ISBN

9798191600987

Author

Roman Irizarry, Alexandra

Publication Date

2026-09-21

UNIVERSITY OF CALIFORNIA,
IRVINE

Morphosyntactic Variation in Spanish Heritage Speakers: Studies on Clitic Pronoun Processing and the Comprehension and Production of Gender-Inclusive –e Across the Animacy Hierarchy

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Education

by

Alexandra Román Irizarry

Dissertation Committee:
Professor Judith F. Kroll, Chair
Professor Elizabeth Peña
Professor Julie Washington
Professor Brandy Gatlin-Nash
Professor Connor J. Mayer
Professor Anne L. Beatty-Martínez
Professor Eleonora Rossi

2026

DEDICATION

A todas las personas que, con sus historias de vida, rompen con los esquemas tradicionales del lenguaje.

‘To all the people who, with their life stories, break with the traditional schemes of language.’

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

To my committee members, thank you all for the time and effort you dedicated to guiding me through this process. Thank you for believing in me and my potential as a language scientist, and for recognizing the importance of my work from our very first conversations. Your words not only comforted me during difficult times but also expanded my theoretical framework and positively shaped the outcome of this dissertation.

To Drs. Julie Washington, Elizabeth Peña, and Brandy Gatlin-Nash, thank you for being socially conscious language researchers in the School of Education. Your presence in the School gave me a space where I felt safe to explore and develop my original research ideas.

I would like to extend special thanks to my committee chair, Dr. Judith F. Kroll, for her unwavering support and mentorship over the past five years. Thank you, Judy, for taking me under your wing, for believing in the potential of studying gender-inclusive language in Spanish, and for pushing me to think critically about how this phenomenon intersects with bilingualism.

I also owe special thanks to Dr. Eleonora Rossi for generously sharing the materials from her original clitic pronoun study, and for her guidance and patience as I learned to work with the event-related potential technique.

To Dr. Anne Beatty-Martínez, thank you for your invaluable advice on how to best approach the study of gender-inclusive Spanish through the lens of grammatical gender asymmetries.

To Dr. Connor Mayer, thank you for your mentorship in navigating access to the Meta Content Library API. Although that work is not included in this dissertation, it provided me with a strong foundation for incorporating it into future research.

I would also like to extend my sincere gratitude to the undergraduate research assistants who contributed to this project. Their support in revising materials, assisting with participant recruitment, data collection, and data cleaning was invaluable. Their names, in order of recent involvement, are: María Torres, Ytzel Delgadillo, Yuxin Liu, Samantha Coronado, Su Lai Pyone, Sharon Chang, Francisca Castañeda, Yunzhi Zhang, Xinlei Zhu, Reina Samuels, Alexander Randall, Andrea Artunian, Andrea Sustaita, and Vanessa Quiroz.

Last but not least, I would like to extend my gratitude to my loved ones. To my domestic partner Victor, thank you for your unwavering support and love throughout the past ten years and for teaching me the importance of work-life balance. To my parents in Puerto Rico thank you for supporting me from 3,000 miles away and for welcoming me back home as I finished writing my dissertation. I love you all dearly.

Financial support was provided by National Science Foundation Graduate Research Fellowship Program Grant DGE-1839285 and the Deborah Lowe Vandell Dissertation Research Award from the School of Education at the University of California, Irvine.

VITA

Alexandra Román Irizarry

2026 Ph.D. in Education, University of California, Irvine

2024 M.A. in Education, University of California, Irvine

2021 B.A. in Foreign Languages and Hispanic Linguistics,
Universidad de Puerto Rico, Río Piedras

FIELD OF STUDY

Psycholinguistics, Cognitive Neuroscience of Language, Bilingualism, Gender-Inclusive Language

PUBLICATIONS

Román Irizarry, A., & Guzzardo Tamargo, R.E. (2025). Processing of Pronouns with Gender-Inclusive –x in Spanish: An Eye-Tracking Study. *Social Psychological Bulletin, 20,* 1-39. Special issue: The Gendered Language (R)Evolution: New Insights Into the Ever-Evolving Interaction Between Gender and Language. https://doi.org/10.32872/spb.13611

Román Irizarry, A., Beatty-Martínez, A. L., Torres, J., & Kroll, J. F. (2025). "Todes" and "Todxs", linguistic innovations or grammatical gender violations? *Cognition*, *257*, 1–25. https://doi.org/10.1016/j.cognition.2025.106061

Román Irizarry, A. (2021). ¿Todxs, todes, o todos? Actitudes de la juventud puertorriqueña hacia los morfemas de lenguaje inclusivo. In B. Llenín Figueroa (Ed.), *Proceedings of the VIII Coloquio ¿Del otro la'o?: Arte y activismo cuir en el Puerto Rico contemporáneo* (pp. 282 – 310). Editora Educación Emergente. https://www.researchgate.net/publication/361980392

Rodríguez Cruz, T., Acevedo Vélez, L. D., Román Irizarry, A. & Ríos Ghigliotty, J. N. (2020). Syntactic preferences in cases of code-switching by bilingual Puerto Rican college students of UPRRP. In N. Faraclas, R. Severing, C. Weijer, E. Echteld & W. Rutgers (Eds.), *Negotiating Crosswinds: Trans-linguality, trans-culturality and trans-identification in the Greater Caribbean* (Vol. 2, pp. 197-208). University of Curaçao.

ABSTRACT OF THE DISSERTATION

Morphosyntactic Variation in Spanish Heritage Speakers: Studies on Clitic Pronoun Processing and the Comprehension and Production of Gender-Inclusive –e Across the Animacy Hierarchy
by
Alexandra Román Irizarry
Doctor of Philosophy in Education
University of California, Irvine, 2026
Professor Judith F. Kroll, Chair

Heritage bilingualism is the most common form of bilingualism, yet it remains the most understudied. The existing literature on heritage languages has mainly employed offline measures, but these measures might not accurately reflect heritage language knowledge. This dissertation employs online methods to examine Spanish heritage speaker's knowledge of complex structures that are traditionally complex (i.e., clitic pronouns) or complex innovations (i.e., the use of –e as a gender-inclusive morpheme). Study 1 examined gender and number agreement in clitic pronouns by means of an acceptability judgment task with EEG recording. At the group level, both the behavioral and electrophysiological data showed heritage speakers processed gender and number together. Such a finding supports the representational economy claim (Scontras et al., 2018), in which heritage speakers shrink their feature inventory and prune their structure so that it includes fewer nodes. The bundling of number and gender is possible because these feature categories are close to each other nominally. Study 2 examined how Spanish heritage speakers comprehended (Experiment 1) and produced (Experiment 2) gender-inclusive –e across the animacy hierarchy. Experiment 1 employed a self-paced reading task with sentences manipulating nouns for gender (masculine vs inclusive) and animacy (humans, animals, and objects). Inclusive nouns resulted in longer RTs than masculine nouns, an effect that was smallest for human nouns (e.g., *les latines*). Higher Spanish dominance facilitated the processing of nouns with the masculine generic, while better inhibitory control reduced RTs across all conditions. Experiment 2 employed three narrative

tasks where heritage speakers had to tell a story regarding a day with their families and an event from their university's Latine Resource Center. The third oral narrative consisted of retelling a story with seeded instances of gender-inclusive –e. Although the masculine –o was the preferred choice, the few instances where –e appeared only referred to humans. This suggests that, despite limited exposure, Spanish heritage speakers can integrate –e as a gender-inclusive morpheme that encodes human referents. Taken together, the results suggest heritage grammars are flexible and open to restructuring traditional features, an openness that also seems to aid in the comprehension and production of innovative forms.

INTRODUCTION

Imagine walking into a busy restaurant in Southern California. The host greets you in English, and you respond in kind. But as you are guided to your table, you notice something remarkable: the family to your left is speaking Spanish, switching seamlessly to English when the waiter approaches. Behind you, a group of friends laughs together in Mandarin. Across the room, you catch snippets of Korean, Vietnamese, and Cantonese. Each table seems to exist in its own linguistic world, yet everyone navigates the shared space effortlessly: ordering in English, then returning to the home language or mixing both languages. In this moment, you are witnessing something extraordinary that's also completely ordinary, the daily reality of bilingualism. This daily reality of speaking one or more languages is true for more than half of the world's population (Grosjean, 2024). Despite how natural and representative this experience is, understanding bilingual minds has only been the product of recent work from the past 30 years (Kroll & Dussias, 2024). To reach a better understanding of how language functions as a linguistic, cognitive, and sociocultural phenomenon, research on bilingual populations is crucial. Furthering research on bilingualism and its consequences will further our understanding of how dynamic the mind and brain can be when speaking two or more languages. By studying how bilingual individuals navigate between linguistic systems, whether in a bustling restaurant or in their everyday lives, researchers can unlock insights into cognitive flexibility, neural plasticity, and the interaction between language and sociocultural contexts that would not be so clearly identified in monolingual contexts.

The main goal of this dissertation is to employ online methods to study how Spanish heritage speakers process two types of complex grammatical structures in the heritage language. Unlike offline measures, like questionnaires and acceptability judgments, which rely heavily on metalinguistic awareness to collect information after processing has occurred, online methods collect information on moment-by-moment comprehension (e.g., word-by-word or phrase-by-phrase) allowing researchers to tap into participants' implicit knowledge (e.g., Keating & Jegersky, 2015). Because heritage speakers are known to underestimate their abilities in the heritage language regardless of proficiency (e.g., Vargas Fuentes & Torres, 2026), heritage language researchers call for the use of online measures as a more suitable approach to examining implicit knowledge of the heritage language (e.g., see Jegerski & Keating 2023 for a review). By complex structures, I am referring to grammatical structures that were acquired later in life. One is a complex structure that already exists in the heritage language, grammatical gender and number processing in clitic pronouns, but that is acquired relatively late in the native language, while the other is an innovative morphological change in Spanish, the case of gender-inclusive morpheme –e (e.g., Latine), which may be acquired during young adulthood. This gender-inclusive morpheme serves a dual purpose: 1) to serve as a truly gender-neutral alternative to the masculine generic, and 2) to serve as a marker of a non-binary gender identity. The use of innovative morphemes (i.e., one of which is –e), although not explicitly tested before, appears to follow the animacy hierarchy. Spanish speakers appear to restrict innovative morphemes to humans (e.g., *les latines* 'the$_{\text{INC}}$ Latines') rather than animals (e.g., *les perres* 'the$_{\text{INC}}$ dogs$_{\text{INC}}$' [unattested]) or objects (e.g., *les botelles* 'the$_{\text{INC}}$ water bottles$_{\text{INC}}$' [unattested]). Therefore, knowledge of this innovatively complex structure will be assessed by testing the extent to which Spanish heritage speakers, who may have little experience with gender-inclusive morphemes, are sensitive to this observed pattern in comprehension and production.

The rest of the dissertation is structured as follows. First, Chapter I discusses cross-language dynamics in bilingual language processing with the purpose of setting the stage of important dynamics that are in play when bilinguals process information in either one of their two languages. Second, Chapter II problematizes the native speaker concept as it has been used in the literature and argues for better characterization of different native speaker types. The third Chapter introduces readers to the structural changes heritage languages undergo, while the fourth Chapter introduces important background information on the event-related potential technique, the method used in the clitic pronoun study. Chapter V contains the clitic pronoun study, while Chapter VI contains the study on the comprehension and production of gender-inclusive –e across the animacy hierarchy. The final chapter contains the general discussion and conclusion.

CHAPTER I: Cross-Language Dynamics in Bilingual Language Processing

The purpose of this chapter is to provide the reader with a general understanding of cross-language processing dynamics in bilingual language processing. By doing so, I hope to showcase how the bilingual experience can leave both languages 'open' to the influence of one another, as opposed to fixed unchangeable systems. More than thirty years of bilingualism research shows us that not only can the first language (L1) influence the second language (L2), but the L2 can also influence the L1 (Kroll & Dussias, 2024). This bidirectional relationship arises from the fact that a bilingual's two languages are always active, even when only one language is required and independent from modality (e.g., Kroll et al., 2015). Therefore, to successfully navigate the joint activation of both languages, bilinguals must regulate the use of each language by recruiting a network of cognitive control. Below, I review literature on lexical and syntactic processing in bilinguals that showcase how the two languages influence each other. The result is an openly flexible dual language system.

Lexical Processing in Bilinguals

Early models of bilingual lexical processing assumed the direction of the transfer between both languages happens from the native or dominant L1 to the second or less dominant L2. One of these models was the Revised Hierarchical Model by Kroll and Stewart (1994). In this model, the dominant L1 is used as a mediator for the less dominant L2. This was tested by having bilinguals translate words in two directions: from the L1 to the L2, known as forward translation, and from the L2 to the L1, known as backward translation. The results showed that forward translation was significantly slower than backward translation. Critically, only forward translation was sensitive to the semantic properties of the words being translated, suggesting that it requires accessing conceptual meaning. In contrast, backward translation appeared to rely on direct lexical connections from L2 to L1, bypassing deeper semantic processing. This supports the notion that in highly proficient unbalanced bilinguals the L1 serves as a mediator for the less dominant L2, with forward translation proceeding through meaning (L1 word → concept → L2 word) while backward translation uses more direct word-to-word links (L2 word → L1 word).

Aside from operationalizing language transfer effects as L1 mediation effects, researchers have also operationalized transfer as cross-language parallel activation. This capitalizes on the fact that, for bilinguals, the influence of the non-target language is evident even when reading or listening to words in one language only. To examine this, a lot of work has relied on lexical ambiguity, orthographical or phonological, across the two languages (e.g., Dijkstra et al., 1998; Marian & Spivey, 2003). When translations are cognates (i.e., convergence between meaning and form), a facilitation effect is typically encountered (e.g., *hotel* in English and Dutch). In contrast, when words are interlingual homographs or false friends, with similar form but different meaning (e.g., *room* means cream in Dutch), interference arises. This effect has also been found when words appear in sentential contexts (e.g., Libben & Titone, 2009), when translated words share the same meaning but not the same form (e.g., Thierry & Wu, 2007), during early stages of L2 acquisition (Bice & Kroll, 2015) and in highly proficient speakers (Schwartz et al., 2007). These findings not only provide evidence of cross-language interactions by showing robust effects regardless of modality, experimental setting, age of acquisition, and proficiency, they also exemplify how dynamic the language system is when an L2 is present. Finding that cross-language activation is not eliminated in sentence context (e.g., Schwartz & Kroll, 2006) also suggests that the findings at the lexical level are not restricted to the processing of words in isolated contexts but reveal the openness of the language system itself at both the word level and the sentence level.

