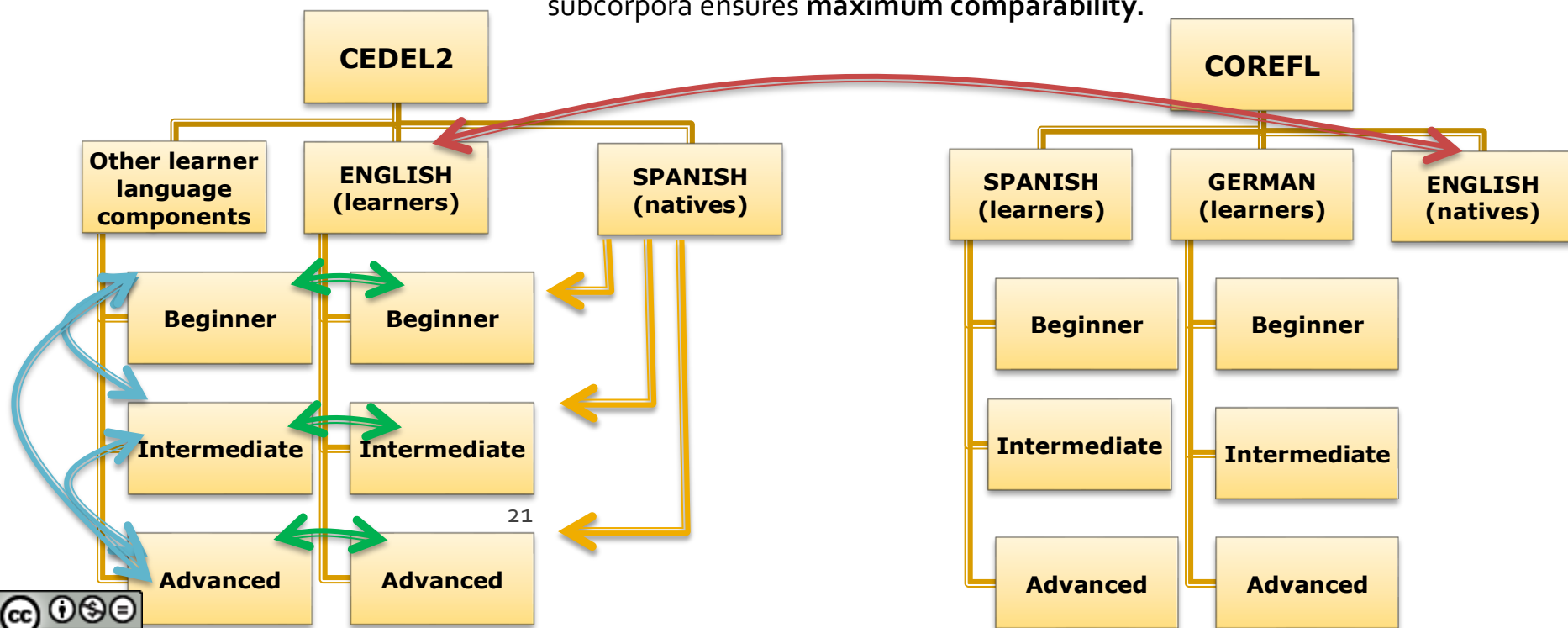
Beyond Contrastive Interlanguage Analysis (CIA):

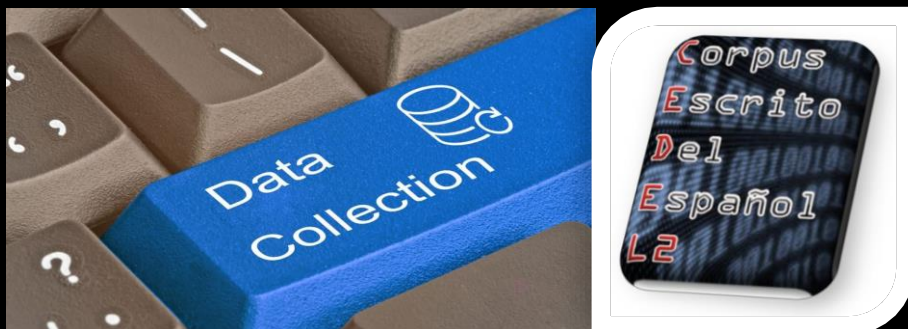
- **Developmental factors** across same L1.
- **Developmental factors** between diff. L1s.
- **L2 acquisition (against the native benchmark)**
- **L1 effects (Eng) on L2 Spa.**

Bi-modal: oral vs written

- **Within- vs between-subject** (same task)
- **Between-task effects** (different tasks)

Similar design across corpora (CEDEL2 vs COREFL) and their subcorpora ensures maximum comparability.





DATA COLLECTION

CEDEL2 (version 2): Data collection instrument: Google Forms

www.learnercorpora.com

- 1 Learner profile: biodata
- 2 Learner profile: L2 profile
- 3 Task (written/oral)
- 4 Placement test

Challenge in LC design: Learner profile

CHALLENGE

- While **learner variables** are acknowledged to be essential, not all corpora include enough ling variables that are crucial for SLA researchers.
 - “these variables have not been controlled consistently across corpora and are seldom incorporated in metadata” (Díaz-Negrillo & Thompson 2013:13)
- **SLA-motivated variables** are recorded in CEDEL2 so as to answer **SLA-relevant questions**.

