

Past Tense Use in L1 and L2 Spanish Narratives: A Corpus-Based Study of Form, Meaning, and Context



Maura Cruz Enríquez
Anahí Alba de la Fuente
François Lareau

Corpus Approaches to Lexicogrammar (LxGr)

July 3rd, 2026

Context

- Past tense verbs pose a major challenge in second language acquisition (SLA) (Bardovi-Harlig, 2000; Slabakova and Montrul, 2002; González, 2003; Izquierdo, 2009; Salaberry, 2011).
- Much of the research on the acquisition of tenses in Spanish has focused on the contrast between the simple tenses of the past: the imperfect *cantaba* and the preterit *cantó*.
- They are usually studied in a paradigmatic relationship, when the change in form implies a change in meaning:
 - *Los barcos llegaban/llegaron cada media hora* (González, 2003)
Boats arrived_IMP/arrived_PRET every half hour.

Context

- These tenses are rarely studied in a paradigmatic relationship with other tenses to express the same meaning:
 - *Vio que la chica **robaba/ estaba robando** un pan.*
 - *He saw that the girl stole_IMP/ was stealing a loaf of bread.*
 - *La señora dijo que la chica **robó/había robado** un pan.*
 - *The lady said that the girl stole_PRET/had stolen a loaf of bread.*

Context

- Recent research has emphasized the importance of large-scale learner corpora for investigating linguistic phenomena in context (N. C. Ellis, 2019).
- Recent advances in Natural Language Processing (NLP) allow researchers to process large amounts of data efficiently.

However:

- most L2 corpus studies still rely on relatively small manually annotated datasets.

PROJECT OVERVIEW

A corpus-based study of past tense use in L1 and L2 Spanish narratives, integrating morphological, semantic, and discourse information.

THE CORPUS

PILOT CORPUS (current study)

- 40 participants
 - 20 L1 Spanish
 - 20 Advanced L2
- Oral & written narratives
 - Modern Times
 - Anecdote task
- ~27,000 words
- Manually annotated subset

TARGET CORPUS (long-term goal)

- Large-scale dataset
 - Hundreds of participants
- Multiple proficiency levels
 - A2 → C2
- Multiple tasks and modalities
 - Oral narratives, written narratives, classroom writing
- Complementary sources
 - Learner corpora (e.g., CEDEL2)
 - University language courses in Quebec

PROJECT STAGES

1. DATA COLLECTION

- Elicited and authentic narratives in Spanish (L1 & L2)
- Oral and written

2. AUTOMATIC PROCESSING

- Python + spaCy Pipeline
- Clause segmentation
- POS tagging & parsing
- Verb form extraction
- Temporal markers
- Syntactic relations

RESOURCES

- ADESSE database (process types and argument structure)
- Rule-based modules and lexical lists

3. MANUAL ANNOTATION

- Semantic-pragmatic annotation
- Process type
- Discourse function
- Temporal relations
- Aspectual meaning
- Subjectivity / perspective
- Ambiguity resolution

ITERATIVE REFINEMENT

- Update annotation scheme and rules based on observed patterns and errors

4. ANALYSIS

- Quantitative & qualitative analyses
- Distribution of past tense forms
- Mixed-effects modeling
- Cross-group and developmental comparisons

5. APPLICATIONS

- Informing automatic models & pedagogy
- Improved annotation and NLP tools
- Pedagogical materials based on corpus evidence

OUTPUTS & IMPACT

- Annotated corpus
 - Publicly usable resource

- Scientific insights
 - On tense polysemy and development

- Methodological contributions
 - Hybrid annotation frameworks and processing tools

- Pedagogical impact
 - Data-driven materials and teaching strategies

ITERATIVE CYCLE: CORPUS GROWTH AND MODEL IMPROVEMENT



THE CORPUS

PILOT CORPUS (current study)



40 participants

20 L1 Spanish

20 Advanced L2



Oral & written
narratives

Modern Times



~27,000 words

Manually annotated subset

L1 Spanish:

cantó – ha cantado contrast

Cuba (n = 12),

Spain (n = 4),

Colombia (n = 2), and

Mexico (n = 2).