Cognitive Control and Bilingual Lexical Processing

Cross-language transfer effects at the lexical level show bilinguals experience widespread competition between their two languages. This competition, which is characteristic of bilingual experience, requires that bilinguals regulate their two languages to enable fluent speech. One model that accounts for bilingual language regulation is Green's inhibitory control model (1998). This model argues that the more dominant language is inhibited to enable speech planning and production in the less dominant language. It has been supported by a body of research that has reported an asymmetric switch cost in language switching studies: bilinguals are slower to speak their L1 following the L2. In both behavioral (Meuter & Allport, 1999) and neuroscience studies (Misra et al., 2012), the evidence suggests that producing words in the L1 after speaking the L2 is associated with inhibitory processes. The explanation is that the more dominant L1 must be suppressed to enable the less dominant L2 to be spoken. When bilinguals have to then speak the L1, it requires upregulation which carries a processing cost. The issue of bilingual language regulation also raises the question of how regulation differs from domain general cognitive control.

Cognitive control is hypothesized to work in different ways (Braver, 2012). Two of these are proactive control and reactive control. While the former involves early selection of goal relevant information, the latter involves late correction mechanisms that are engaged once an unexpected event occurs. Another subdomain of cognitive control is inhibitory control, which involves suppression of distracting information. Language regulation engages a network of cognitive control, but not exactly in the way that executive function tasks reflect that network (Prior et al., 2017). Both contexts engage competition and conflict, but there is no one-to-one mapping. In cases of bilingual language competition, regulation depends on environmental context and individual differences in language experience. A theoretical account that attempts to explain this is the Adaptive Control Hypothesis (Green & Abutalebi, 2013).

The Adaptive Control Hypothesis states that language control processes adapt to the demands placed on them by the environment and the context of recurring patterns of conversational exchanges within a community of speakers. It distinguishes between three different interactional contexts based on how languages are used every day: a single language context, a dual language context, and a dense code-switching context. In the single language context, one language is used in one environment while the other is used in a distinct environment, and no switching between the two languages occurs. In the dual language context, both languages may be used with different speakers and language switching occurs within conversation but not within an utterance. In the dense code-switching context, both languages are used and speakers frequently mix both languages within a single utterance. According to each context, selecting a language for speaking may be done so cooperatively (i.e., dense code-switching) or competitively. In cases where the two languages compete, such as in single and dual language contexts, the demands on the cognitive system increase to be able to properly use the target language. However, the dual language system is hypothesized to further increase demands on the cognitive system because individuals must not only monitor the environment they are in, they must also be on the lookout for the arrival of a new interlocutor that speaks the other language.

Bilingual lexical processing studies support this hypothesis (Beatty-Martínez et al., 2020; Zhang et al., 2021). Beatty-Martínez et al., examined how three different groups of Spanish-English bilinguals differentially engaged cognitive control, as measured by the AX-Continuous Performance task (AX-CPT; Braver et al., 2001), when completing a category verbal fluency task and a picture naming task in Spanish and English. One group used their two languages separately and lived immersed in the L1 (separate context), another used their two languages interchangeably and lived immersed in the L1 with many opportunities to use the L2 (integrated context), and a third group was immersed in an L2 context with very few opportunities to use the L1 (varied context). The verbal fluency results supported the context-driven differences imposed by the language environments: those in the separated context produced more exemplars in Spanish, those in the integrated context produced a similar amount of exemplars in both languages, while those in the varied context produced more exemplars in English. Regarding performance on the AX-CPT, bilinguals' cognitive control performance aligned with the Adaptive Control Hypothesis: bilinguals in separated contexts relied more on reactive control, bilinguals in the varied context relied more on proactive control, while bilinguals in the integrated context exhibited varied cognitive control strategies. When examining the relationship between interactional context, cognitive control, and lexical access, they found that bilinguals in the varied context (L2 immersed) used proactive control mechanisms to maintain fluid lexical access in Spanish in an environment with very few opportunities to use it, while bilinguals from the separated context relied more on reactive control strategies when maintaining lexical fluidity in Spanish, a beneficial control strategy when keeping languages separate. Similarly, Zhang et al. (2021) compared highly proficient Mandarin-English bilinguals in China (L1 immersion) and the US (L2 immersion) by means of a language switching paradigm and the AX-CPT. They found an association between proactive control and increased L1 inhibition in the switching task for the L2 immersed group. Other studies have also found a link between the diversity of bilingual social networks and the recruitment of cognitive control (Gullifer & Titone, 2020; Gullifer et al., 2018). Taken together, these studies show how the high level of competition between the two languages gives rise to a language system that must adapt to dual language use in the environment via the recruitment of domain general cognitive resources.

Syntactic Processing in Bilinguals

At the syntactic level, studies on bilingual language processing show that exposure to an L2 can influence syntactic processing in the L1. Native language reconfiguration studies provide an example. Some of these studies have been conducted on syntactically ambiguous sentences and codeswitched sentences. For example, Dussias and Sagarra (2007) examined how extensive immersion in English would impact how Spanish-English bilinguals resolved syntactically ambiguous sentences in their L1, Spanish. The study used syntactically ambiguous relative clauses preceded by a complex noun phrase*: Alguien disparó al hijo de la actriz que estaba en el balcón*. 'Someone shot the son of the actress who was on the balcony.' When asked 'Who was on the balcony?', Spanish speakers typically respond, 'the son' while English speakers typically respond 'the actress.' Spanish-English bilinguals immersed in the L2 preferred the English interpretation when they were parsing sentences in their L1, Spanish, in comparison to a non-immersed proficiency matched bilingual group who preferred the Spanish interpretation.

Syntactic changes to the L1 can also be observed in bilinguals who codeswitch. Codeswitching is the systematic interleaving of two or more languages in a single utterance, requiring a high level of bilingual proficiency. For example, in Spanish-English bilinguals there is an asymmetric use of Spanish determiners in mixed noun phrases (NP; Beatty-Martínez et al., 2019). First, preference is given to Spanish determiners followed by an English noun (e.g., *el* fork 'the$_{\text{MASC}}$ fork') rather than English determiners followed by Spanish nouns (e.g., the *tenedor* 'the fork$_{\text{MASC}}$'). Second, Spanish determiners with masculine grammatical gender may follow English nouns whose Spanish translation is either grammatically masculine (e.g., *el* fork) or feminine (e.g., *el* spoon, 'the$_{\text{MASC}}$ spoon', spoon in Spanish is *la cuchara* 'the$_{\text{FEM}}$ spoon$_{\text{FEM}}$'). In contrast, Spanish determiners with feminine gender are not as common and only occur with English words whose Spanish translation is feminine (e.g., *la* spoon). This asymmetry has been documented in Spanish-English bilingual corpus data (e.g., Parafita Couto & Gullberg, 2017), psycholinguistic (e.g., Valdés Kroff et al., 2017) and neurolinguistic studies (e.g., Beatty-Martínez & Dussias, 2017). In an ERP study, Beatty-Martínez and Dussias examined how experience with codeswitching affected the processing of the grammatical gender asymmetry in mixed NPs. They found that habitual codeswitchers had difficulty integrating incongruent NPs such as *la* fork, an unacceptable combination in codeswitched speech, but not congruent NPs such as *el* spoon, an acceptable combination in codeswitched speech. In contrast, non-habitual codeswitchers elicited P600[1] effects, when reading congruent NPs like *el* spoon. These findings show that the bilingual comprehension system can adapt to speaker's codeswitching experience by modulating sensitivity to grammatical gender as an anticipatory cue.

Chapter Summary

The research presented in this chapter shows that a bilingual's two languages are open to one another at both the lexical and syntactic level. Such a notion challenges traditional views of the L1 as unchanging, while also questioning critical periods for syntactic learning by demonstrating how the native L1 can adapt to linguistic and cultural contexts (Hartshorne et al., 2018; Hernandez et al., 2021).

[1] The P600 is an effect associated with detection and repair of morphosyntactic violations. See chapters IV and V for more information on electrophysiological measures of language processing.

CHAPTER II: Redefining the Native Speaker Concept: Variation in Monolingual Speakers

In an Introduction to Hispanic Linguistics class I recently taught, most of my students were heritage speakers of Spanish. They were born into a Spanish-speaking household in the United States and learned Spanish at home as their L1. Despite being exposed to Spanish since birth, many of them did not feel confident enough to call themselves 'native speakers', which is a term they reserved for their parents and other Spanish speaking relatives in the heritage home country. Regardless of their self-perception as 'non-native' speakers, these students performed exceptionally well in class and could easily be differentiated from the advanced L2 speakers in the classroom. This anecdote mainly highlights two important aspects regarding the heritage language experience. First, that heritage speakers do not consider themselves native speakers of their heritage language, and second, that they perceive themselves differently from speakers born and raised in the heritage country. To shed light on these aspects, this chapter demystifies 'native speaker' ideals that have been strongly held in the literature, by showing how there can be different types of native speakers who can vary in their L1 knowledge. In doing so, I hope to illustrate how heritage speakers fall within a native speaker variation continuum which includes both monolinguals and multilinguals alike.

Deconstructing Native Speaker Ideals

In the literature, the term 'native speaker' has been defined in different ways, with some linking it to someone's language history since birth, and others conflating the term with monolingualism altogether (see Cheng et al., 2021 for a detailed discussion). One way to define the term 'native speaker' is to think of it as a status awarded to someone by other native speakers from the same speech community in a given place (Bonfiglio, 2010; Montrul, 2022). As intuitive as this might seem at first glance, place of birth does not capture the full spectrum of native speakers. What should instead be used to determine whether someone is a native speaker is: 1) exposure from birth to the language, and 2) growing up speaking the language (Montrul, 2022). These two criteria allow us to have different types of native speakers which are illustrated in Figure 1: monolingual native speakers and bilingual native speakers.

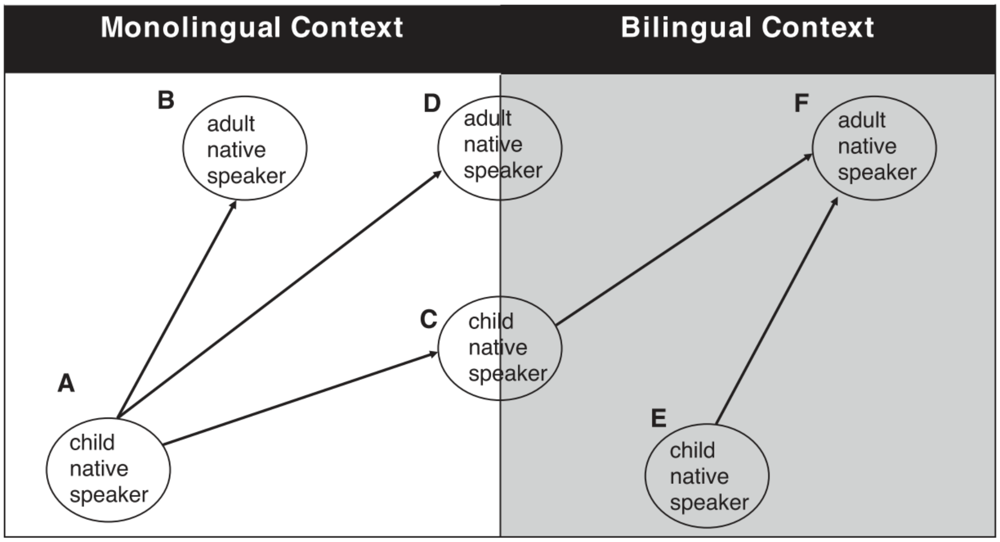

Figure 1: Native Speaker Typology (Montrul, 2022)

Within these native speaker types there are children who are born into monolingual contexts (child A) and stay in monolingual contexts throughout adulthood (adult B), and there are children who are born into bilingual contexts (child E) and stay in bilingual contexts throughout adulthood (adult F). Within this typology, movement from monolingual to bilingual contexts is also quite frequent. For example, a child born in a monolingual context (usually the home country) moves to a bilingual context during childhood and starts acquiring a second language (child C) and later grows up to become a bilingual adult (adult F). Similarly, a child may grow up in a monolingual context (child A) and move as an adult to a bilingual context where they start acquiring a second language as an adult (adult D). Taken together, the typology shows that being a native speaker can look differently than what the literature typically considers 'prototypical native speakers' (child A —> adult B), and that within bilingual native speakers some acquire both languages simultaneously (child E —> adult F) while others acquire their languages sequentially (child A —> child C —> adult F). Therefore, this typology allows researchers to classify heritage speakers as native speakers of their heritage language who may grow up either in a monolingual environment first or grow up in a bilingual context altogether, the latter which enables the acquisition of two native languages (Rothman & Treffers-Daller, 2014). Although Montrul's typology (2022) is useful for showing different native speaker experiences, bilingual upbringing should not be conflated with the heritage language experience alone, as there are language contact situations in the world where children acquire two languages simultaneously in an environment that supports the use of the two (e.g., French and English in Montréal, Canada, and Spanish and English in San Juan, Puerto Rico). It is also not uncommon for bilinguals to move between monolingual and bilingual contexts, such as bilinguals who live on both sides of the US-Mexico border (e.g., Hernández Santacruz et al., 2025) or to grow up in a bilingual environment and then move to a monolingual environment. What typically sets apart the heritage language experience from other bilingual upbringing experiences, is the lack of support of the heritage language in the broader society (more on how this environmental support affects heritage languages in chapter 3).

If all we need for a native speaker definition is exposure from birth and growing up speaking the language, why aren't heritage speakers considered by themselves and others as native speakers? The answer lies in how the term native speaker has been conceptualized. To understand how this term has been used in the literature, it is worth considering Noam Chomsky's (1965) ideal speaker-listener:

> Linguistic theory is concerned primarily with an ideal speaker–listener, in a completely homogeneous speech community, who knows its language perfectly and is unaffected by

> such grammatically irrelevant conditions as memory limitations, distractions, shifts of attention and interest, and errors in applying his knowledge of this language in actual performance. (p. 3)

While Chomsky's ideal speaker-listener is a platonic concept meant to isolate the study of language knowledge, and his definition does not identify monolinguals as the ideal speaker-listener, researchers have often operationalized the 'homogenous speech community' as referring to a group of monolingual speakers. To this point, Rothman and Treffers-Daller (2014) note "Despite the fact that most linguists would agree that the terms monolingual and native should not in principle be synonymous, the fact remains that linguists often use these terms interchangeably in writing, in practice, and in mindset" (p. 93). Under such operationalization, only adult B would be considered an ideal speaker-listener. According to Montrul (2022), the operationalized ideal speaker-listener is typically someone who: has been exposed to the language since birth, has completed schooling in the language (the period where linguistic consolidation occurs), and has reached a period of cognitive and linguistic maturity in young adulthood which leads to a 'stable' mental grammar, also known as 'ultimate attainment' (e.g., Birdsong, 1992). It is this stable mental grammar, also called interior language, which has been the primary concern of generative linguistics and has served for many years as the basis for our understanding of a theory of language (Montrul, 2022). However innovative the ideal speaker-listener concept was at the time, its operationalization proves problematic for a theory of language. First, the conflation of monolingualism with native speaker hood as the norm is not ecologically valid for a theory of human language given that most of the world is bi-/multilingual (Grosjean, 2024). Second, the assumption that ultimate attainment implies a stable/unchanging language that is independent from cognitive variables is not supported by empirical studies.