1a. Biodata

Biodata:

- Sex
- Age [CHRONOLOGICAL AGE FACTOR]
- Educational profile [EDUCATIONAL LEVEL FACTOR]
- Mother's and Father's L1 [L1 TRANSFER FACTORS vs UNIVERSAL FACTORS]
- Language spoken at home [LANGUAGE USE/DOMINANCE FACTORS]

DATOS PERSONALES

(1) Tus iniciales *

Tu respuesta

(2) Sexo *

- Hombre
- Mujer

(3) Edad *

Tu respuesta

(4) Universidad / colegio / institución a la que perteneces (si procede)

Tu respuesta

(5) Carrera universitaria que cursas o has cursado (si procede)

Tu respuesta

(6) Curso universitario / en el colegio (si procede)

(7) Tu lengua(s) materna(s): *

En "Otro" puedes escribir más de una. Por ejemplo: español/francés; español/inglés; etc.

- Español
- Other:

(8) Lengua(s) materna(s) de tu padre *

- Español
- Other:

(9) Lengua(s) materna(s) de tu madre *

- Español
- Other:

(10) Lengua(s) que hablas en casa *

- Español
- Other:

1b. Linguistic profile

Learners' linguistic profile:

- **AGE** at which L2 learning started [**AGE-OF-EXPOSURE FACTOR, CRITICAL PERIOD**]
- **Years** you have been learning the L2 for [**LENGTH-OF-EXPOSURE FACTOR**]
- **Stays** in Spanish-speaking country: **duration**, where [**LENGTH-OF-RESIDENCE FACTOR, NATURALISTIC EXPOSURE FACTOR**]

PERFIL LINGÜÍSTICO

(11) Age at which you started to learn Spanish (in years) *

Your answer _____

(12) Number of years studying Spanish *

For example 0.5 / 1 / 2.5 etc.

Your answer _____

(13) Have you stayed in a Spanish-speaking country for one month or longer? *

If your answer is NO, you can move on to question (17) Certificates.

Yes

No

(14) If yes, where?

Your answer _____

(15) When?

State the year(s)

Your answer _____

(16) For how long (in months)?

If you have been several times and/or in several countries, please write the total number of months.

Your answer _____

1b. Linguistic profile

Profile about the learners' L2:

- Linguistic level: **SELF-EVALUATION** in the 4 skills; language certificates [**PROFICIENCY-LEVEL FACTOR; DEVELOPMENTAL FACTORS**]
- **Other foreign languages** apart from Spanish and **self-evaluation** [**L2 vs L3 FACTOR**]

High **correlation** between placement test and self-evaluation (Lozano 2016)

(17) Certificates. If you have any official certificates stating your level of Spanish, state which one(s) and the level

For example, DELE (B2)

Your answer

(18) Please estimate your ability in Spanish. *

Please indicate your real language level regardless of any language certificates you may have.

	Lower Beginner (A1)	Upper Beginner (A2)	Lower intermediate (B1)	Upper Intermediate (B2)	Lower Advanced (C1)	Upper Advanced (C2)
Speaking	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Understanding	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Reading	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Writing	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

(19) Do you speak a third language (in addition to Spanish and your native English)?

If yes, please write it down and then estimate your abilities in the next question. Otherwise, write 'no' and scroll down and go to the next section.

Your answer

Challenge in LC design: task variables

CHALLENGE

- Task variables must be also taken into account, as task differences have been shown to affect learner output.
- Tasks have to be varied so as to elicit different types of language.

Tasks

14 tasks to choose from:

Text type variety: descriptive, narrative, argumentative.

Tasks elicit **all types of verbal tenses** (present, different types of past, future)

Tasks elicit a **wide range of linguistic structures**.

- 1. Describe the **region where you live**.
- 2. Talk about a **famous person**.
- 3. Describe a **film** you have recently seen.
- 4. What did you do last year during your **holidays**?
- 5. Which are your plans for the **future**?
- 6. Describe a **trip** you have recently made.
- 7. Narrate an **experience** you have lived.
- 8. Talk about the problem of **terrorism** in the world.
- 9. What do you think about the **anti-tobacco law**?
- 10. Do you think **gay couples** have the right to get married and adopt children?
- 11. Do you think **marihuana** should be legalised?
- 12. Analyse the main aspects of **immigration**.
- 13. Describe the **frog story** (picture-task).
- 14. Describe the **Charles Chaplin** silent video clip.

2. Tasks

Additional info about the task:

- duration
- location
- tools/resources used

[TASK CONDITION FACTORS]

(21) ¿Cuánto has tardado en realizar la tarea aproximadamente? *

*

h min s

: :