PROJECT STAGES

1. DATA COLLECTION



Elicited and
authentic narratives
in Spanish (L1 & L2)

Oral and written

2. AUTOMATIC PROCESSING

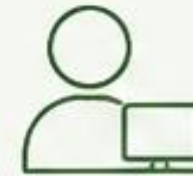


Python + spaCy
Pipeline

- Clause segmentation
- POS tagging & parsing
- Verb form extraction
- Temporal markers
- Syntactic relations

Termino → present
Terminó → preterit

3. MANUAL ANNOTATION



Semantic-pragmatic
annotation

- Process type
- Discourse function
- Temporal relations
- Aspectual meaning
- Subjectivity / perspective
- Ambiguity resolution

What it means in context
and what function it fulfills
in the discourse.

4. ANALYSIS



Quantitative &
qualitative analyses

- Distribution of past
tense forms
- Mixed-effects
modeling
- Cross-group and
developmental
comparisons

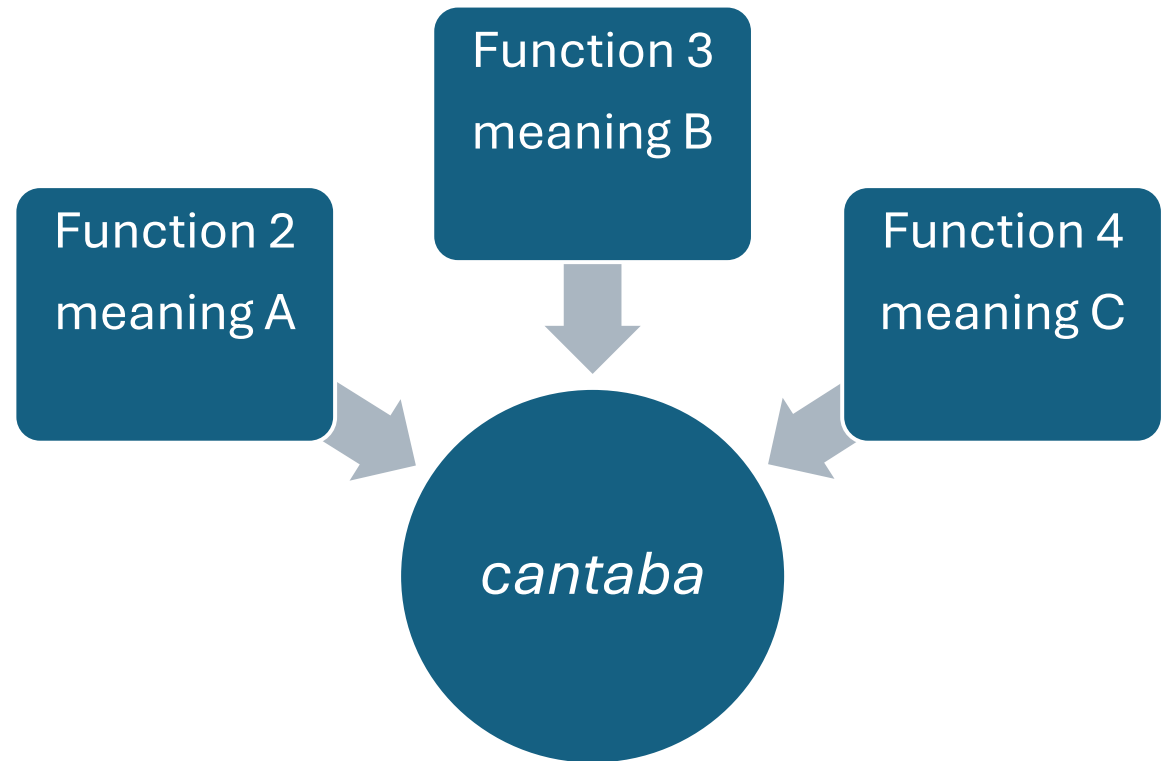
Theoretical framework

Based on the **Function-meaning-form model (FMF)** (Cruz Enríquez, 2019, 2022; Cruz Enríquez and Alba de la Fuente, 2024; Cruz Enríquez and Lareau, 2025):

- It integrates insights from:
 - Functional linguistics: process types (M. Halliday, 1985; M. A. K. Halliday & Matthiessen, 2014), and
 - Enunciation theory: distinction between commentary and narration (Weinrich, 1973).
- It accounts for key dimensions in the interpretation of verbal forms: tense, grammatical aspect, lexical aspect, process type and subjectivity, as well as distinctions between narrative and commentative discourse.