While traditional approaches to the study of language attribute different forms of language variation to performance, something they consider to be imperfect and non-essential to the core linguistic system, the emergentist framework offers a different view. Instead of treating variation as noise that obscures the true nature of linguistic knowledge, the emergentist approach treats variation as the central data that any comprehensive language theory must explain (Caldwell-Harris & MacWhinney, 2023). Such a centering on variation shows the native languages of bilinguals and monolinguals alike are not privileged fixed systems, but rather the knowledge of the native language can change according to different environmental factors. The subsection below reviews how some of these factors affect knowledge of the L1 in monolinguals. The presence of such variation in monolingual speakers highlights the importance of moving away from ideal speaker-listener criteria to instead focusing on understanding how different environmental factors affect L1 knowledge across different monolingual and bilingual groups.

Variation in Monolingual Native Speakers

There are many ways researchers can approach the study of language variation. Traditionally, linguists distinguish four dimensions of variation within a language (Coseriu, 1981). These are: diachronic (language variation throughout time), diatopic (language variation according to regions), diastratic (language variation according to social class), and diaphasic (language variation in different registers). These distinctions help language researchers study externalized language (E-language) as a window into the internalized language (I-language) and are typically used within Labovian variationist sociolinguistics (Montrul, 2022). In reality, all four dimensions of variation are present within speech communities and it is the duty of researchers to properly identify and characterize them. In this subsection, I review the effects of education level (diastratic), cognitive variables (diaphasic) and knowledge of another language (diastratic/diatopic) on monolingual language variation.

Education.

Most of the early work of education effects on monolingual language variation comes from studies conducted by Ewa Dąbrowska (2012). These studies show that adult monolingual native speakers do not share the same mental grammar. She does this by reviewing studies that examine two fundamental aspects of universal grammar: productivity of morphological constructions (Dąbrowska, 2008a, Dąbrowska, 2008b) and syntactic constructions from varying difficulties (Chipere, 2001; Dąbrowska, 1997; Ferreira, 2003; Dąbrowska & Street, 2006; Pakulak & Neville, 2010; Street & Dąbrowska, 2010). An important thing to know is that all the monolinguals included in her review met ideal speaker-listener criteria. In other words, they were 18 years of age or older, had no known disability, did not learn any languages growing up, and had completed at least some level of secondary education.[2]

For the productivity of morphological inflections, Dąbrowska tested a highly irregular morphological marking system, the Polish genitive singular (2008a), and a less irregular morphological marking system, the Polish dative singular (2008b), by means of a nonce word inflection task. In the genitive study, only two of the participants consistently applied the correct semantic rule for inflection, while the rest (n = 24) performed at chance. In the dative study, the participants' educational background varied from 8 to 20 years of formal education. Those with higher education performed better at the inflection task. This leads Dąbrowska to conclude that speakers extract different rules when learning to inflect irregular marking systems and that the nature of the extracted rules depends on educational level.

[2] Dąbrowska (2008b) was the only study in which the lowest level of education was 8 years of formal schooling. All other studies in the review had around 11 years of formal schooling.

Regarding syntactic constructions, Dąbrowska reviews studies with complex sentences (Chipere, 2001; Dąbrowska, 1997) and simple sentences (Ferreira, 2003; Dąbrowska & Street, 2006; Street & Dąbrowska, 2010; Pakulak & Neville, 2010). In her complex sentences study, Dąbrowska (1997) tested four different types of educational levels: an 'unskilled' group consisting of janitors and cleaners, undergraduates, postgraduate students, and lecturers from the same university in Glasgow, UK. The conditions of the study were: 1) subordinate clauses with a noun complement in the subject position, 2) 'tough movement' sentences, 3-4) two types of sentences with parasitic gaps, and 5) slightly longer control sentences with simple structures.

1) Paul noticed that the fact that the room was tidy surprised Shona.
2) John will be hard to get his wife to vouch for ___.
3) The nervous-looking student that Chris met ___ after being told his girlfriend wanted to jilt ___ took the 11 o' clock train.
4) It was King Louis who the general convinced ___ that this slave might speak to ___.
5) Alex decided that the easiest way to find out whether or not the plan would work was to ask the man who played the guitar at the party.

Participants heard the sentences and answered simple questions afterwards. Robust education related differences were found between the experimental conditions, but not for the control condition (Figure 2) with the unskilled group achieving the lowest accuracy and often performing below chance. Despite the robust education effects, Dąbrowska acknowledges the difficulty of the complex sentences might have placed heavy demands on the system, and the resulting differences could instead reflect differences in processing capabilities. A follow up study by Chipere (2001) addressed this issue by examining complex sentences in a Low Academic Achievement Group (LAA) and a High Academic Achievement Group (HAA). After

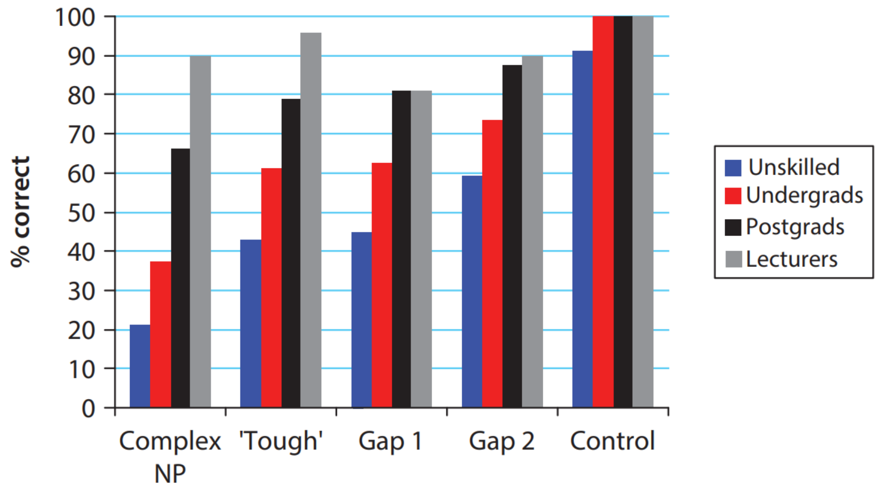

Figure 2: Education-related differences in comprehension (Dąbrowska, 1997)

hearing the sentences, participants had to repeat the whole sentence out loud and answer simple questions like the ones used by Dąbrowska (1997). The HAA group performed above 70% accuracy in recall and comprehension, while the LAA group obtained 60% accuracy during recall and 10% accuracy during comprehension. Then, to assess whether comprehension and recall performance was related to working memory, the LAA participants were randomly assigned to complete a memory training or a comprehension training. While the memory training improved recall and not comprehension, the comprehension training improved both outcomes. Thus, Chipere (2001) concludes that LAA performance is attributable to a lack of linguistic knowledge rather than limited processing capacities.

To examine whether educational differences in linguistic ability would arise with simpler sentences (i.e., single clause, no embeddings or long-distance dependencies), Dąbrowska and Street (2006) tested four types of simple sentences in two different groups: a less educated group (maximum completion was secondary school) and a highly educated group (at least 15 years of formal education). The conditions under study were plausible passives (‘The man was bitten by the dog’), implausible passives (‘The dog was bitten by the man’), plausible actives (‘The dog bit the man'), and implausible actives ('The man bit the dog'). This study was a partial replication and extension of Ferreira (2003) who tested undergraduates and found they performed at ceiling when reading active sentences, but struggled with passive sentences, particularly implausible passives for which they obtained a 74% accuracy, a process which Ferreira attributed to 'good-enough' processing. Aside from testing whether this difference differed by educational attainment, Dąbrowska and Street were interested in examining whether educational effects were also found depending on the type of experience with the structure. Therefore, they also recruited non-native speakers with low and high levels of educational achievement. Both high academic achievement groups and the non-native low academic group performed at ceiling across all conditions, while only the native low academic achievement group struggled with implausible sentences, particularly the passive ones. Dąbrowska and Street conclude these differences are attributable to having less experience with the particular structures, something the non-native low academic achievers did not due to other factors related to their experiences as L2 learners of English: greater exposure to a large number of exemplars in a short amount of time, enhanced metalinguistic awareness, and possibly greater motivation to learn a new language. In addition, these findings also challenge native like assumptions by showing non-native speakers can also achieve native like processing. They also highlight that not only is general education important for achievement in the L1, but that in cases of general low achievement, education in the L2 is of great importance when assessing competency in it. This is important to consider in the current study as heritage speakers can vary greatly in terms of the education they receive in their heritage language.

Another example of educational differences in the grammatical knowledge of understanding simple sentences is the experiments by Street and Dąbrowska (2010) which included the quantifier ‘every’ as well as active and passive sentences. The quantifier conditions included a locative variant (Q-is) ‘Every fish is in a bowl’ and a possessive locative (Q-has) ‘Every bowl has a fish in it’, while the other conditions contained a passive ‘The girl was kissed by the boy’ and its corresponding active ‘The boy kissed the girl.’ In experiment 1, HAA (at least 17 years of education) performed at ceiling across all conditions, while the LAA (at most 11 years of education) only performed at ceiling for the active condition. For the passive and Q-is condition, the LAA’s accuracy was 88% and 78%, while for the Q-has condition accuracy was below chance (43%). A second experiment, with a different sample of LAA whose results patterned with the LAA group of experiment 1, randomly assigned low scoring participants to a passive or a quantifier training. Results improved performance only for the trained conditions. Taking into account the findings across both experiments, Street and Dąbrowska (2010) conclude that engagement with particular structures is crucial for linguistic knowledge, but that not every instance of exposure is a learning moment, as LAAs would have encountered many exemplars of passive and quantifier constructions throughout their life. Instead, it is increased years of formal education which consolidates linguistic knowledge of said structures.

An example of processing differences in sentences with canonical word order as a function of education and proficiency level can be found in the event-related potential study conducted by Pakulak and Neville (2010). They auditorily presented canonical sentences such as ‘Timmy can ride the horse at his farm’ and sentences with insertion violations ‘Timmy can ride the horse at my his farm.’ The participants were university students and non-university students. Using a standardized measure, participants were split into a high proficiency group and a low proficiency group. These proficiency groups were positively correlated with education level and socioeconomic status. In their results, they find a larger and broader P600 amplitude for the highly proficient group, an electrophysiological component related to morphosyntactic sensitivity. Pakulak and Neville conclude that when examining language processing in monolingual native speakers, taking into account proficiency, education and socioeconomic status is important.

Cognitive Variables.

Other important variables that moderate language variation in monolingual speakers are working memory, a cognitive system that actively maintains and processes information in the face of distracting information (Conway et al., 2007), and cognitive control, a cognitive mechanism that includes inhibition for goal maintenance. Individual differences in these variables prove problematic for Chomskyan native speaker ideals which classify cognitive variables such as memory and attention as "grammatically irrelevant conditions." Moreover, the findings from ERP studies examining the role of working memory and cognitive control in monolingual native speaker processing of simple sentences are present despite speakers being college students, which should be comparable with Dąbrowska's highly educated groups (e.g., Beatty-Martínez et al., 2021; Kim et al., 2018; Zirnstein et al., 2018). Kim et al. (2018) found that monolingual native speakers of English showed differences in processing semantic anomalies: while some produced N400 responses (reflecting a conflict with world knowledge), others were more P600 dominant, an effect that differed by working memory. Similarly, the English monolinguals in Zirnstein et al., (2018) showed different ERP signatures according to predictability of the sentence, an effect that was moderated by cognitive control. In other words, inhibitory control reduced the processing costs of prediction errors. Finally, Beatty-Martínez et al. (2021) used ERPs to study whether masculine and feminine grammatical gender elicit qualitatively different brain responses. Individual's ERP responses differed according to the type of gender violation and were moderated by lexical knowledge for the feminine violations and by cognitive control for the masculine violations.

Learning Another Language.

The last piece of evidence I will discuss that provides evidence against stable native language ideals is changes in the first language after learning a second language. Aside from the studies that discuss how L2 exposure can influence syntactic processing in the L1 (see chapter 1), studies on language immersion provide insight as to how increased exposure to the L2 can influence the L1 (Bice & Kroll, 2015; Brice et al., 2021; Botezatu et al., 2022; Linck et al., 2009). One of the first studies to examine what happens to the L1 during immersion was conducted by Linck et al. (2009). They compared two groups of monolingual native speakers of English who were learning Spanish as their L2. One group was in their third month of study abroad in Spain, while the other group was enrolled in an intermediate level Spanish course and had no prior immersion experience. Both groups came from the same university and were matched in English and Spanish proficiency as well as working memory. The results showed immersed learners outperformed classroom learners in both comprehension and production in the L2. However, the immersed learners showed a reduction in the lexical access to their L1, which lead Linck and colleagues to conclude this was due to the L1 being inhibited during immersion.

A more recent study by Botezatu et al. (2022) examined whether the L1 performance effects that result from immersion in the L2 are present when the L2 is typologically different than the L1 or whether the effects are more related to the potential overlap between the languages. For this purpose, they recruited monolingually raised native speakers of English matched on cognitive resources and years of education who were learning a second language. Participants were recruited in two groups: a typologically different immersed group who had spent 6 months in China, and a typologically similar classroom group who was learning a romance language (Spanish or French). The immersed group produced: fewer words in the native language, more semantic errors, benefitted from higher lexical frequency items when retrieving the L1, and had less competition from phonologically dense neighbors. However, no differences in native language word reading or phonemic decoding were found between the two groups. This led the researchers to conclude that L2 immersion reduces access to the L1, but that some effects, particularly those related to reading abilities, depend on cross-language differences.