(22) ¿Dónde has realizado la tarea? *

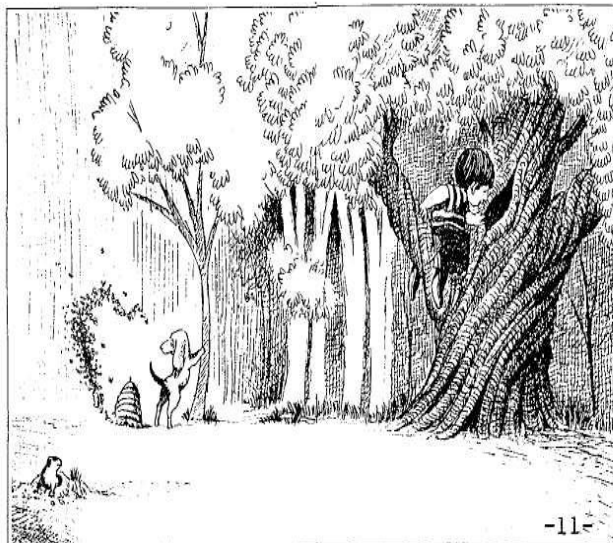
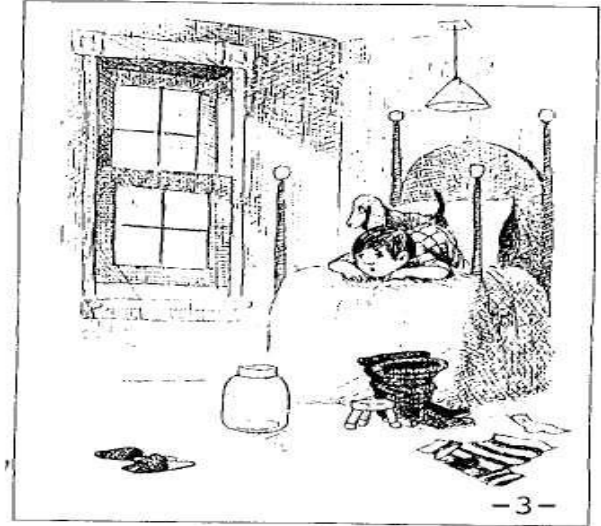
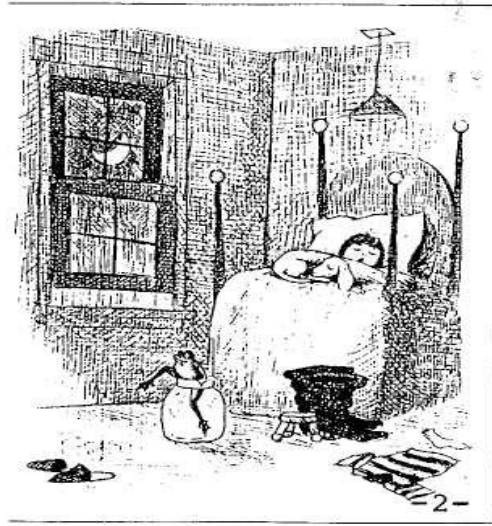
- En clase
- Fuera de clase (en casa, en la biblioteca, en el tren...)
- En clase y fuera de clase.

(23) ¿Has usado alguno de los siguientes recursos?

Marca todos los que hayas usado.

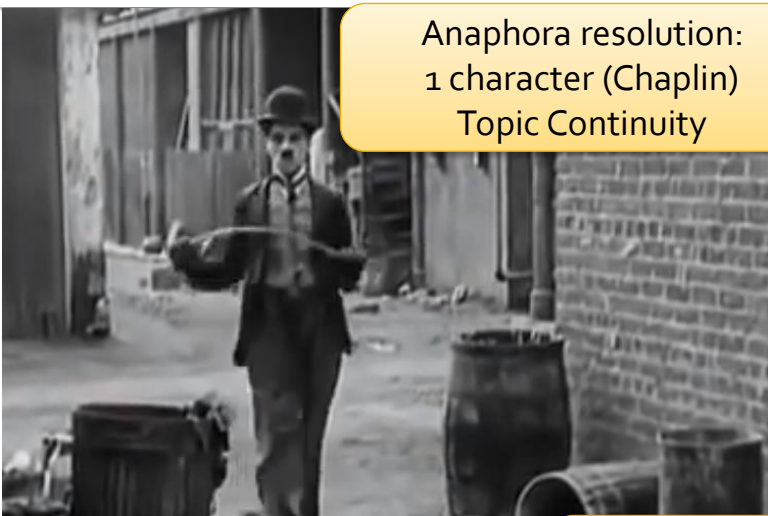
- Ayuda de un nativo
- Diccionario bilingüe (español-inglés / inglés-español)
- Diccionario monolingüe
- Corrector ortográfico
- Gramática
- Materiales relacionados con el tema de la tarea (periódicos, internet, TV, etc.)

Task: Frog, where are you?



12 pictures in total

Task: Chaplin (The kid): ideal to investigate anaphora



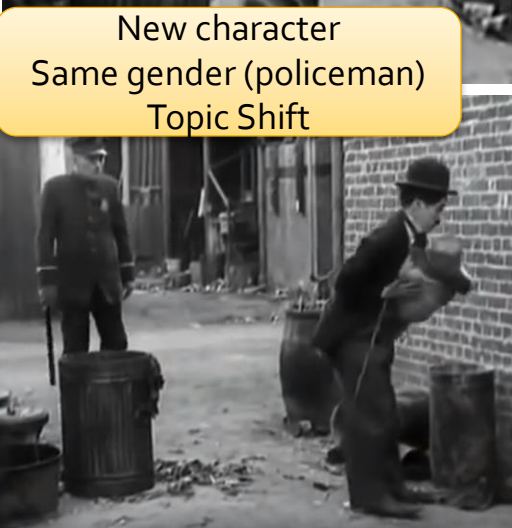
Anaphora resolution:
1 character (Chaplin)
Topic Continuity