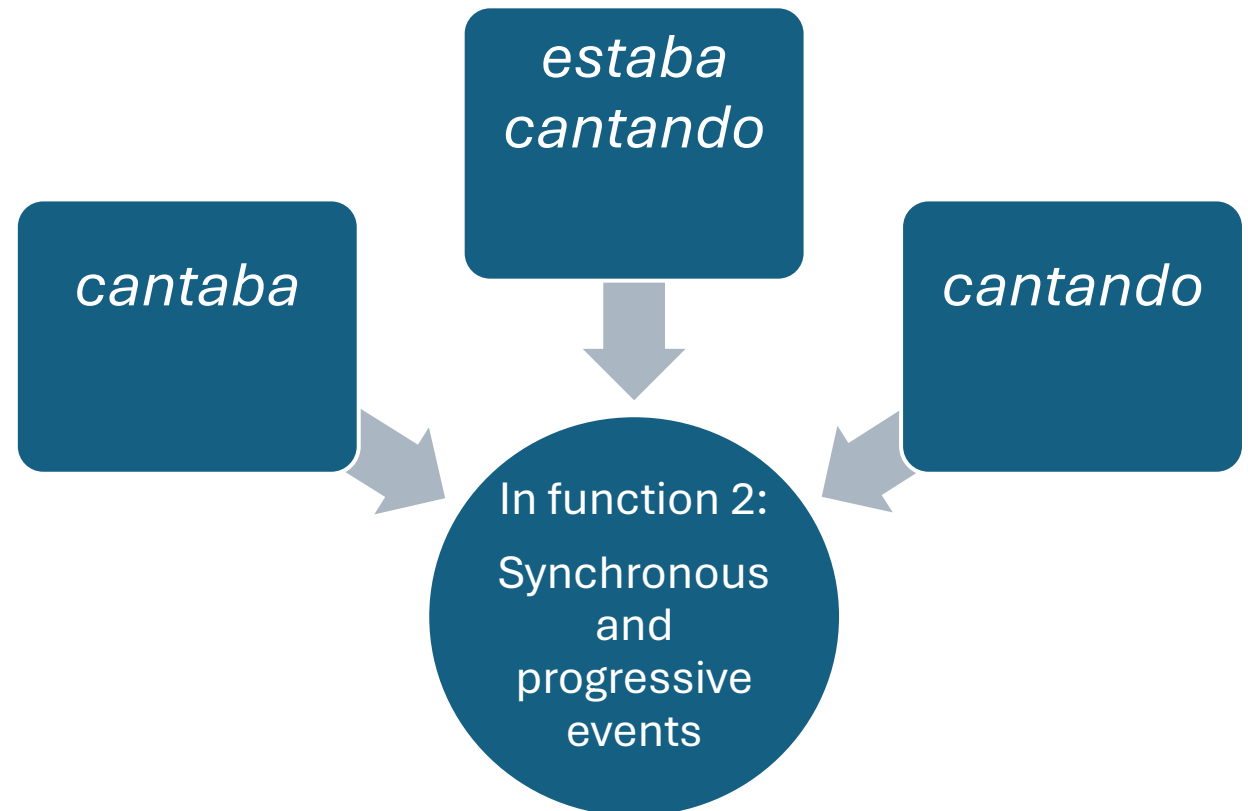
Function-meaning-form

- Four narrative functions that allow:
- to study the polysemy or contextual uses of verb tenses,
- to connect each verbal form to one of its meanings.



Function-meaning-form

- Four narrative functions that allow:
- to study the polysemy or contextual uses of verb tenses,
- to connect each verbal form to one of its meanings.





Sequence of narrative processes (F1)

*La chica **robó** una barra de pan y **huyó**. Luego **tropezó** con Chaplin y ambos **cayeron** al suelo.*

The girl stole a loaf of bread and fled. Then he tripped over Chaplin and they both fell to the ground.



Projecting the commentary material onto the background (F2)

- “**Soy inocente, no he hecho nada**”

"I'm innocent, I haven't done anything"

- *La chica dijo que **era** inocente y que no **había hecho** nada.*

The girl said she was innocent and had done nothing.



Projecting the commentary onto the background from a character's point of view (subjective)(F2)

*Una señora vio que la chica **robaba** una barra de pan.*

A lady saw that the girl stole_IMP a loaf of bread.



Time frame (F3)

*Quando **terminó/ terminaba** de comer, llamó a un policía.*

*When he finished_ **PRET/IMP** eating, he called a policeman.*



Subjective Narrative Process Sequence (Real-Time Narration) (F4)

*Chaplin **tomaba** la leche de una vaca, se **comía** unas uvas del patio y luego él y la chica **se sentaban** a comer.*


Chaplin drank_IMP milk from a cow, ate_IMP some grapes from the courtyard and then he and the girl sat_IMP down to eat.