Neurolinguistic studies also provide evidence for changes in the native L1 both during early stages (Bice & Kroll, 2015) and long-term immersion in the L2 (Brice et al., 2021). Bice and Kroll (2015) examined what happened to the native language during early stages of L2 learning by means of the event-related potential technique. They included a group of functionally monolingual native speakers of English as well as native English speakers who were learning an L2 for the first time. The L2 bilinguals were either enrolled in a beginner or intermediate Spanish course. All groups performed a lexical decision task that included cognates and non-cognates. A reduced N400, which is associated with processing facilitation, was observed for the cognates in the L1, an effect that was present in both groups of learners with intermediate learners showing the greatest effect. At the individual differences level, slower reaction times in the L1 lexical decision task were observed as the N400 to cognates increased and as inhibitory control increased. These individual differences were not found for the monolingual group. Taking the results together, Bice and Kroll conclude that the L2 can affect processing in the L1 even as skills in the L2 are in early development. Regarding long terms effects of L2 immersion in the L1, Brice et al. (2021) conducted a longitudinal fMRI (functional magnetic resonance imaging) study that followed two groups of bilinguals who were recent immigrants up until they had resided in the L2 country for two years. One group was composed by Hebrew speakers learning English in the United States while the second group included American English speakers learning Hebrew in Israel. For both groups, the neural processing of the L2 was relatively stable despite proficiency and fluency gains. However, the neural processing of the L1 was more affected by L2 immersion. From this, Brice at al. conclude that L1 processing is not stable during adulthood and that reading strategies in the native language can shift over time as a result of the linguistic environment and specific language exposure an individual receives.

In summary, the goals of this chapter were: 1) to redefine problematic native speaker ideals, and 2) to show there is great language variation even in prototypical native speakers. As part of this process, I have used Montrul's native speaker typology (2022) to show how heritage speakers are a type of native speaker and should be treated as such in the literature. I have also showcased how variation in native language knowledge can be affected by education, cognitive variables, and knowledge of another language. This by no means is an exhaustive list that encompasses the full spectrum of native language variation, as the L1 can also vary by other factors (e.g., for L1 attrition a summary is found in Montrul, 2022; for 'gender' see Eckert, 1989; for 'ethnicity' see Rosa & Flores, 2017). The next chapter focuses on some of the changes heritage languages go through with the purpose of further understanding how different factors shape the heritage language experience.

CHAPTER III: Structural Changes in Heritage Languages

The previous chapter established heritage speakers as a type of native speaker. In this chapter, I provide a brief overview of some of the structural changes heritage languages undergo as a result of the heritage language experience based on Montrul (2022) and Polinsky (2018). Despite having exposure to the heritage language since birth, and fulfilling the native speaker criteria, heritage grammars can differ from prototypical native speaker grammars and the grammars from balanced bilinguals.

In the realm of phonological abilities, heritage speakers tend to perform well in both perception and production, with the former being more baseline ‘like’ (i.e., similar to prototypical native speakers) than the latter. In a perception study, Chang (2016) tested the extent to which perceptual benefits of the heritage language Korean, a language with mandatory unreleased stops, could help when perceiving unreleased stops in English, a language where released stops are considered canonical. Korean heritage speakers showed native-like perception of Korean unreleased stops while also outperforming native English speakers in the perception of American-English unreleased stops. In a perception and production study, Kim (2020) tested Spanish heritage speaker’s perception and production of Spanish lexical stress. Similar to Chang (2016), Kim found Spanish heritage speakers performed similar to baseline speakers and outperformed L2 learners in Spanish lexical stress perception. However, in lexical stress production, Spanish heritage speakers performed more similarly to L2 learners. Kim (2020) concludes that early exposure to a heritage language is beneficial for perception, but it does not guarantee baseline production.

Heritage speakers have been found to differ the most from prototypical native speakers in morphosyntax and the discourse-syntax interface. In her seminal study on Mexican-American adults in Los Angeles, Silva-Corvalán (1994) used sociolinguistic interviews aimed at eliciting simple and complex verb tenses across different Spanish verbal moods (i.e., indicative, subjunctive, and conditional). The heritage speakers in her sample lacked the past-perfect subjunctive, simple and complex conditionals, and confused the preterit and imperfect in the indicative. Simplification of the present subjunctive forms and the simple futures was also encountered. Regarding the discourse-syntax interface, heritage speakers overproduced overt subjects in null subject contexts and showed weakening of the pragmatic aspect of using overt subjects in topic shift contexts.

Since the publication of Silva-Corvalán's (1994) study, a lot of work has been done on the morphosyntactic changes different heritage languages undergo. A good overview is presented in Polinsky (2018). In her book, Polinsky outlines that morphological changes in heritage languages are related to perceptual salience, overregularization, and increased analyticity. Starting with perceptual salience, heritage speakers have a strong preference for morphologically salient forms. For example, in Russian, nouns like 'name' have a truncated nominative accusative form *imja* that use a perceptually heavier stem -*men*- in all other cases (e.g., *__imen__i* [genitive singular], *__imen__a* [nominative plural]). Russian heritage speakers prefer the use of the perceptually heavier stem -*men*- in the nominative accusative form over the truncated form (i.e., *imenja* over *imja*). In contrast when the forms are perceptually light, they may disappear completely or overgeneralize to a new context. An example of this is differential object marking (DOM), that is the different marking of animate and specific direct objects. In Spanish, this is done with the preposition *a*, a phonetically light and nearly inaudible marker when the object noun begins with a vowel. Spanish heritage speakers omit the *a* marker with high rates and accept its omission in comprehension (Montrul & Sánchez Walker, 2013; Montrul et al., 2015).

In contrast, Romanian has a phonetically more salient DOM marker *pe*, and Romanian heritage speakers retain *pe* and overextend its use to mark inanimate and other nonhuman objects (Montrul & Bateman, 2020a, 2020b). Two examples of overregularization can be seen in Arabic and Spanish. In Arabic, plural forms have a regular concatenative form (i.e., *fallaħ* 'farmer', *fallaħ-iin* 'farmers') and an irregular nonconcatenative form (i.e., *kalb* 'dog', *klaab* 'dogs'). Arabic heritage speakers have significantly more difficulty with irregular plurals than regular plurals, but when they produce an incorrect irregular plural, they overextend the most familiar irregular patterns (Benmamoun et al., 2014). In Spanish, heritage speakers treat the reflexive *se* as a marker of a well-formed predicate rather than a reflexive marker and thus produce it in ungrammatical contexts (e.g., *El lobo *se persiguió el conejo*. 'The wolf chased the rabbit'). Finally, when it comes to the use of more analytical (i.e., periphrastic) over synthetic (i.e., inflectional) constructions, heritage speakers seem to prefer analytical structures. An example of this preference can be found in Moroccan Arabic heritage speakers in the Netherlands (Boumans, 2006) who use more frequently the analytical genitive than the synthetic genitive.

Other examples of morphosyntactic vulnerabilities arise when relationships need to be encoded between constituents, such as establishing grammatical gender and case marking. For example, the supplementary material in Keating (2022) presents a summary of 16 studies examining grammatical gender agreement in children and adult Spanish heritage speakers. Thirteen of the studies reviewed found Spanish heritage speakers varied greatly when establishing grammatical gender agreement, especially when the noun was feminine (i.e., the marked grammatical gender), inanimate, and possessed a non-transparent ending (i.e., an ending that deviates from the typical masculine and feminine endings). For similar results with other heritage languages with grammatical gender, see Bianchi (2013) and Polinsky (2018).

When it comes to case marking, Polinsky (2018) discusses two tendencies that characterize case marking strategies in heritage languages. The first one was pointed out by O'Grady et al. (2011) as a systematic disregard for case marking, where heritage speakers reject case-marking variation and instead settle on one or two default forms when marking case in nouns. The choice of default forms seems to be informed by frequency and salience, not necessarily the nominative case. The second tendency is the use of adpositions in lieu of case markers. An example of this is the omission of the Korean accusative *-lul* being replaced by the directional postposition *hanthey*. Table 5.5 from Polinsky (2018) contains a long list of studies that document case-marking changes across different combinations of heritage and societal languages. The general trend for now seems to be that case attrition occurs across different language typologies.

Some final examples worth mentioning are those that concern heritage languages with little to no morphology. For these languages, it might seem like there would not be many changes in the heritage language, but this is not the case. For example, in Mandarin and Cantonese, the use of a classifier is mandatory when numerals or demonstratives are used, but heritage speakers of these languages tend to omit or use the wrong classifiers (e.g., Ming & Tao, 2008; Wei and Lee, 2001). Other examples from heritage Mandarin include the overgeneralization of the preposition *zài* and the inappropriate use or omission of the verbal marker *le* (Ming & Tao, 2008).

In summary, this chapter has presented a brief introductory overview on the structural changes heritage languages undergo. In doing so, I presented evidence that shows heritage languages are most vulnerable in the morphosyntactic domain. For more detailed overviews on the structural changes of heritage languages the reader can consult Polinsky (2018), Jegerski and Sekerina (2021), and Montrul (2022).

CHAPTER IV: Electrophysiological Measures of Language Processing

The purpose of this chapter is to briefly introduce the event-related potential technique along with the most common electrophysiological measures of language processing. Doing so, will better equip readers to understand why this online methodology was chosen to study clitic pronoun processing in Spanish heritage speakers.

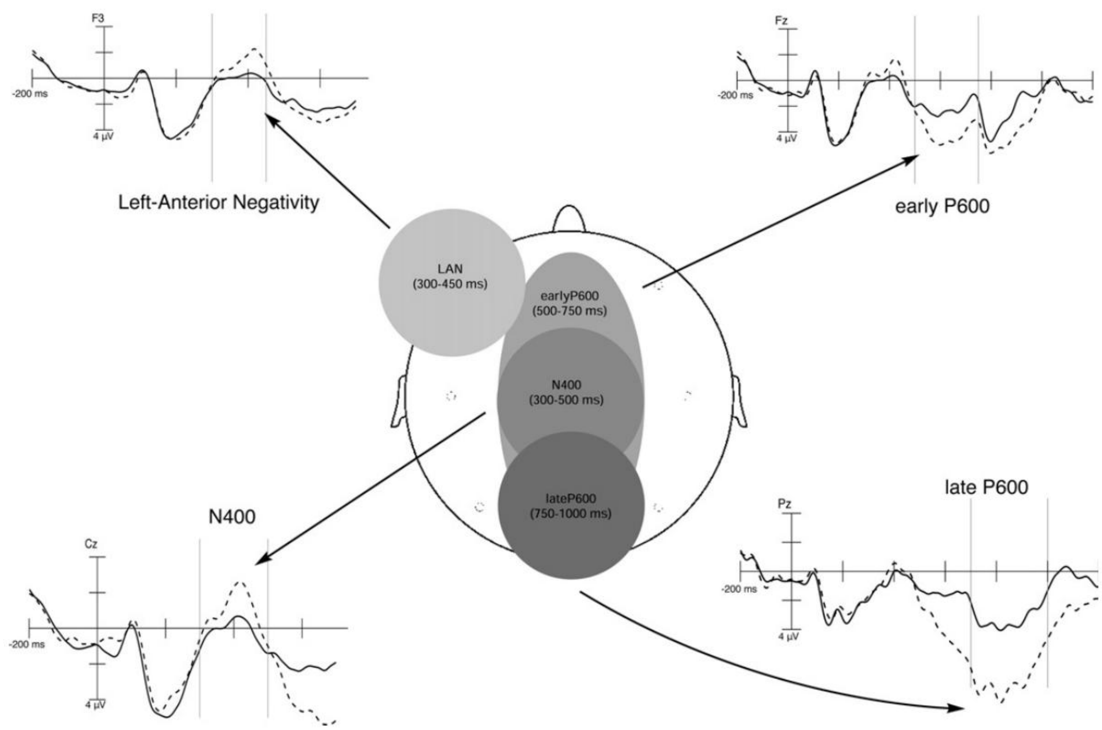

Figure 3: Main Language Related ERP Components (Molinaro et al., 2011)

Event-related brain potentials (ERPs) are scalp recorded voltage changes in the electroencephalogram (EEG) which are time locked to the onset of an event (Osterhout et al., 1997). They are considered an online measure of real time processing along with other behavioral measures such as RTs and eye tracking. However, unlike the previous behavioral measures which cannot collect data between the stimulus and response period, ERPs provide a continuous measure of processing that begins prior to the stimulus onset and extends past the response time (Luck, 2014). This allows researchers to zoom into subtle nuances in fine grained language processing and identify ERP components. Operationally, an ERP component is as a set of potential changes that are functionally related to an experimental variable (Donchin et al., 1978). Three of the most prominent language related ERP components are the left-anterior negativity (LAN), the N400 and the P600 (see Figure 3). The LAN is a left-frontal negativity associated with the early detection of a morphosyntactic violation usually reported in the 300 – 450 ms time window (e.g., Münte et al., 1997). LAN effects have also been found for grammatical sentences that require working memory resources (e.g., Kluender & Kutas, 1993), but unlike this working memory LAN which shows a sustained effect, the LAN to morphosyntactic violations tends to return to baseline after 450 ms (Fiebach et al., 2002).

The N400 is a negative-going waveform usually found in the 300 – 500ms time window that peaks at around 400 ms post-stimulus and is typically linked to lexical-semantic processing (Kutas & Federmeier, 2011). It can be elicited when researchers manipulate conceptual and linguistic retrieval of categorical or event-knowledge stored in long-term memory and its amplitude can be moderated by factors that affect retrieval from long-term memory such as frequency, novelty, and relatedness (Boudewyn et al., 2015). While both the LAN and the N400 appear in the same time window, they can be topographically dissociated: the former appears in left-anterior frontal regions, while the latter is strongest in central-parietal areas.

The P600 is a positive-going waveform that peaks around 600 ms post-stimulus when grammar or morphosyntax are manipulated and people are required to detect, reanalyze, or repair a sentence (Cafarra et al., 2015). It can also be found in grammatical sentences that contain syntactic ambiguities or processing difficulties (Fiebach et al. 2002; Kaan et al., 2000; Phillips et al., 2005). Some researchers distinguish between an early P600 in the 500 – 750 ms time window and a late P600 in the 750 – 1000 ms time window. In this study, I use the early P600 time window. Both the N400 and the P600 have unique yet overlapping scalp distributions in midline central and posterior electrodes, with the former being more anterior/central and the latter being more posterior.