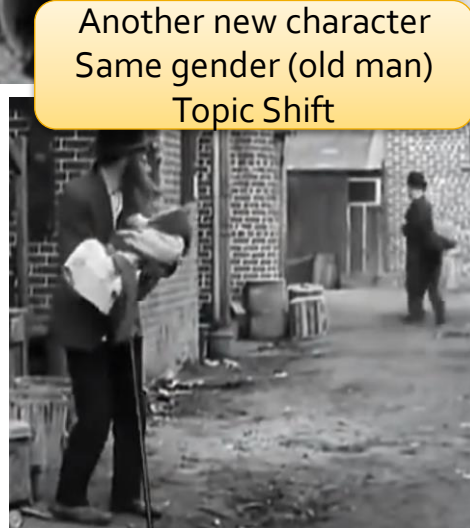
New character (baby)



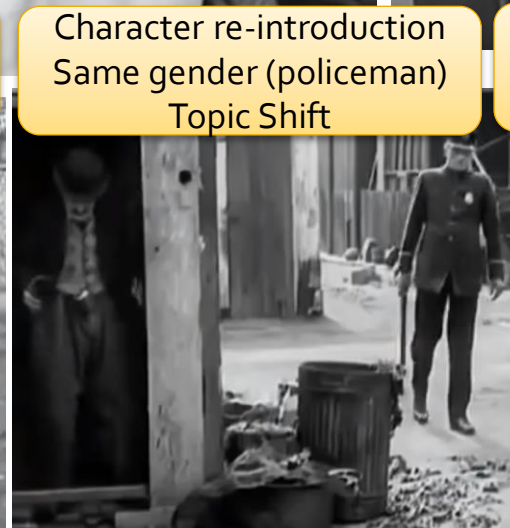
New character
Different gender (woman)
Topic Shift



New character
Same gender (policeman)
Topic Shift



Another new character
Same gender (old man)
Topic Shift



Character re-introduction
Same gender (policeman)
Topic Shift



Character re-introduction
Different gender (woman)
Topic Shift

<https://www.youtube.com/watch?v=4QkTNJFhu-g>

Challenge for LC design: Standardised proficiency level

CHALLENGE

- A problem in many LC (e.g., ICLE, SPLLOC):
 - 'Ad hoc' measures of proficiency: **institutional status** (e.g., classroom/year level).
 - **Myles (2015:316)**:
 - "Being in the same year group at school is not always a sufficiently rigorous indication, and it is advisable to carry out **independent measures of proficiency**".

3. Placement test

CEDEL2

- **Standardised placement test** (University of Wisconsin Placement Test)
- Also: **self-proficiency** ratings in the 4 skills (speaking / writing / listening / reading).
- Also: oficial lang proficiency **certificates**.

Bottom line: CEDEL2 includes learners from **6 different proficiency levels** as to maximize its usefulness to study **development** in L2 grammar

SPANISH PLACEMENT TEST

INSTRUCTIONS:

(Test from the University of Wisconsin, 1998)

This test is important, as it helps us better understand your level of Spanish grammar. You will get your result: **FOR FREE** after completing the test.

There are two multiple choice sections. Each question has only one correct response. In some sentences, you don't have to change anything, so select "--".

Please answer as many questions as you can and do **NOT** use any dictionaries, grammar resources, or other help.

Section 1: Multiple choice

1. No veo ___ los muchachos.

1 punto

a

--

2. ¡Pobre Pablo! Hoy ___ enfermo.

1 punto

está

es

3. A: "¿Te costó mucho el libro?" B: "Sí, pagué veinte dólares ___ este libro."

1 punto

para

por

4. Tomás siempre escuchaba la radio mientras___.

1 punto

leía

leyó

Self-evaluation of proficiency (apart from standardized UWPT)

(17) **Certificados:** Si tienes algún certificado de inglés, indica cuál y el nivel.

P.ej., First Certificate (B2), Proficiency (C2), etc.

Tu respuesta

(18) **Valora tu nivel de inglés ***

Indica aquí cuál crees que es tu nivel de inglés real, independientemente de los títulos que poseas.

	Principiante bajo (A1)	Principiante alto (A2)	Intermedio bajo (B1)	Intermedio alto (B2)	Avanzado bajo (C1)	Avanzado alto (C2)
Expresión oral	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Comprensión auditiva	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Comprensión lectora	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Expresión escrita	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>


(19) **¿Hablas algún otro idioma extranjero además de inglés?**

¿Hablas algún otro idioma extranjero además de inglés? Si es así, indica qué idioma y valora tu nivel en la siguiente pregunta. Si no, escribe 'no' y pasa a la siguiente sección.

Self-evaluation measures **significantly correlate** with LoI (Length of Instruction) and LoR (Length of Residence abroad) → (Lozano 2016)

Sample data (Google Sheets)

- Manual cleaning of data.
- Standardization protocol.