[Context: The character imagines this scene]

Research Questions

In **function 2 (F2)**: Projecting the comr

For these two **meanings**:

- Synchronous and progressive 
- Anterior and narrative



Una señora vio que la chica robaba / estaba robando una barra de pan.

A lady saw that the girl stole IMP/was stealing a loaf of bread.

Question:

- Which **verbal forms** are used to express these meanings?
- Are there differences between groups (L1 and L2)?
- Are there differences based on modality (oral vs written)?

Research Questions

In **function 2 (F2)**: Projecting the comment

For these two **meanings**:

- Synchronous and progressive
- Anterior and narrative



- *El panadero dijo que la chica **robó** / **había robado** una barra de pan.*

The baker said the girl stole/had stolen a loaf of bread.

Question:

- Which **verbal forms** are used to express these meanings?
- Are there differences between groups (L1 and L2)?
- Are there differences based on modality (oral vs written)?



Data analysis

brms package (Bürkner, 2017) in R (R Core Team, 2022)



- Verb_tense as the dependent variable
- Subject_id as the random intercept
- Group, modality as fixed effects

- Bayesian mixed-effects models reveal differences between groups in the distribution of past tense forms in function 2.

Results

Synchronous and progressive events



Una señora vio que la chica robaba / estaba robando una barra de pan.

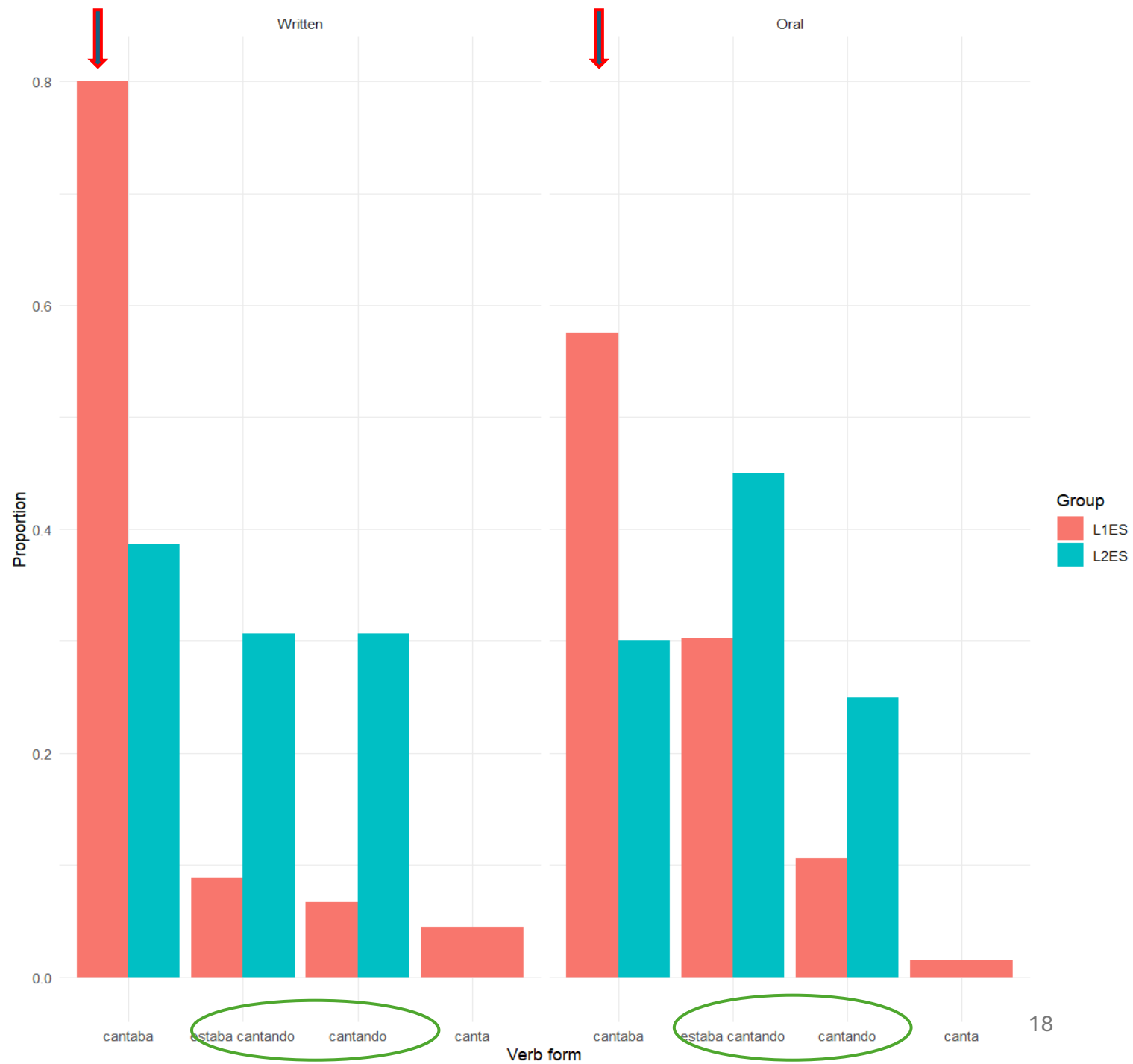
A lady saw that the girl stole IMP/was stealing a loaf of bread.