CHAPTER V: Grammatical Gender and Number Agreement in Spanish Heritage Speakers: An ERP Study on Clitic Pronouns

As previous chapters have explained, heritage speakers are native speakers of their heritage language whose unique pattern of language acquisition leads to potentially different language outcomes than prototypical native speakers born, raised and schooled in the home country. This chapter focuses on the first study of the dissertation whose main goal is to study Spanish heritage speaker's processing of gender and number agreement in clitic pronouns, a more complex grammatical structure, by means of the event-related potential (ERP) technique, a technique with high temporal resolution well suited for measuring implicit language knowledge. In doing so, this study will contribute to our understanding of native heritage grammars. Below, I first introduce the agreement phenomenon generally, what it is and its different types as well as how it has been examined using ERPs. Then, I discuss ERP studies that have examined gender and number processing in prototypical native speakers. The section that follows introduces how heritage speakers process gender and number. Because the use ERPs to study heritage languages is fairly recent, I present findings from behavioral and other online methods (i.e., eye tracking) to supplement our understanding of how gender and number are processed in real-time in Spanish heritage speakers. Afterwards, the choice of clitic pronouns is justified. Finally, the goals and predictions of the study are outlined before detailing the methods, results and discussion of dissertation study 1.

Agreement

Agreement is the covariation of inflectional morphology between related words, and it is a phenomenon that affects 50% of all languages (Bybee, 1985). It involves variation of three main features: number, gender, and person (Wechsler, 2009). Number is related to the quantity of discourse referents and can be semantically interpreted in nouns. For example, English has two values, a singular 'unmarked' form (e.g., 'cat'), and a plural 'marked' form which is created by adding –s in nouns (e.g., 'cats'). In most languages, number agreement is reflected in the verb (e.g., 'The black cat**s** **are** sleeping'), while in Romance and Slavic languages number agreement also affects determiners and adjectives (e.g., *Lo**s** gato**s** negro**s** están durmiendo*). On the other hand, gender agreement manifests itself by classifying nouns into abstract categories. While European languages may have two categories denoted 'masculine' and 'feminine' (e.g., Spanish), or three gender categories denoted as 'masculine', 'feminine', and 'neuter' (e.g., German), other African languages have more complex gender systems (e.g., Wolof has 17 distinct gender values: Babou & Loporcaro, 2016). Grammatical gender and semantic sex may overlap in animate nouns (e.g., some words with masculine grammatical gender referring to male beings are *el gato* 'the[masc] cat[masc]', *el rey* 'the[masc] king[masc]'), but this is not always true (e.g., *la persona* uses feminine grammatical gender to denote 'the[fem] person[fem]', *el puma hembra* uses masculine grammatical gender to denote 'the[masc] female puma[masc]'). For this reason, grammatical gender and semantic sex/gender are separate things. In cases of inanimate nouns, grammatical gender is a pure morphosyntactic feature. Although it affects all nouns of a grammatical gender language, its classes are not evenly distributed. In Spanish, the masculine, typically marked with –o (e.g., *el libro* 'the[masc] book[masc]'), serves as the unmarked and default, while the feminine, typically marked with –a (e.g., *la silla* 'the[fem] chair[fem]'), serves as the marked and exclusive category. Evidence of the masculine serving as the default gender can be found in: mixed gender groups (e.g., *los niños* 'the[masc] children[masc]'), noun phrase gender conflict (e.g., *la rosa y el clavel blancos* 'the white[masc] rose[fem] and carnation[masc]), verb infinitives (e.g., *el saber* 'the[masc] knowledge[masc]'), compound words (e.g., *el sacapuntas* ‘the[masc] pencil sharpener[masc]’), loanwords (e.g., *el kayak* ‘the[masc] kayak), and codeswitched noun phrases (e.g., *el* spoon is more common than *la* spoon when referring to *la*[fem] *cuchara*[fem]). For a review that presents psycholinguistic and neurolinguistic evidence of how these gender asymmetries affect processing in Spanish see Beatty-Martínez and Dussias (2019).

Another feature that requires agreement marking in languages is person. Person is related to the role a participant has in a speech act. In English and Spanish, 1st person includes the speaker, 2nd person includes the addressee, and 3rd person excludes both speaker and addressee. Person agreement is also triggered at the verb level. Greenberg (1963) proposes all three features are organized hierarchically according to how often they appear across languages: person > number > gender. While some behavioral work has found evidence for this feature hierarchy (Carminati, 2005; De Vincenzi, 1999), a review of agreement processing using ERPs found no evidence of the feature hierarchy (Molinaro et al., 2011). Instead, Molinaro and colleagues found that person and gender violations were harder to reanalyze than number. Below, I review studies that assessed gender and number processing using ERPs.

Gender and Number Agreement in Prototypical Native Speakers: Evidence from ERPs

In this section, I first report the gender and number findings from Molinaro et al. (2011) which review agreement processing in 29 studies including prototypical native speakers from the early ERP literature (1983 – 2010). Then, I review findings from more recent ERP work on gender and number processing in Spanish.

In their review of number agreement mismatches, Molinaro et al. found 23 studies that focused on three types of agreement: subject-verb within-clause relation (*n* = 17), determiner-noun (*n* = 3), and noun-adjective (*n* = 3). A LAN-P600 to number mismatches was found in 74%

of the subject-verb agreement studies, and in 66% of the determiner-noun and noun-adjective studies. They also find that when number is not morphologically expressed or when it is expressed in another clause, the LAN is not triggered. Regarding gender agreement, Molinaro et al. reviewed 9 studies that focused on two types of agreement: determiner-noun (*n* = 6) or noun-adjective (*n* = 4). Out of these studies, 80% of them also showed a LAN-P600 to gender mismatches.

Since the publication of Molinaro et al.'s review, many ERP studies have been published on the topic of gender and number processing. In Spanish, these studies have compared gender and number processing (Alemán-Bañón et al., 2012; Alemán-Bañón & Rothman, 2016; Rossi et al., 2014) or have focused more on different nuances related to gender processing (Beatty-Martínez et al., 2021; Caffarra & Barber, 2015). Alemán-Bañón et al. (2012) examined the extent to which structural distance (e.g., within = *un edificio muy seguro* 'a very safe building', across = *el cuento es anónimo y el manuscrito también* 'the story is anonymous and so is the manuscript') impacts the processing of gender and number agreement between nouns and adjectives. They found gender and number violations elicited a P600 effect regardless of structural distance, but no LAN effect nor differences between gender and number violations were found. In a follow-up study, Alemán-Bañón & Rothman (2016) examined how asymmetry in feature values (i.e., feminine for gender and plural for number are marked features) affected agreement processing in noun adjective relative clauses (e.g., *un uniforme que parecía sucio* 'a uniform that seemed dirty'). Gender and number violations elicited the typical LAN-P600 response. Further examination of the P600 response showed it emerged earlier when the feature was marked (i.e., feminine or plural), and that it was strongest for number marked conditions.

Rossi et al. (2014) also examined gender and number processing, but they did so in clitic pronouns, the target structure of the current study. Their goal was to test the extent to which advanced L2 Spanish speakers could acquire features (i.e., grammatical gender) not present in their native language (i.e., English). For this purpose, they recruited native Spanish speakers and advanced L2 Spanish speakers. The native speakers elicited a P600 response to gender and number mismatches, while the L2 speakers only elicited a P600 to number mismatches. When taking proficiency into account, only individuals with very high proficiency (i.e., accuracy score > 0.90) exhibited a P600 response to gender mismatches. Rossi et al. conclude that native like processing can be achieved in very highly proficient L2 learners even when the features are not present in their L1 (see Alemán Bañón et al., 2017 for similar findings in L2 Spanish speakers).

Regarding gender processing, Caffarra and Barber (2015) examined whether the linguistic system takes advantage of formal gender cues during sentence processing by examining determiner noun agreement in transparent versus opaque nouns. Transparent nouns have reliable gender-related endings that follow the –o masculine and –a feminine grammatical gender pattern in Spanish (e.g., *el libro* ‘the$_{\text{MASC}}$ book$_{\text{MASC}}$’), while opaque nouns have ambiguous endings (e.g., *el tren* ‘the$_{\text{MASC}}$ train$_{\text{MASC}}$’). Caffarra and Barber found greater central-anterior negativities to transparent nouns and a biphasic LAN-P600 to both types of gender violations. From this, they concluded that noun transparency affects processing, but it emerges early and does not seem to affect agreement processing. Differential processing of masculine and feminine agreement has also been examined at the determiner noun level (e.g., *el vestido* ‘the$_{\text{MASC}}$ dress$_{\text{MASC}}$’ vs **la vestido* ‘the$_{\text{FEM}}$ dress$_{\text{MASC}}$’). Beatty-Martínez et al. (2021) examined whether the asymmetrical distribution of grammatical gender in Spanish (Beatty-Martínez & Dussias, 2019) produces qualitatively different neural responses in agreement processing. They found masculine noun violations (e.g., **la vestido* 'the$_{\text{FEM}}$ dress$_{\text{MASC}}$') elicited the biphasic LAN-P600 response, while feminine noun violations (e.g., **el cartera*, 'the$_{\text{MASC}}$ purse$_{\text{FEM}}$') elicited a P600 response. Regarding individual differences, masculine noun violations were moderated by inhibitory control while feminine noun violations were moderated by lexical knowledge. Taken together, these findings show that linguistic features from the same class (i.e., grammatical gender) exhibit patterns of brain activity that correspond to their distributional use.

Gender and Number Agreement in Heritage Speakers: Evidence from ERPs, Psycholinguistics, and Offline Studies

Compared to ERP studies on prototypical native speakers, ERP studies on heritage speakers are currently emerging. In Spanish, Bice and Kroll (2021) offer insights on subject-verb number agreement while Luque et al. (2023) and Hao et al. (2025) provide insight on gender processing (for processing of complex Wh- questions see Martohardjono et al. 2017). Bice and Kroll tested subject-verb number agreement (e.g., *Este sobre contiene información muy importante*, vs *Este sobre *contienen información muy importante* "this envelope *contains/contain very important information") and found a marginally significant P600 to number mismatches. Follow up analysis showed that as proficiency[3] in the heritage language increased, there was a decrease in the N400 response and an increase in the P600 response to number mismatches. Luque et al. (2023) examined grammatical gender agreement processing between the noun and adjective. They specifically compared transparent nouns (e.g., *Mateo visitó un pueblo pequeño con sus amigas.* 'Mathew visited a$_{\text{MASC}}$ small$_{\text{MASC}}$ town$_{\text{MASC}}$ with his friends$_{\text{FEM}}$.) against opaque nouns (e.g., *Carla pidió un postre dietético.* 'Carla ordered a$_{\text{MASC}}$

[3] Bice and Kroll (2021) created a composite measure of proficiency using measures of: production fluency (verbal fluency), comprehension (d' scores on grammaticality judgment task), self-rated proficiency averaged for speaking, understanding, and reading, and age of acquisition.

low-calory$_{MASC}$ dessert$_{MASC}$). At the group level, they found a biphasic N400-P600 to gender mismatches regardless of noun transparency. Luque and colleagues claim this new biphasic pattern reflects the notion that heritage speakers are sensitive to overt morphology: while the P600 indexes errors in agreement, the N400 seems to reflect enhanced sensitivity to morphology due to grammatical gender's lexical nature. To further investigate the individual differences that could lead to this biphasic N400-P600 and explore whether it varied between unmarked (e.g., **un silla* '*a$_{MASC}$ chair$_{FEM}$') vs marked (e.g., **una libro* '*a$_{FEM}$ book$_{MASC}$') agreement errors, Hao et al. (2025) reanalyzed the data from their previous study (Luque et al., 2023). They found that higher Spanish proficiency, indexed by the LexTale, increased the P600 response to both error types, but showed no relationship with the N400 response. The lack of an inverse relationship between proficiency and the N400 (i.e., increased proficiency reducing the N400), supports the notion the authors made in their original paper. In other words, at least some participants employ both an N400 and a P600 route when processing gender agreement errors. For unmarked errors, a stronger N400 was found as Spanish exposure and use increased, while for marked errors as formal instruction in the heritage language increased, there was an increase in the N400 response and a decrease in the P600 response. Hao et al. argue that exposure and use of the heritage language enhances sensitivity to morphological regularities, while formal schooling seems to enforce pattern-based processing of less salient patterns while reversing markedness in grammatical gender processing.

Taken together, the emerging ERP research on gender and number processing seems to suggest that Spanish heritage speakers are qualitatively similar to prototypical native speakers in their agreement processing. This finding is also supported by online psycholinguistic studies. In number processing, most of the work has focused on subject-verb number agreement (e.g., Foote, 2011; Jegerski & Fernández Cuenca, 2025; Sagarra & Rodríguez, 2022). In general, this work has found Spanish heritage speakers are sensitive to subject-verb number agreement, with some finding the effect is strongest when agreement is adjacent (Foote, 2011), while others find this sensitivity is independent of the distance between the agreement source and the agreement target (Jegerski & Fernández Cuenca, 2025). Moreover, the sensitivity to subject-verb number agreement seems to be independent of age of acquisition and proficiency (Sagarra & Rodríguez, 2022).

Regarding gender agreement, while studies using offline methods show heritage speakers exhibit a great amount of variation when compared to prototypical native speakers (see the supplementary material in Keating, 2022 for a review), online methods show a different picture. A lot of this evidence comes from eye-tracking studies, especially those with facilitative processing. In these studies, participants view pairs of pictures as they hear a sentence instructing them to find a target object (i.e., *Encuentra el*… 'Find the[masc]…', *Encuentra la*… 'Find the[fem]…'). When the grammatical gender of the determiner is informative of the target's gender, facilitative processing occurs if participants orient their eyes quicker towards the target object, than when the determiner is not informative of the target's gender. Fuchs (2021) found that while heritage speakers were overall slower than prototypical native speakers, they fixated faster on trials where the gender of the article was informative versus when it was not informative, suggesting heritage speakers also rely on grammatical gender to facilitate processing. This confirms findings from previous studies[4] that included child and adult heritage

[4] In these studies, the L1 children and L1 adults included (except for the parents in the 2007 study) fall under the criteria of Spanish heritage speakers, meaning they grew up being exposed to mostly Spanish in the US. In Lew-Williams and Fernald (2007), the children were on average 37.7 months old and had 85% exposure to Spanish. The children from Lew-Williams (2017) were in dual language immersion schools with 50% of Spanish instruction. Finally, Lew-Williams and Fernald (2010) compared adult heritage speakers to the Spanish heritage speakers from their first study (2007).

speakers but did not compare them to prototypical native speakers from the same age group (Lew-Williams, 2017; Lew-Williams & Fernald, 2007, 2010). Furthermore, it underscores the fact that both groups of native speakers can exploit gender cues during processing. Sensitivity to noun adjective agreement has also been found in Spanish heritage speakers, with one eye-tracking study (Keating, 2022) finding this effect regardless of agreement proximity, and another self-paced reading study finding reduced sensitivity in non-adjacent conditions (Foote, 2011).