CEDEL2 L1 English-L2 Spanish ☆ 
File Edit View Insert Format Data Tools Form Add-ons Help [Last edit was 1 hour ago](#)

100% £ % .0 .00 123 Arial 10 B I A            

fx FILENAME:




	A	B	C	D	E	F	G	H	I	J	K
1	FILENAME:	CERTIFICA	Timestamp	Score	PLACEMI	PROFICIENCY I	If you agr	(1) Your i	(2) Sex	(3) Age	(4) School/University/Institu
2	EN_WR_20_20_1_2_MD		16/11/2016 22:45:30	20 / 43	46.5	upper beginner		MD	Female	20	Denison University
3	EN_WR_15_19_1_7_JRM		17/11/2016 00:37:23	15 / 43	34.9	upper beginner		JRM	Female	19	Denison University
4	EN_WR_29_18_6_2_LAC		17/11/2016 19:17:07	29 / 43	67.4	upper intermediate		LAC	Female	18	Denison University
5	EN_WR_20_21_4_7_JM		18/11/2016 06:30:12	20 / 43	46.5	upper beginner		JM	Male	21	Denison University
6	EN_WR_32_18_6_1_LD		18/11/2016 15:55:45	32 / 43	74.4	upper intermediate		LD	Female	18	Denison University
7	EN_WR_28_20_4.5_1_HH		18/11/2016 17:53:13	28 / 43	65.1	lower intermediate		HH	Female	20	Denison University
8	EN_WR_27_19_5_4_JR		18/11/2016 19:24:44	27 / 43	62.8	lower intermediate		JR	Female	19	Denison University
9	EN_WR_24_19_8_1_KL		19/11/2016 04:08:28	24 / 43	55.8	lower intermediate		KL	Female	19	Denison University
10	EN_WR_32_18_6_10_MMQ		19/11/2016 22:11:31	32 / 43	74.4	upper intermediate		MMQ	Female	18	Denison University
11	EN_WR_24_20_4_2_BYS		23/11/2016 20:30:42	24 / 43	55.8	lower intermediate		BYS	Female	20	Claremont Mckenna College
12	EN_WR_14_19_3_3_GC		28/11/2016 17:08:33	14 / 43	32.6	upper beginner		GC	Female	19	Denison University
13	EN_WR_39_22_10_10 EMC		01/12/2016 23:25:43	39 / 43	90.7	lower advanced		EMC	Female	22	Denison University
14	EN_WR_27_19_5_3_CEW		05/12/2016 23:09:50	27 / 43	62.8	lower intermediate		CEW	Female	19	Denison University
15	EN_WR_38_18_6_1_CW		12/12/2016 02:24:09	38 / 43	88.4	lower advanced		CW	Female	18	Denison University
16	EN_WR_24_19_4_2_JNS		16/12/2016 06:07:20	24 / 43	55.8	lower intermediate		JNS	Female	19	Denison University
17	EN_WR_41_20_7_14_CP		19/12/2016 15:25:03	41 / 43	95.3	upper advanced		CP	Male	20	Haverford



STATS

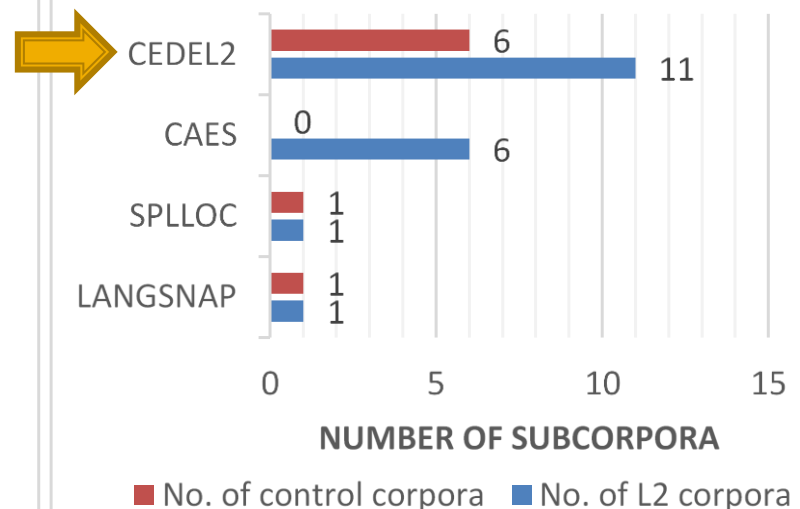
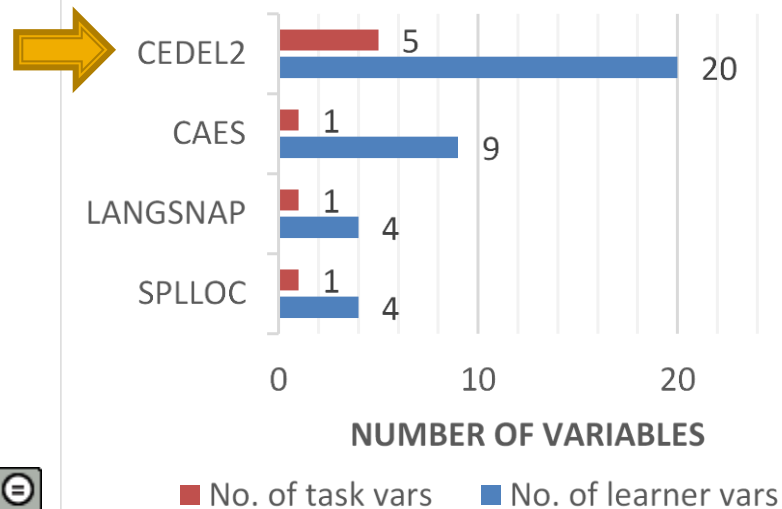
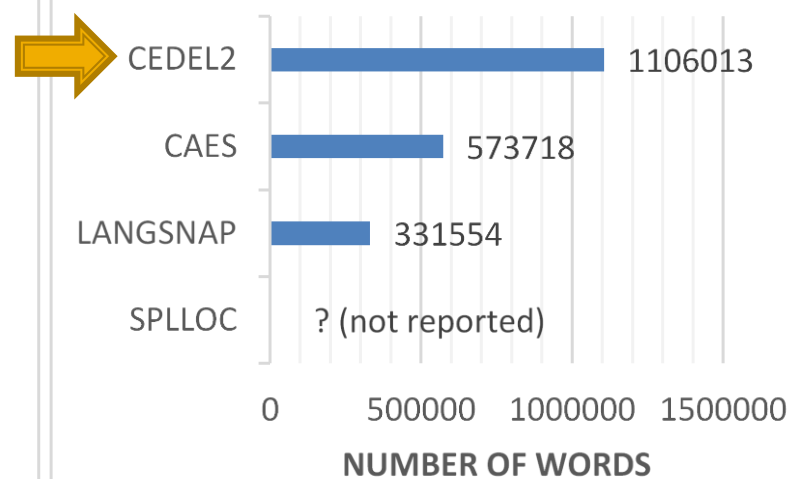
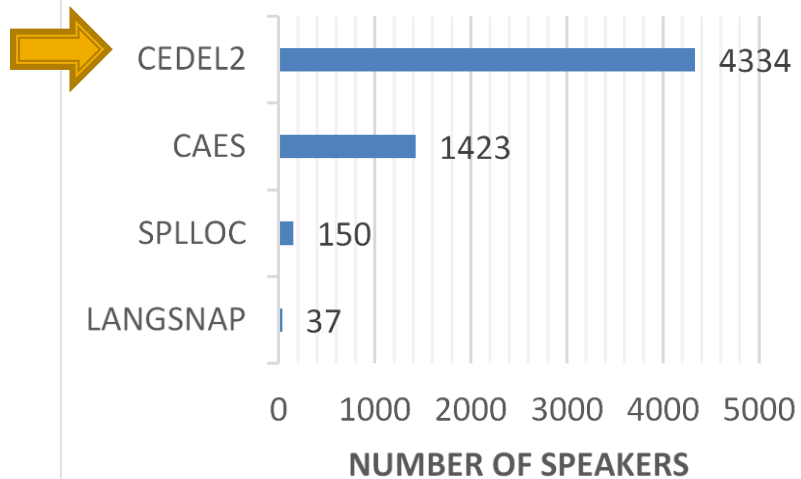
CEDEL2 version 2