Estaba_cantando_L2:

(Est. 1.59, 95% CI [0.59, 2.64])

Cantando_L2:

(Est. 1.87 , 95% CI [1.01, 2.81])



Results

Modality: written vs oral

Estaba_cantando_Oral:

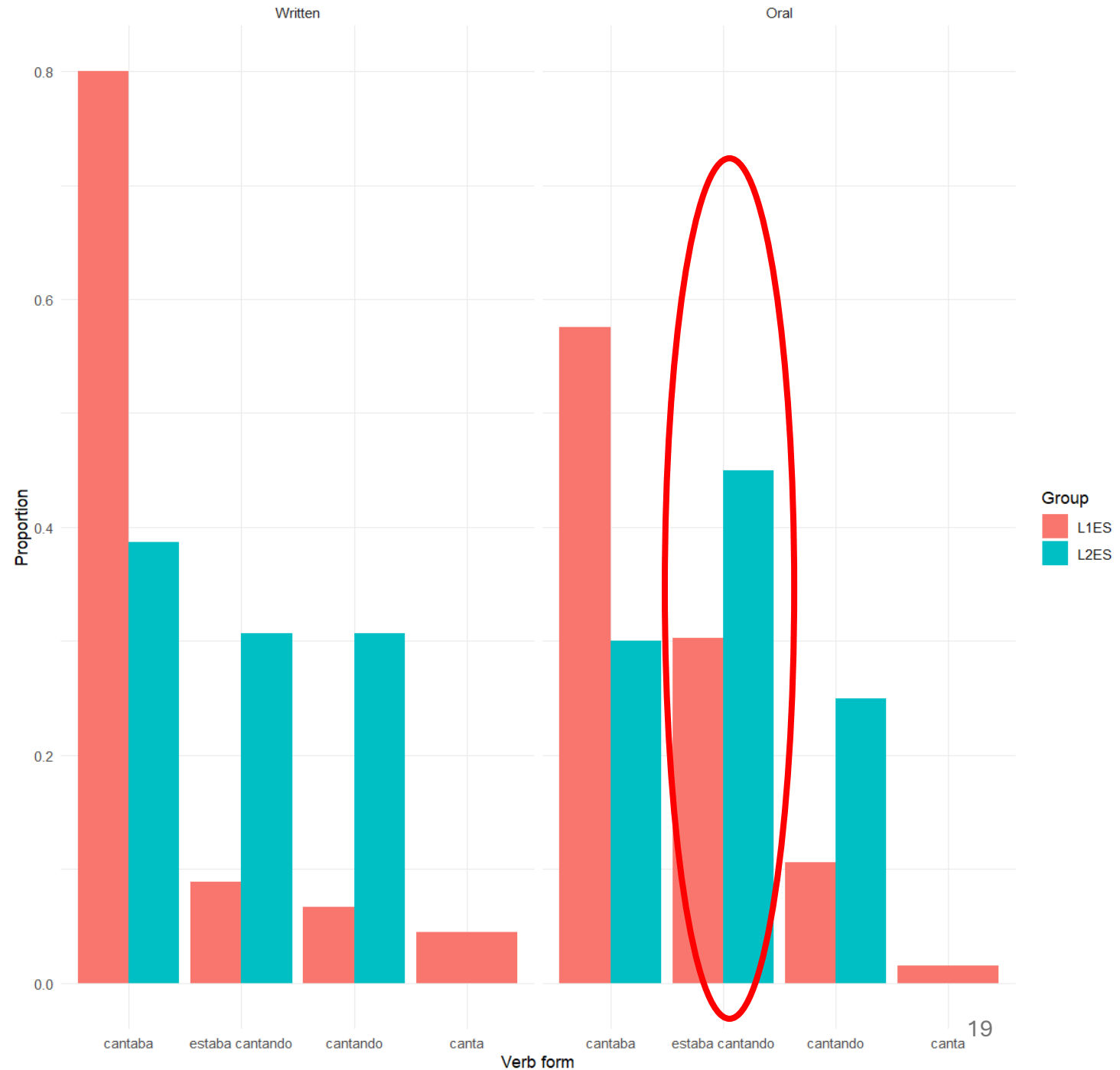
(Est. 1.27, 95% CI [0.53, 2.01])