Regarding the processing of gender and number together, prototypical native speakers have been found to process each feature independently (e.g., Barber & Carreiras, 2003; Fuchs et al., 2015). In heritage speakers, the work of Scontras et al. (2018) and Fuchs et al. (2026) stand out. Scontras et al. (2018) employed metalinguistic judgments to examine agreement errors in gender and number by means of an agreement attraction paradigm. They found that heritage speakers bundle the processing of gender and number together, a process they attributed to representational economy. According to representational economy, heritage speakers shrink their feature inventory to include fewer nodes and bundle gender and number together because the two features are close to each other on the nominal scale, a process that follows predictable structural changes. However, a recent visual world eye-tracking study by Fuchs et al. (2026) found that heritage speakers process gender and number independently to facilitate the processing of upcoming words. Nevertheless, when only one feature uniquely identified the target, prototypical native speakers attended more to the gender feature while heritage speakers attended more to the number feature.

Clitic Pronouns

The review of online studies examining gender and number processing in heritage speakers seems to suggest that they are not that different from prototypical native speakers.

Something to note is that most of the gender work has been conducted within the noun phrase, with one study testing embedded clauses (i.e., Scontras et al., 2018), while number agreement seems to be widely studied in subject-verb agreement. While this is informative, it remains to be seen how Spanish heritage speakers process gender and number in more complex grammatical structures like clitics. Clitics represent a difficult structure to acquire for all types of speakers, including monolinguals as well as early and late bilinguals (Rossi et al., 2014; Satterfield, 2021). In L1 acquisition, object clitics are acquired relatively late, at around 22 months of age (e.g., Guasti, 1994; Kaiser, 1994; Domínguez, 2003). They are also highly vulnerable in clinical populations (e.g., for aphasia see Sánchez-Alonso et al., 2011, for children with developmental language disorder see Castilla-Earls et al., 2023). According to Satterfield (2021), these structures are complex because cliticization involves a complex integration of: general syntactic knowledge, language-specific knowledge of lexical items (e.g., grammatical gender), interactions between interfaces, and a continuity of extra-grammaticality factors (i.e., processing complexity, input frequency). Accusative clitic pronouns, which replace a direct object, distinguish between two processes: proclisis (b), where the clitic pronoun precedes simple finite verbs, and enclisis where the clitic pronoun may only follow a finite verb when the imperative is used (c). Non-finite verbs may also be used with clitic pronouns, but only enclisis is possible in these cases (d). Expressions such as *leyó lo* and *lo leer* constitute violations of these rules and are ungrammatical. The morphosyntactic knowledge required for correct clitic pronoun use also requires speakers to match the pronoun with the grammatical gender and number of the direct object (i.e., masculine singular in examples a-d).

a) *Alex leyó <u>el libro</u>*. 'Alex read <u>the book</u>.'

b) *Alex <u>lo</u> leyó*. 'Alex read <u>it</u>.'

c) *Alex, ¡léelo!* 'Alex, read it!'

d) *Alex vendió el libro sin leerlo.* 'Alex sold the book without reading it.'

Some scholars argue the exact frequency of proclisis versus enclisis use is dialect specific (Sinnott & Smith, 2007). Corpus-based work by Davies (1995) found proclisis was used more often in oral speech, but its frequency usage did not differ across ten different Spanish dialects. Regarding heritage speakers, mixed findings exist on whether they use one process more over the other when compared to prototypical native speakers. Montrul (2010) finds heritage speakers use more proclisis than enclisis when compared to L1 monolinguals and advanced L2 learners, while Pérez-Leroux et al. (2011) finds heritage speakers prefer more enclisis, and Shin et al. (2017) and Requena and Dracos (2018) find no difference between heritage speakers and community norms. In this study, I examine clitic pronoun processing when proclisis is used.

The Current Study

The current study addresses a gap in our understanding of how grammatically complex structures are acquired during native language acquisition in heritage language contexts. Specifically, it will contribute to the literature that seeks to understand how multiple factors related to the heritage language experience affect real time processing of agreement features. For this purpose, I ask two main questions: 1) How do Spanish heritage speakers process gender and number in clitic pronouns?, 2) How do individual differences in Spanish proficiency, language use, schooling, and language dominance affect heritage speakers' processing of gender and number in clitic pronouns?

Given the wide range of variability heritage speakers possess in the heritage language, I predict sensitivity to gender and number agreement will vary as a function of individual differences. In other words, as Spanish proficiency, Spanish use, schooling in Spanish, and

Spanish language dominance increase, so will sensitivity to gender and number. Behaviorally, these individual differences should predict higher accuracy and faster reaction times (RTs) for both type of mismatches, while electrophysiologically they should predict greater LAN and P600 effects to mismatches. Given the findings from previous ERP work on Spanish heritage speakers, higher Spanish use and higher Spanish schooling should also predict greater N400 effect to agreement mismatches.

Methods

Participants

Thirty-seven[5] heritage speakers of Spanish from Southern California were recruited and compensated for participation from the student population at the University of California, Irvine in accordance to IRB approval. To participate in the study, participants had to: be 18 years or older, have grown up speaking Spanish at home in the U.S., have corrected-to-normal vision, have no serious eye or neurological conditions, and have no prior history of head trauma. One participant was excluded from the analysis due to incomplete behavioral data, while another was excluded for being a heritage speaker of an indigenous language who learned Spanish as their L2. The remaining sample (n = 35) had an average age of 20.7 years (SD = 1.8) and reported using more English (M = 6.1, SD = 0.7) than Spanish (M = 3.7, SD = 1.3). Regarding gender, 6 self-identified as male, 28 self-identified as female, and 1 self-identified as non-binary.

Materials

Diploma de Español como Lengua Extranjera (Adapted Test)

The *Diploma de Español como Lengua Extranjera* (DELE; Diploma of Spanish as a Foreign Language) is a standardized proficiency exam developed and administered by the

[5] An additional 19 heritage speakers recruited at a California university will be incorporated into this dataset prior to publication.

Instituto Cervantes on behalf of the Spanish Ministry of Education. The exam assesses learners' overall proficiency in Spanish and is aligned with the proficiency levels established by the Common European Framework of Reference for Languages. The adapted test used in this study was based on materials from the DELE C1 examination and included three sections. The first section was worth 20 points and assessed reading comprehension through a gap-filling task where participants completed a continuous text by selecting the most appropriate lexical or grammatical option for each blank. The second section was worth 10 points and assessed vocabulary by asking participants to identify the meaning of low-frequency words, idiomatic expressions, and fixed phrases presented in sentences. Finally, the third section was worth 20 points and assessed grammatical proficiency through multiple-choice items that target morphosyntactic structures characteristic of upper-intermediate and advanced Spanish proficiency (i.e., verb tense and mood selection, subjunctive use, pronoun placement, relative clauses, conditional constructions, and discourse connectors). Each item was worth 1 point, for a maximum score of 50. On average, the sample included in the study obtained an average DELE score of 31.2 (*SD* = 6.5).

Language History Questionnaire

A short online version of the Language Experience and Proficiency Questionnaire (LEAP-Q) was used to assess participants' language background (Marian et al., 2007). The complete language background information for the analyzed sample can be found in Table 1. The questionnaire was also modified to include questions related to language brokering experience, that is the experience of serving as a translator or interpreter without formal training. Almost all of the participants reported serving as language brokers during childhood (91%) and in the last 6 months (89%). Following the approach by Hao et al. (2025), the average self-

Table 1: Language History and Language Proficiency for Clitic Pronoun Study Participants

	Spanish	English
Age of acquisition (in years)	.0 (.0)	4.1 (3.1)
Years in a country where ______ is spoken	7.8 (8.8)	18.6 (3.8)
Years in a family where______ is spoken	20.8 (1.8)	13.1 (8.8)
Years in a school/work where ______ is spoken	6.0 (6.2)	16.7 (3.7)
Average language use[a]	3.7 (1.3)	6.1 (0.7)
Average language exposure[b]	29.0 (11.6)	71.0 (11.4)
Reading preference[b]	20.0 (14.2)	80.0 (14.2)
Speaking preference[b]	36.0 (15.6)	64.0 (16.2)
Self-rated proficiency[c]		
Speaking	7.4 (1.5)	9.3 (0.9)
Understanding	8.2 (1.4)	9.5 (0.6)
Reading	7.2 (1.6)	9.3 (0.9)
Objective proficiency measures		
Verbal Fluency[d]	32.4 (10.8)	42.9 (10.9)
Elicited Imitation Task[e]	5.5 (0.6)	5.7 (0.2)
Meaning	5.5 (0.6)	5.6 (0.3)
Syntax	5.5 (0.5)	5.7 (0.2)
Morphology	5.5 (0.6)	5.9 (0.1)
Vocabulary	5.6 (0.5)	5.7 (0.2)
DELE[f]	31.2 (6.5)	NA

[a] Average language use on a typical day. Ranges from 1 = 'No usage at all' to 6 'Usage all the time.'
[b] Averages expressed in percentages that sum up to 100%.
[c] Ranges from 0 = 'None' to 10 = 'Perfect.'
[d] Reflects raw counts of lexical items produced.
[e] Ranges from 0 = 'No response' to 6 = 'Perfect response'
[f] *Diploma de Español como Lengua Extranjera* 'Diploma of Spanish as a Foreign Language'

reported frequency use across four contexts (home, work, school, free time) was first calculated to obtain an average language use variable. This language use variable is on a 7-point Likert scale ranging from 'no usage at all' to 'usage all the time.' Then, the English use score was divided by the Spanish use score to obtain a composite language use ratio. A language use ratio of 0 indicates a participant uses both languages equally, while a score greater than 1 indicates English is used more, and a score lesser than 1 indicates Spanish is used more. On average, the sample obtained a language use ratio of 1.9 (*SD* = 0.9), which reflects the typical experience heritage speakers have of using the societal language more than the heritage one (Rothman & Treffers-Daller, 2014).

Verbal Fluency Tasks

Participants completed a category verbal fluency task in each language. They were given 30 seconds to name as many words as possible that applied to a specific semantic category (e.g., Zirnstein et al., 2018). Eight semantic categories were presented, for a total of four semantic categories for each language (i.e., 'musical instruments', 'animals', 'vegetables', and 'clothes' for one language; 'colors', 'body parts', 'fruits', and 'school supplies' for the other language). Semantic categories and language order were counterbalanced across participants. All the responses were recorded and transcribed for data analysis. The difference between participant's English and Spanish verbal fluency scores was taken as a proxy for language dominance. On average, participants produced more exemplars in English (M = 42.9, SD = 10.9) than Spanish (M = 32.4, SD = 10.8).

Elicited Imitation Tasks

An elicited imitation task (EIT) was administered in each language. In this task, participants hear a series of sentences that increase in syllable length and complexity. This process entails forming a mental representation of the sentence, keeping this representation in short-term memory, and afterwards producing it based on the initially formed mental representation (Bley-Vroman & Chaudron, 1994). The skills required to comprehend and produce sentences in the EIT reflect participants' implicit knowledge and strongly correlate with cloze tests and other proficiency measures (see Kostromitina & Plonsky, 2022 for a review). In this study, I administered the English and Spanish EIT created by Ortega et al. (2002). The English version included 30 sentences ranging from 7 to 19 syllables, while the Spanish version included 36 sentences ranging from 7 to 27 syllables. The Spanish version included complex constructions like clitic pronouns, reflexive *se* and the Spanish subjunctive. Because the goals of this study focus on examining agreement features in Spanish as a heritage language, I used the scoring developed by Gaillard and Tremblay (2016). This scoring evaluates productions on four[6] different 7-point subscales that assess participants' meaning, syntax, morphology, and vocabulary. A score of 0 indicates no response was given, while a score of 6 indicates the production is identical to the original sentence. If more than half of the sentence was correct, participants could receive a score ranging from 4 to 6 on each scale. In contrast, if more than half of the sentence was missing, participants' score could range from 1 to 3. On average, participants performed at ceiling on both EIT tasks, with the Spanish EIT scores exhibiting greater variation (M = 5.5, SD = 0.6) than the English ones (M = 5.7, SD = 0.2).

Acceptability Judgment Task (AJT)

This study used the same materials from Rossi et al. (2014), which contained 384 sentences: 192 experimental items and 192 fillers. The experimental items always began with a preamble that introduced the antecedent noun phrase followed by a clitic pronoun. The clitic pronoun was manipulated in a 2 x 2 gender and number design: a correct condition where gender and number match (+ Gender, + Number), a gender violation condition where only gender mismatched and number matched (– Gender, + Number), a number violation where only number mismatched and gender matched (+ Gender, – Number), and a double violation condition where both gender and number mismatched (– Gender, – Number). An example of the experimental items across conditions can be found in Table 2.

[6] Gaillard and Tremblay (2016) also include pronunciation as their 5th subscale. This dimension was not of interest to the current study on morphosyntactic variation, so it was not included.

Table 2: Example of the Experimental Sentences Used in the Acceptability Judgment Task

Number	Grammatical Gender	
	Match (+)	Mismatch (-)
Match (+)	Antes de leer el libro, Ana lo sacó de la envoltura de plástico.	Antes de leer el libro, Ana la sacó de la envoltura de plástico.
Mismatch (-)	Antes de leer el libro, Ana los sacó de la envoltura de plástico.	Antes de leer el libro, Ana las sacó de la envoltura de plástico.

Note. Translation = 'Before reading the book, Ana took it out of the plastic wrapper.'

In total, each condition had 48 sentences. Within these 48 sentences, the gender and number of the antecedent was balanced across gender and number: 12 masculine plural, 12 feminine plural, 12 masculine singular, and 12 feminine singular. Filler sentences served a dual purpose. First to counterbalance the overall ratio of correct (i.e., 48) and incorrect (i.e., 134) trials, and second, to expose participants to different structures than clitic pronouns in the experiment. Filler sentences included the following additional structures: subject-verb agreement (e.g., *El regalo para los bebés estaba/*estaban en el suelo*; the present for the babies was/*were on the floor), tense agreement (e.g., *Carolina le ha ordenado a María que limpie/*limpia su escritorio con un trapo*; Carolina ordered Maria to clean her desk with a cloth), and sentences with high and low attachment preferences (e.g., *Amelia fotografió a la prometida del príncipe que fue jugador de fútbol en Italia*; Amelia photographed the fiancée of the prince who was a soccer player in Italy). All sentences were counterbalanced across four lists using a Latin square design. Presentation of the materials was pseudorandomized so that consecutive trials would have 1) no more than two sentences of the same condition and 2) no more than two sentences with a masculine or feminine clitic. To diminish tiredness effects, four additional lists were created with the opposite order of items to the original ones. An individual participant viewed only one of the eight lists. Finally, to ensure participant's attention, twenty comprehension questions were interspersed at regular intervals unbeknownst to the participants. Comprehension questions only referred to the filler sentences.