Total stats

	4,334
	111
	4,400
AA	1,106,013

- Detailed stats: <http://cedel2.learnercorpora.com/statistics>
- Largest corpus of L2 Spanish

COMPARISON: CEDEL2 vs. other L2 Spa corpora



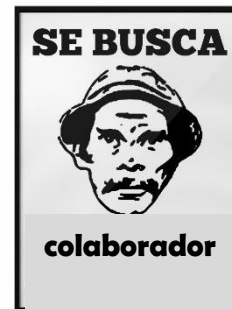
《《CALL FOR PARTICIPANTS》》



CALL FOR PARTICIPATION

Participation (online)

- www.learnercorpora.com
- Learners can get
 - Their level in the Spanish placement test
 - Certificate of participation
- Collaboration
 - Institutional
 - ELE / EFL teachers
- Looking for collaborators:



Participate in ...

▶ **CEDEL2**

Corpus
Escrito
Del
Español
L2

Corpus Escrito Del Español L2
L2 Spanish Written Corpus

Contact details:
Cristóbal Lozano
Universidad de Granada
cristoballozano@ugr.es

CEDEL2: Corpus Escrito del Español L2 (version 2) v2.0
Sept. 2020

CEDEL2 Search / Download User guide Statistics About Contact

Corpus to search

Subcorpus
Learners of L2 Spanish

Sensitivity

Case
 Orthographic accents

Result (Output)

Result type: Concordances (KWIC)
Result subtype: Words
Sorting: Concordance match
Page size: 50


Filters

L1: L1 any - L2 Spanish
Medium: Any
Sex: Any
Proficiency level: Any
Placement test score (%):
Proficiency (self-assessment):

Task title: Any
Filename: Any
Age:
Age of exposure to Spanish:
Years studying Spanish:
Stay abroad (months):

Words

Five words maximum



CEDEL2 web interface

Web interface

CEDEL2 (v. 2): <http://cedel2.learnercorpora.com>

CEDEL2: Corpus Escrito del Español L2 (version 2) v2.0
Sept. 2020

[CEDEL2](#) [Search / Download](#) [User guide](#) [Statistics](#) [About](#) [Contact](#)

Corpus to search

Subcorpus
Learners of L2 Spanish

Result (Output)

Result type: Concordances (KWIC) Result subtype: Words

Sorting: Concordance match Previous element Page size: 50

Sensitivity

Case
 Orthographic accents

Filters

L1: L1 any - L2 Spanish Task title: Any

Medium: Any Filename: Any

Sex: Any

Proficiency level: Any

Placement test score (%):

Proficiency (self-assessment):

Age:

Age of exposure to Spanish:

Years studying Spanish:

Stay abroad (months):

Words

Five words maximum

Basic search

(this is developed in the following slides)



hands-on

- Words/strings (with wildcards): `hab1*`
- Filters
- Downloads
- Grammatical elements
 - Lemmas: `SER` / `ESTAR`
 - Concordance errors: `la climaa`, `la poetaa`, `la díaa`
 - Word order: `Adj+N` vs. `N+Adj`

Búsqueda avanzada: LEMA

¿Ser o estar + locativo?