Group × Modality interaction

Stronger effect for L1:

Native speakers increase *estaba cantando* substantially in oral mode.

L2 also use it a lot, but because they already overuse it in written mode, the oral boost is smaller.



Results

Anterior and narrative events



- El panadero dijo que la chica **robó** / **había robado** una barra de pan.

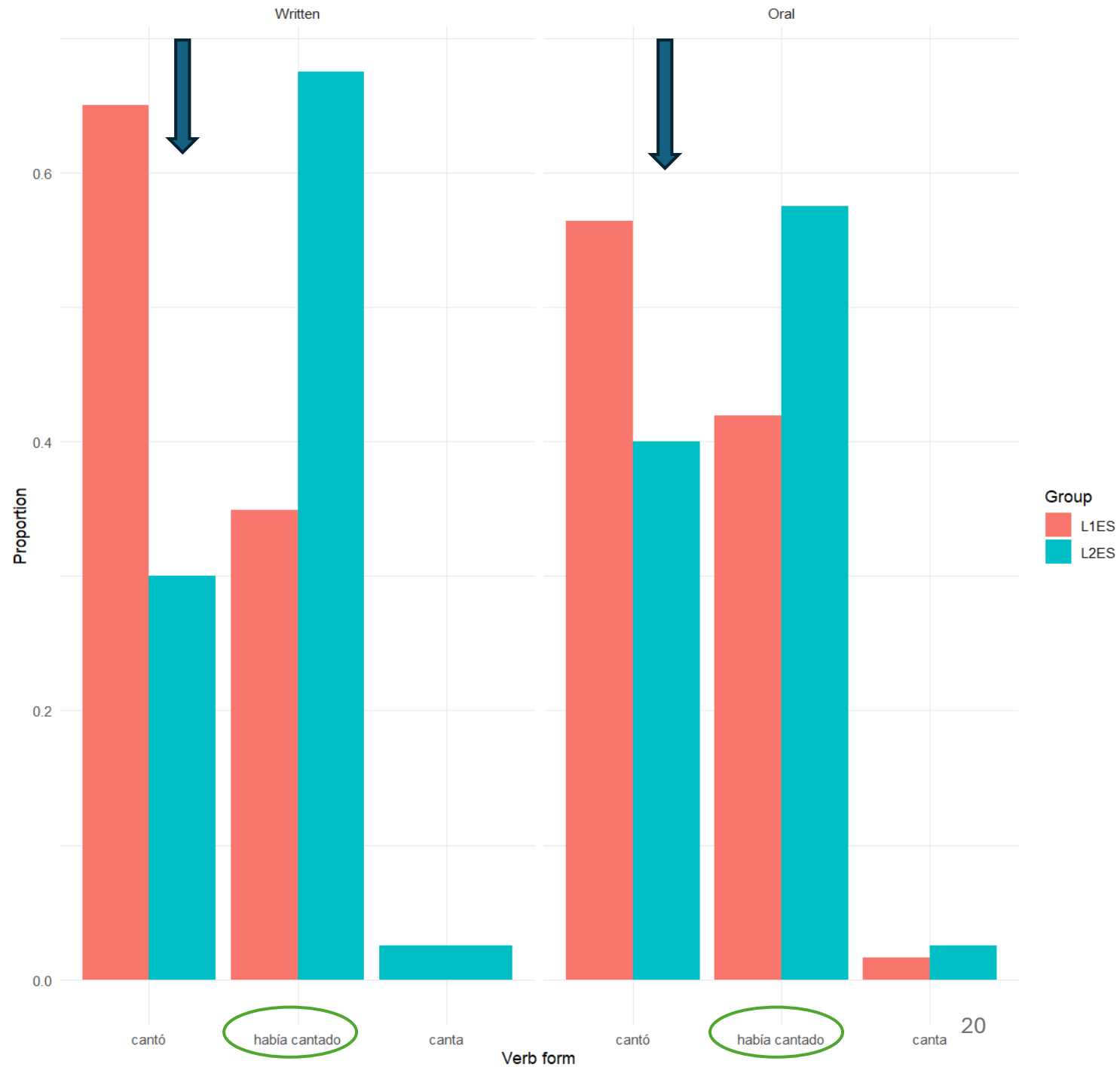
The baker said the girl stole/had stolen a loaf of bread.