Procedure

Eligible participants first provided their informed consent via email. Before coming into the lab for an EEG session, they first completed the DELE. Once they arrived at the lab for the EEG session, informed consent was given again in person. Next, participants were taken to an attenuated sound-proof chamber optimized for EEG recordings and an appropriate EEG cap was placed on their heads. Participants completed the language history questionnaire as the EEG preparation procedures were conducted. Once impedances were below 10 kilo-Ohms (kΩ), common EEG artifacts were explained to participants so they could avoid them as much as possible during the experiment (e.g., excessive blinking, tightly clenching the jaw). Afterwards, participants were given instructions on how to complete the experimental task. They were instructed to read each sentence carefully and perform an acceptability judgment at the end of each sentence as quickly and as accurately as possible. If the sentence was acceptable, they had to press a green response key and if the sentence was unacceptable they had to press a red response key. Participants were also told comprehension questions would appear periodically to ensure they were closely reading the sentences, and they were to respond to them using the same keys. Sentences were presented using E-Prime 3.0 (Psychological Software Tools, Inc., Pittsburgh, PA) through a computer monitor connected to the stimulus presentation computer inside the EEG control room. At the start of each trial, a fixation cross appeared for 2500 ms. Then, sentences were presented word by word using rapid-serial-visual-presentation (RSVP) at a fixed rate of 300 ms and an interstimulus interval (ISI) of 350 ms. Accuracy and reaction times to the acceptability judgments were collected for behavioral analysis.

EEG Data Acquisition and Pre-processing

Continuous EEG was acquired from 32 Ag/AgCl scalp active electrodes (BrainVision, Brain Products GmbH, Gilching, Germany) placed in accordance with the 10-20 system. The signal was amplified using a BrainVision LiveAmp wireless amplifier with 32 channels (Brain Products GmbH, Gilching, Germany) and continuously recorded with a 1000 Hz sampling rate without online filters. All pre-processing steps were conducted on Brain Vision Analyzer version 2.22 (Brain Products GmbH, Gilching, Germany). First, continuous EEG data were zero-phased filtered using 4th order Butterworth IIR filters, with a high-pass cutoff of 0.1 Hz and a low-pass cutoff of 30 Hz. Then, the data were manually inspected for artifact rejection, and an independent component analysis (ICA) was used to identify and remove vertical and horizontal eye movements. Furthermore, three participants elicited a lot of alpha wave activity across their EEG signal, so their ICA identified an alpha component that was also removed. Re-referencing was conducted on the average of the left and right mastoids. Finally, the data were segmented into epochs relative to the clitic pronoun from −200 ms to 950 ms, baseline corrected to −200 ms to 0 ms, and averaged per condition per participant for all trials.

In contrast to Rossi et al. (2014) who only analyzed RTs to correct responses, the current study analyzes RTs regardless of accuracy with the goal of understanding differences between explicit metalinguistic judgments and implicit online language processing measures. To control for outliers, reaction times (RT) below 300 ms and above 5,000 ms were excluded in accordance to Rossi et al. (2014). This resulted in 21% data loss and evenly affected all four conditions. Additionally, RTs that were 2.5 SDs above the mean of their respective condition were windsorized[7], a process that affected 15% of the remaining data points and was also evenly

[7] A process by which RTs > 2.5 SDs above the mean of each condition were replaced with their respective cutoff values.

distributed across conditions. To determine whether it was necessary to transform the RT data to meet normality and homoscedasticity assumptions, the *boxcox* function from the *MASS* R-package was used (Ripley et al., 2013). The optimal transformation could be achieved with the log scale, so a log10 transformation was applied to RTs in accordance with suggestions from Winter (2019).

Statistical Modeling

All the statistical analysis were conducted in R (Version 4.6.0; R Core Team, 2026) using the *lme4* package for mixed-effects modeling (Bates et al., 2015). The gender and number variables were sum coded as 2-level factors where matches were assigned a -0.5 value and mismatches were assigned a +0.5 value. To answer research question 1, a model with maximal random effects[8] structures (Barr et al., 2013) was built for each dependent variable using maximum likelihood (models a and b). For the accuracy model, a generalized linear mixed-effects logistic regression model was used with binomial as the type of probability distribution used.

a) DV ~ Gender * Number + (1 + Gender * Number | Subject) + (1 | Item)
b) DV ~ Gender * Number * Electrode + (1 + Gender * Number | Subject)

For the electrophysiological data, an electrode variable was created (model b) by clustering electrodes into three main sites of interest: left-anterior ('Fp1', 'F7', 'F3', 'FC5', 'FC1'), central ('C3', 'Cz', 'C4', 'Pz', 'CP2', 'CP1', 'P3', 'P4'), and posterior ('O1', 'O2', 'P7', 'P8', 'P3', 'P4', 'CP5', 'CP6', 'CP1', 'CP2'). The first time window of interest aimed to capture potential LAN and N400 effects at 350 ms – 550 ms. Therefore, the electrode variable for this time window captured raw voltage differences between central and left-anterior sites, which would indicate a LAN or an

[8] Because of the ERP grand averaging process, item-level information is lost

N400, or central vs posterior sites, which would indicate a more central or posterior N400. A second time window was extracted for the P600 at 550 ms – 750 ms. Because the P600 tends to be distributed in central posterior regions, only the central vs posterior contrast was included in the electrode variable of the second time window.

To examine how different measures of Spanish proficiency affect Spanish heritage speakers' processing of gender and number, the following individual differences (ID) were added in separate models (models c and d): DELE score, Spanish EIT score, language use ratio, language dominance, and schooling difference. All continuous predictors were standardized except for the schooling difference variable, which represents the raw difference between years of schooling in English and years of schooling in Spanish. To understand the relationship between the individual differences of interest, Pearson correlations were run.

c) DV ~ Gender * Number * ID + (1 + Gender * Number | Subject) + (1 | Item)

d) DV ~ Gender * Number * Electrode * Proficiency +
   (1 + Gender * Number | Subject) + (1 | Item)

If the model failed to converge, control = lmerControl(calc.derivs = FALSE) and optimizer = "bobyqa" was added to the model. If it still did not converge, the random effect that contributed the least variance was removed and if the model still did not converge, one random slope at a time was removed until convergence was achieved (Matuschek et al., 2017). This last step was only applied to the reaction time language dominance model, which converged with gender and number random slopes, and subject and item random intercepts. The *lmerTest* package (Kuznetsova et al., 2017) was used to compute all *p* values and its ANOVA function was used to employ a likelihood ratio test to determine the best fitting model. For non-nested models, or cases in which the test was not informative, the best fitting model was selected based on: lower

Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), higher log likelihood, and smaller deviance from a perfect model.

Results

In this section, I first present the results from the correlations between individual differences measures with the goal of understanding how these relationships might influence the interpretation of the regression results. Then, in each subsection below, I first report the results from the best fitting models. In cases where the best fitting model did not include an interaction between the conditions of interest (i.e., gender, number) and individual differences measures, I present emerging trends from alternative models. All the models built along with their model fit indices can be found in Appendix A.

Correlations Between Individual Differences Measures

Pearson correlations between individual differences measures are shown in Figure 4. DELE was significantly positively correlated with Spanish EIT, $r(33) = .64$, $p < .001$ and language dominance, $r(33) = .35$, $p < .05$, and negatively correlated with years of English schooling, $r(33) = -.43$, $p < .01$. The positive correlations between the DELE, Spanish EIT and language dominance are expected, since higher scores in each of them reflect higher Spanish proficiency. In contrast, the negative correlation between DELE and English schooling suggests that higher years of English schooling are associated with lower Spanish proficiency scores. This interpretation is further supported by the negative correlation between Spanish EIT and English schooling, $r(33) = -.36$, $p < .05$, and suggests that this relationship holds regardless of language test modality. Spanish EIT scores were also significantly positively correlated with English EIT scores, $r(33) = .58$, $p < .001$, and language dominance, $r(33) = .39$, $p < .05$, thus showcasing the highly proficient nature heritage bilinguals exhibit in both of their languages,

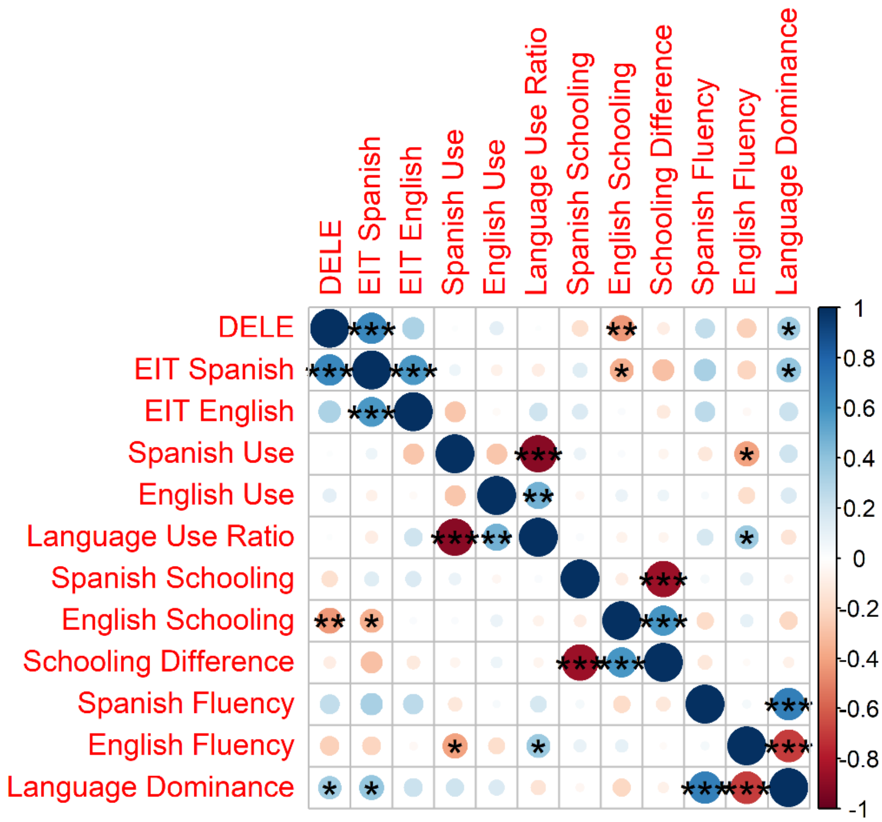

Figure 4: Correlations Between Study 1 Individual Differences

Note. The y-axis on the right side contains the correlation strength legend. A correlation of 0 is presented in white, while positive correlations are presented in blue and negative correlations are presented in red. The darker the shade, the stronger the correlation magnitude.

while the relationship between Spanish EIT and language dominance suggests the ability to produce more lexical items in the heritage language relative to English is associated with higher Spanish proficiency. Regarding English fluency, there was a statistically significant positive correlation with language use ratio, $r(33) = .36$, $p < .05$, and a negatively significant correlation with average Spanish use, $r(33) = -.40$, $p < .05$. These correlations indicate that higher language use ratio (using more English) is associated with greater ability to produce lexical items in English, while increased average Spanish use is associated with lower lexical production ability in English. All composite measures—language use ratio, schooling difference, language dominance—correlated as expected with their constructed variables. What is important for the current study is that all three composite measures were not correlated with each other. According to a meta-analysis on individual differences research (Gignac & Szorodai, 2016), correlations greater than .30 are considered large effects, therefore the magnitude of the reported correlations in this study appears to be strong.

Behavioral Results

For accuracy, the best fitting model (see Table 3 and Figures 5 and 6) included main effects for gender, $b = -1.008$, $SE = .128$, $p < .001$, number, $b = -1.238$, $SE = .136$, $p < .001$, and DELE, $b = .589$, $SE = .172$, $p < .001$, as well as statistically significant interactions between gender and number, $b = 3.144$, $SE = .250$, $p < .001$, gender and DELE, $b = .429$, $SE = .130$, $p < .001$, and number and DELE, $b = .285$, $SE = .139$, $p < .05$. To further explore the gender and number interaction, simple effect analyses were conducted. When gender matched, heritage speakers were significantly more accurate on number matched conditions than number mismatch conditions ($p < .001$). This pattern was reversed when gender mismatched: accuracy was significantly higher to number mismatch conditions than number match conditions ($p < .01$). Similarly, when number matched, accuracy to gender matches was higher than gender mismatches ($p < .001$); when number mismatched, accuracy to gender mismatches was higher than in gender matches. Taken together, the simple effects show heritage speakers were less accurate when only one feature mismatched and more accurate when both features matched. In the mismatch context, when both features mismatched, accuracy also improved. Finally, while higher DELE scores significantly increased accuracy across all conditions, they specifically increased accuracy to gender mismatch conditions and number mismatch conditions, an effect whose magnitude was strongest for gender.

Table 3: Logistic Regression for Accuracy in the AJT

	Estimate	*SE*
Fixed Effects		
Intercept	.263	.169
Gender	-1.008***	.128
Number	-1.238***	.136
Gender*Number	3.144***	.250
DELE	.589***	.172
Gender*DELE	.429***	.130
Number*DELE	.285*	.139
Gender*Number*DELE	-.237	.350
Random Effects	Variance	*SD*
Subject	.942	.971
Item	.018	.134
Gender	.372	.610
Number	.454	.674
Gender*Number	1.382	1.176

Note. Estimates are in log odds.
* p < .05. ** p < .01. *** p < .001.

For reaction time, the best fitting model (see Table 4 and Figures 7 and 8) included a main effect for number, $b = .028$, $SE = .009$, $p < .01$, an interaction between gender and number, $b = -.050$, $SE = .015$, $p < .001$, and an interaction between gender and language dominance, $b = -.024$, $SE = .010$, $p < .05$. To further explore the gender and number interaction, simple effects analyses were performed. When gender matched, heritage speakers responded significantly slower to number mismatches than number matches ($p < .001$). In contrast, when gender mismatched no differences were found between number matches and number mismatches.