- ¿Soy en la calle o estoy en la calle?
- Elementos gramaticales:
 - Lema: *ser*
 - Preposición (palabra "en").

Grammatical elements

Grammatical element	Tag	<input type="text" value="ser"/>	Word	-	+
Grammatical element	Tag	<input type="text" value="Lemma"/>	<input type="text" value="en"/>	-	+

1 tags selected ?

... una casa de el bosque , Puff , Raúl y estaban viviendo sonriendo . Puff tenía su propia casa , **era en** una botella , y cuando el niño durmió , Puff también descansó en su pequeña habitación . Ellos era muy pequeño , estaba muy graciosa y sociable . Ella le encantaba cantar y reír . En la universidad , **era en** la sororidad " Zeta Tau Alpha " . Cuando era en las classes , ella gustaba actuar en el teatro ...

... la playa . No es una playa como las de California . hacia fresco y el mar era fria . **era en** esta misma lugar que por la primera vez yo vi el mar . me recordaba de ser un joven y las ...

Torres Gemelas solo horas antes de cuando las eran rompidas por los aviones . Aunque el ataque **era en** todos canales de el televisor , yo y mis compañeros no lo conocimos nada de eso hasta el proximo

... no era su mascota . El animal tocó el chico y su perro entre el agua . Cuando el chico **era en** el agua el chico finalmente encontró su mascota y encontró toda de la familia de su mascota . El ...

de restaurantes fabulosas , de cualquier tipo de comida . Pero , el restaurante mejor de Colombia **era en** Chia , una_hora por coche a los picos afuera de Colombia . Este restaurante se llama " Andrés Car

... nos spagnol , pero y mucho mas facil a entender . Por ahora mi gusto mucho mi vida , como **era en** la passada y como es en este momento , y espero que continuera para el futuro .

... fuimos a Las Vegas , nosotros fuimos a Bryce Canyon y Zion National Park . Yo recuerde cuando **era en** Zion , una persona muero cuando el en un top de el un cliff . Era cientoidos pies de el ...

En esta historia , Había un pobre hombre en un mal lugar en alguna ciudad , cuando **era en** la calle , había hubierto un niño . Este hombre se llevaba y se ponía en un coche con otro ...

... encantaba cantar y reír . En la universidad , era en la sororidad " Zeta Tau Alpha " . Cuando **era en** las classes , ella gustaba actuar en el teatro . Despues de graduación , Ashley mueva a Los Angeles

... ios de el colegio son recuerdos muy preciosos y agradables para la vida de una persona . Cuando **era en** el colegio , no me aburría nunca . Jugaba a la pelota con mis compañeros , hacíamos punto , coque

... ar a los otros amigos en otras ciudades y salíamos todos juntos a los bares de la ciudad . Cuanto **era en** casa me encantaba leer libros para pasar mi tiempo libre alegremente . Generalmente este veran

... . Tambien , ve donde el locacion de los World Trade Centers . Es muy deprimido . Todos los dias **era en** New York , era muy calor y hace sol . Ojala que un dia mi familia y yo pueda visitar ...

... jlesa atrapada en un país franco-flamenco donde no comunicaban con gente fuera y la educación **era en** segundo lugar después de emborrachar se a el estilo ingles . Me sentía inmovilizada y deseaba cc

Búsqueda avanzada

Concordancia artículo-nombre



- *¿El clima o la clima? ¿El idioma o la idioma? ¿El día o la día?*
- Elementos gramaticales (tags):
 - Artículo femenino + Nombre masculino (que termine en "a")

Grammatical elements

Grammatical element Tag Lemma Word - +

1 tags selected ?

Grammatical element Tag Lemma *a - +

2 tags selected ?

... flores y a ver los bonitas colores . en el verano , todos los jovenes va a la playa , **la cinema** , va comprado y etc. Es muchas cosas a hacer en el verano . Atlanta tiene muchos parados y fiestas ...

... la Vacaciones , me visite mi hermana . Ella vive en la ciudad de Neuva York . Nosotros fuimos a **la cinema** a ver peliculas juntos . Entonces Nosotoros fuimos a las tiendas ir de compras , y ella y me compramos ...

... Riehle , Mary Kate Riehle , Elizabeth Riehle , Kristen Riehle , y Joey Riehle . En el invierno , **la clima** hace frio . En el verano lo hace calor . Durante el verano , mi familia viaja a otro citio ...

En Junio de el ano pasodo fui a Sevilla , Espana con mi novia . **la clima** ahi estaba muy buena y la temperatura tambien . El aeropuerto de Sevilla es muy pequena y muy cerca de ...

... cercas de los peotonos . Por fin , me gustaria visitar a Sevilla otra vez . La gente son amables **la clima** estupenda y la comida buena . Me gusto mucho .

... de color . En el invierno hace frio pero tanto como en el norte . A veces tenemos suerte con **la clima** en el invierno . El sur es una region diverso y interesante . Espero que todas puedan visitar lo a ...