Cantó_L2:

(Est. -0.90, 95% CI [-2.04, 0.29])

L2 learners have lower log-odds of using the preterit *cantó*.

This difference is suggestive but not conclusive.

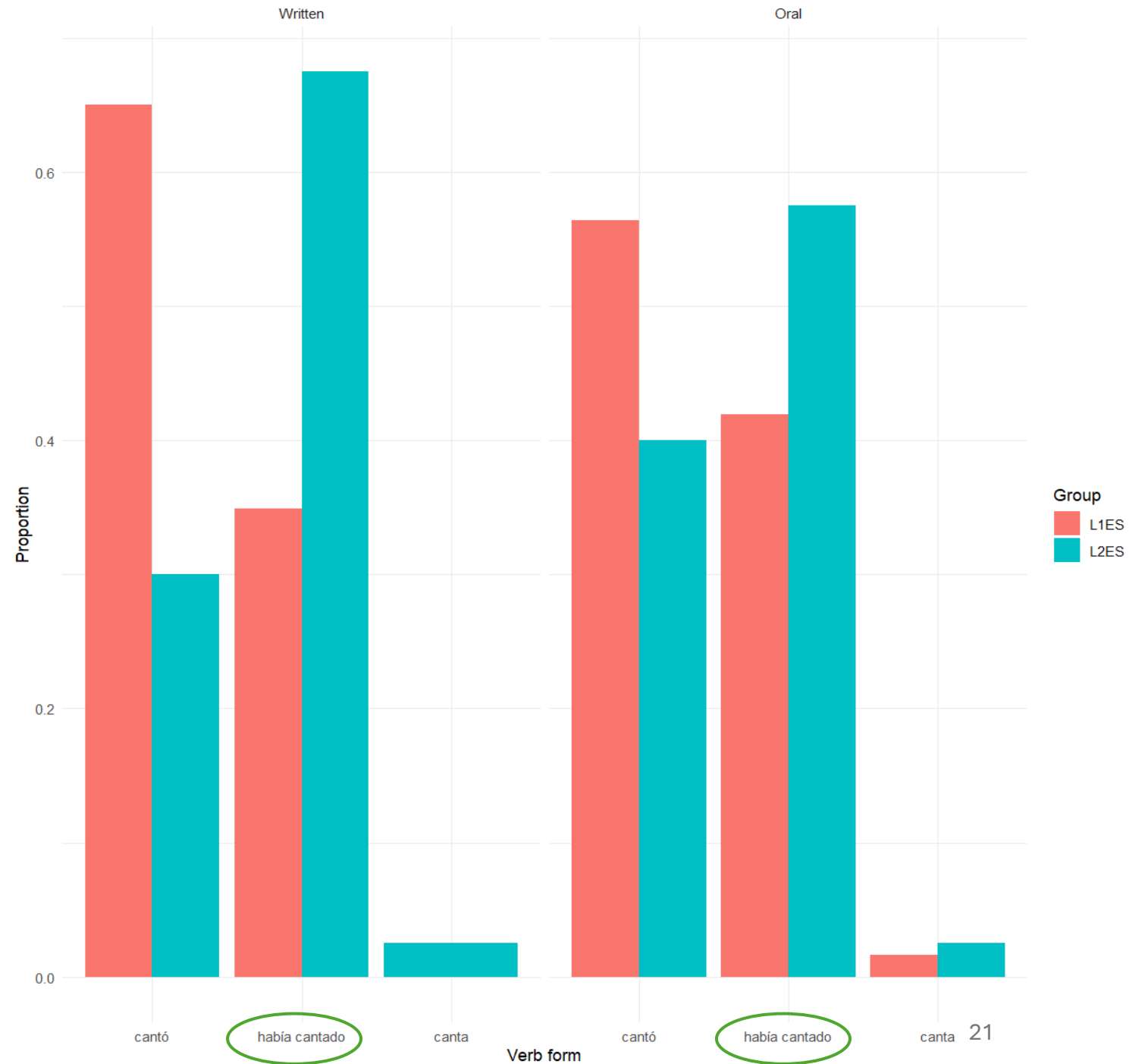


Results

Modality: written vs oral

Group × Modality interaction

The differences between oral and written modalities were not statistically robust.



Conclusion: pilot corpus analysis

- Overall, these findings suggest that observed differences between advanced learners and native speakers are better explained in terms of differences in the discourse-pragmatic control of verbal polysemy than by purely morphosyntactic limitations.
- Furthermore, more data is needed to confirm some tendencies and to be able to consider more predictable variables.

Conclusion

- These results inform the next stage of the project, since the manual analysis makes it possible to identify general tendencies and systematic differences between groups in the way temporal relations are expressed.
- These observations help anticipate the kinds of patterns that need to be captured in a hybrid or automatic approach, especially when dealing with learner data across proficiency levels.
- This provides a basis for developing more robust procedures for the automatic analysis of tense–aspect usage in L1 and L2 narrative corpora.

References

- Bardovi-Harlig, K. (2000). "Tense and aspect in second language acquisition: Form, meaning, and use". *Language Learning: A Journal of Research in Language Studies*, 50(1).
- Benveniste, É. (1966). *Problèmes de linguistique générale I*. Gallimard.
- Cruz Enríquez, M. (2019). *Función-significado-forma: Un modelo para el estudio de los tiempos verbales del español*. <https://papyrus.bib.umontreal.ca/xmlui/handle/1866/22635>
- Cruz Enríquez, Maura et Lareau, François (2025). [Narrative Discourse in Spanish : A Two Reference Points System for Verb Tense Representation](https://doi.org/10.21747/21833958/red16a12). *REDIS: Revista de Estudos do Discurso* (16), 335-375. <https://doi.org/10.21747/21833958/red16a12>
- Cruz Enríquez, Maura et Alba de la Fuente, Anahí (2024). [Past tenses in narrative production. How do advanced speakers of spanish as a second language compare to native speakers?. Semas. Revista de lingüística teórica y aplicada](https://doi.org/10.61820/zr58dg52), 5 (9), 27-46. <https://doi.org/10.61820/zr58dg52>
- Cruz Enríquez, Maura (2022). [Nueva propuesta bipartita de clasificación de los verbos según el aspecto léxico](https://doi.org/10.1038/s41562-017-0163). *Hispanic Studies Review*, 6 (2).
- Ellis, N. C. (2019). *Essentials of a Theory of Language Cognition*. <https://doi.org/10.1038/s41562-017-0163>
- García-Miguel, J. M. (2002). *ADESSE: Base de datos de verbos, alternancias de diátesis y esquemas sintáctico-semánticos del español* (Universidade de Vigo). <https://adesse.uvigo.es>
- González, P. (2003). *Aspects on Aspect. Theory and Applications of Grammatical Aspect in Spanish*. Universiteit Utrecht.
- Halliday, M. (1985). *An Introduction to Functional Grammar* ((1ra ed.).). Edward Arnold.
- Halliday, M. A. K., & Matthiessen, C. (2014). *An introduction to functional grammar*. Routledge.
- Python Software Foundation. (2025). *Python (Version 3.13.4)* [Logiciel]. <https://www.python.org/>
- R Core Team. (2022). *R: The R Project for Statistical Computing* [Logiciel]. <https://www.r-project.org/>
- Salaberry, R. (2011). "Assessing the effect of lexical aspect and grounding on the acquisition of L2 Spanish past tense morphology among L1 English speakers". *Bilingualism: Language and Cognition*, 14(02), pp. 184-202. <http://doi:10.1017/S1366728910000052>
- Slabakova, R. & Montrul, S. (2002). "On viewpoint aspect interpretation and its L2 acquisition". In R. Salaberry & Y. Shirai (Eds.), *The L2 acquisition of tense-aspect morphology* (pp. 363-395). Amsterdam: John Benjamins Publishing Company
- Weinrich, H. (1973). *Le Temps : Le récit et le commentaire*. Éditions du Seuil.

Thank you

Maura.cruzenriquez@teluq.ca

anahi.alba.de.la.fuente@umontreal.ca

francois.lareau@umontreal.ca