... mas o menos , pescando sin suerte . Pasamos casi todo el dia buscando y pescando . Gracias a Dios **la clima** era bien . Tuvimos neblinas pero ni una gota de lluvia . Finalmente mi hermano tenia cuatro salmones pero

... y por la vida mas natural y barata . Me encanta las montanas y las vistas aqui . Me encanta **la clima** y las estaciones distintas . El invierno que hace mucho frio y nieve pero con un cielo azul y claro ...

... Era niño Che Guevara sufría de Asma . Por eso , sus padres movían mucho en su adolescencia para encontrar **la clima** apto . A pesaba de la enfermedad , el jugaba muchos deportes como Rugby . También , era niño el ...

... Muchos son bonitos con gente amables . Pero , el mejor lugar de el mundo está aquí . ¿ Es **la clima** o la naturaleza ? Creo que no . Creo que es porque todo mi familia está aquí y es el ...

... turo incluyen viviendo y enseñando inglés y literatura norteamericano en un país hispanohablante . A mí me molesta **la clima** política en los Estados Unidos , y mi esposa Cubana no puede trabaja a el nivel suficiente por su experien

Búsqueda avanzada: Orden de palabras (Adjetivo+Nombre)



- *¿Alcohólicas bebidas o bebidas alcohólicas?*
 - Tags: Adjetivo + Nombre

Grammatical elements

Grammatical element	<input type="text" value="Tag"/>		Lemma	Word	-	+
	64 tags selected ?					
Grammatical element	<input type="text" value="Tag"/>		Lemma	Word	-	+
	18 tags selected ?					

... es vodka . Su tiene muchos **agradable ropas** , y sus mucha
sa . Yo tiene vino y diferente **alcoholicas bebidas** . Nosotr
-amburgo y está casado con **alemana-mexicana Monique**
le estuvo llena de sabiduría y **alentadoras palabras** . A fines
e hombre en un mal lugar en **alguna ciudad** , cuando era er
sto tres gatos y una pero . Yo **also gusto** to become el doct
ar la comida muy fresca y de **alta calidad** . Tambien , poder
e está muy popular . Un poco **alta montañas** son en Kane , l
jor , una sistema de salud de **alta nivel** . Si una persona vie
uardó con una tecnología de **alta nivel** , y el hombre nunca
ural v su genética indica una **alta probabilidad** de padecim

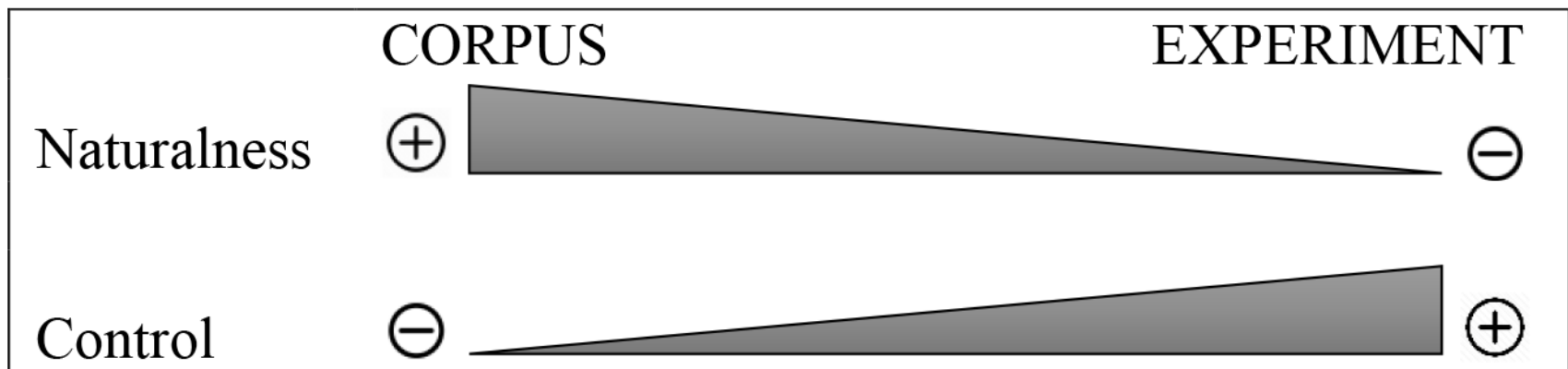
**Final recommendations:
Good corpus design +
Triangulation in SLA**

Conclusions

- **Learner (L2) corpora** can contribute to theoretically-motivated analyses of in SLA provided...
 - The corpus is **designed** according to certain principles.
 - SLA-motivated information about each informant and the task is collected (**metadata**) .
 - The corpus is **tagged/annotated** based on SLA-motivated tagsets (instead of tagging errors wholesale).

Methods to investigate L2 acquisition (Lozano 2021)

- According to the nature of the **data**:



- Ideal world: **triangulation.**

Lozano, C. (2020). How to do research in SLA by combining corpus and experimental methods: An introduction for language researchers and teachers. En M. Planelles Almeida, A. Foucart, & J. M. Licerias (Eds.), *Perspectivas actuales en la enseñanza y el aprendizaje de lenguas en contextos multiculturales / Current Perspectives in Language Teaching and Learning in Multicultural Contexts* (pp. 389-421). Thomson Reuters Aranzadi. <https://doi.org/10.32029/2605-4655.11.01.2020>

THE END

Merci beaucoup!! 😊



UNIVERSIDAD
DE GRANADA