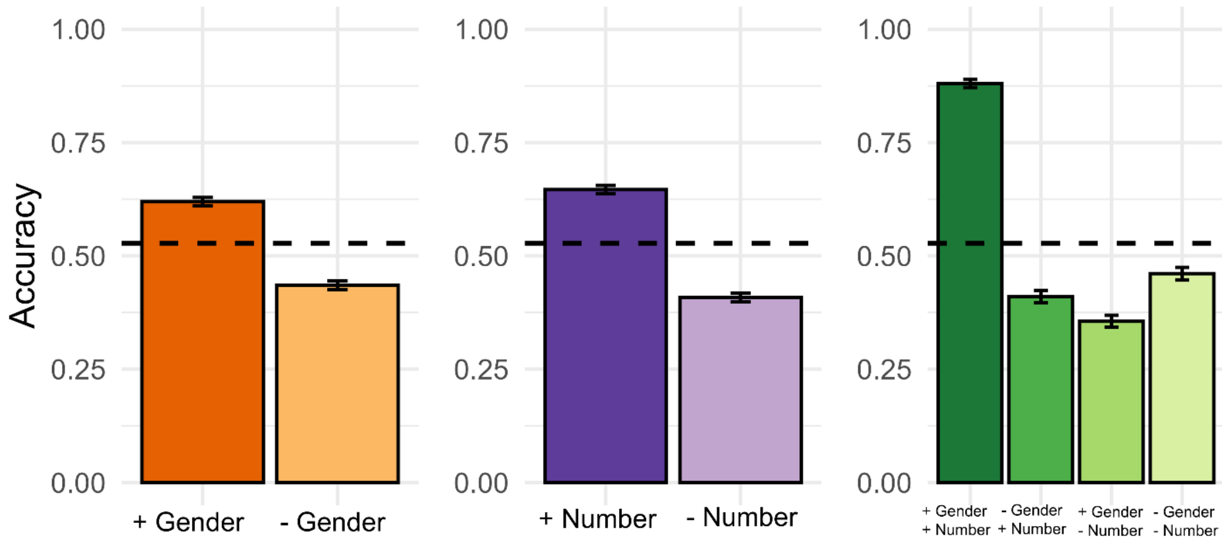

Figure 5: AJT Accuracy by Gender and Number

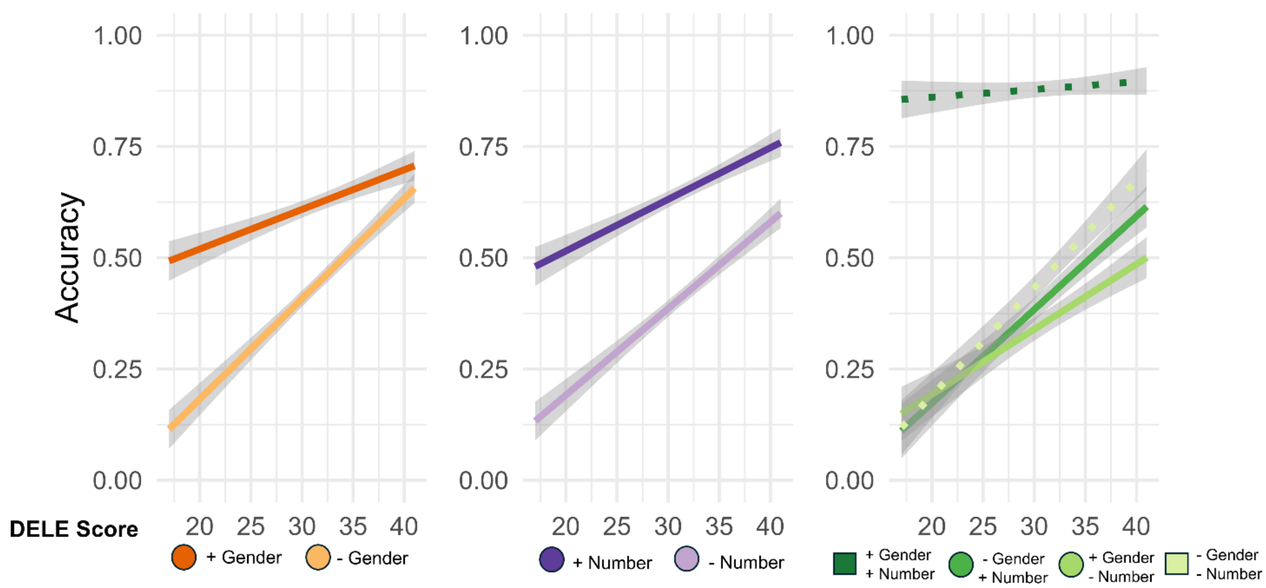

Figure 6: AJT Accuracy by DELE*Gender and DELE*Number

When examining number matches, heritage speakers were significantly slower to respond to gender mismatches than gender matches ($p < .05$). When number mismatched, heritage speakers responded significantly faster to gender mismatches than gender matches ($p < .05$). Taken together, these findings suggest that the processing cost associated with a mismatch in one feature was reduced or eliminated when the other feature mismatched as well. Finally, the significant gender and language dominance interaction suggests that as heritage speakers become more Spanish dominant, responses to gender mismatches are faster.

Table 4: Linear Regression for Reaction Time in the AJT

	Estimate	*SE*
Fixed Effects		
Intercept	3.051***	.029
Gender	.000	.010
Number	.028**	.009
Gender*Number	-.050***	.015
Language Dominance	-.035	.029
Gender*Language Dominance	-.024*	.010
Number*Language Dominance	-.011	.010
Gender*Number*Language Dominance	-.004	.015
Random Effects	Variance	*SD*
Subject	.029	.171
Item	.002	.041
Gender	.002	.040
Number	.001	.035

* p < .05. ** p < .01. *** p < .001.

ERP Results

In the LAN/N400 time window, the best fitting model (see Table 5) included significant main effects for number, $b = .688$, $SE = .293$, $p < .05$, and language use ratio, $b = 1.327$, $SE = .494$, $p < .05$, as well as a significant interaction between gender and number. In other words, heritage speakers with an average language use ratio, elicited stronger positivities to number mismatches and stronger negativities to number matches (see Figure 9). There was also a

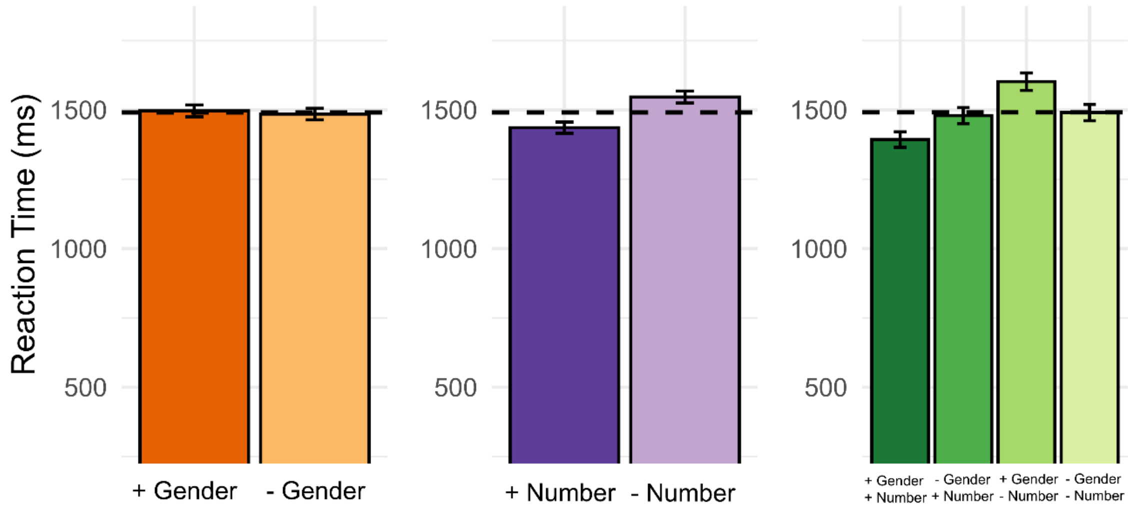

Figure 7: AJT Reaction Time by Gender and Number

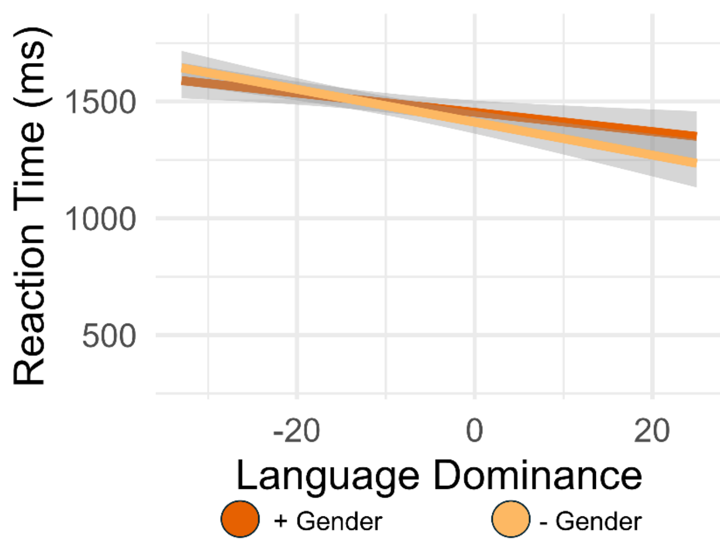

Figure 8: AJT Reaction Time by Gender*Language Dominance

statistically significant interaction between gender and number, $b = 2.045$, $SE = .861$, $p < .05$. Follow-up simple-effects analysis revealed number matches were significantly more negative than number mismatches when gender mismatched ($p < .01$). When number mismatched, gender mismatches were significantly more positive ($p < .05$) than gender matches.

Table 5: LAN/N400 Time Window Best Fitting Model

	Estimate	*SE*
Fixed Effects		
Intercept	-.194	.493
Gender	.029	.307
Number	.688*	.293
C vs LA	.349***	.074
C vs P	-1.068***	.074
LUR	1.327*	.494
Gender*Number	2.045*	.861
Gender*C vs LA	.001	.147
Gender*C vs P	-.080	.147
Number*C vs LA	.044	.147
Number*C vs P	.056	.147
Gender*LUR	.263	.308
Number*LUR	-.039	.293
LUR*C vs LA	.295***	.074
LUR*C vs P	-.087	.074
Gender*Number*C vs LA	.142	.295
Gender*Number C vs P	.054	.295
Gender*Number*LUR	-.563	.863
Gender*LUR*C vs LA	-.076	.147
Gender*LUR*C vs P	.109	.147
Number*LUR*C vs LA	-.167	.147
Number*LUR*C vs P	.059	.147
Gender*Number*LUR*C vs LA	-.147	.295
Gender*Number*LUR*C vs P	.002	.295
Random Effects	Variance	*SD*
Subject	8.419	2.902
Gender	2.930	1.712
Number	2.621	1.619
Gender*Number	24.458	4.945

* $p < .05$. ** $p < .01$. *** $p < .001$.

In other words, when gender mismatched but number matched greater negativities were elicited, while gender and number mismatches together elicited the most positive amplitudes (see Figure 9). In general, using more English than Spanish was associated with stronger positivities overall, an effect that was strongest in central sites, $b = .295$, $SE = .074$, $p < .001$. Moreover, while electrode site did not produce statistically significant interactions with gender or number, central sites were more positive than left anterior sites, $b = .349$, $SE = .074$, $p < .001$, but less positive than posterior sites, $b = -1.068$, $SE = .074$, $p < .001$.

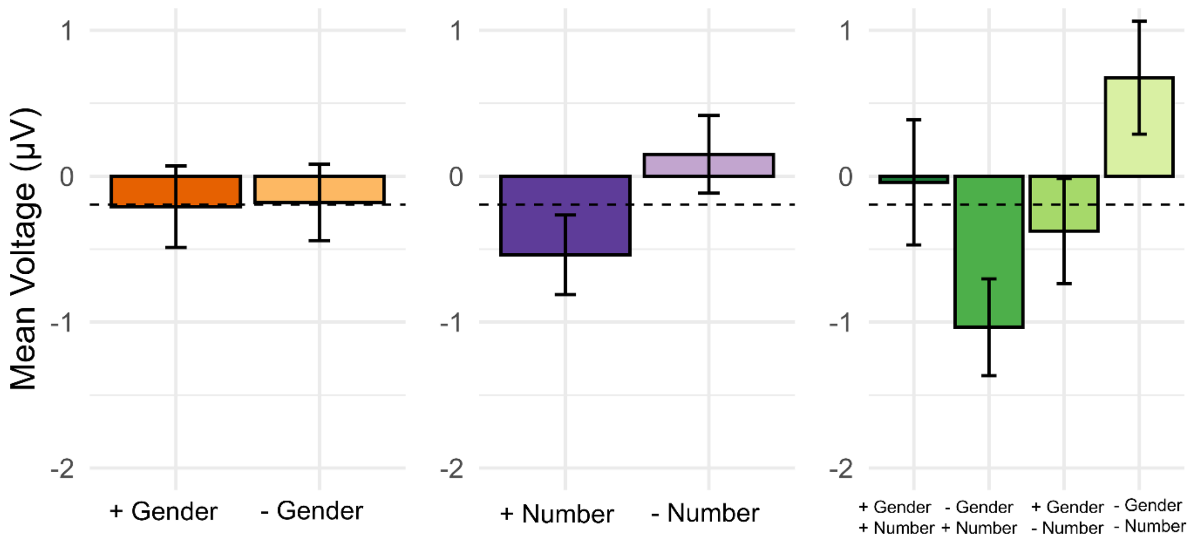

Figure 9: Gender and Number Effects at the N400 Time Window

An alternate model with language dominance (see Table 6), found a marginally significant interaction between number, electrode cluster and language dominance, $b = .275$, $SE = .150$, $p = .068$. In other words, as heritage speakers became more Spanish dominant, the difference between number mismatches and number matches became more negative in left anterior sites than in central sites (see Figure 10). Follow-up estimated marginal means comparisons at the LA site across different language dominance levels showed significant differences between number matches and mismatches for participants who were more English dominant (-2 SD $p < .05$; -1 SD $p < .05$). For these heritage speakers, there was greater positivity to number mismatches. Although the simple effects did not reveal significant number agreement effects at higher levels of Spanish dominance (+2 SD), visual inspection suggests greater Spanish dominance might be related to a LAN effect to number mismatches.

Table 6: LAN/N400 Time Window Language Dominance Model

	Estimate	*SE*
Fixed Effects		
Intercept	-.194	.524
Gender	.029	.311
Number	.689*	.291
C vs LA	.349***	.075
C vs P	-1.068***	.075
Language Dominance	-.808	.523
Gender*Number	2.045*	.867
Gender*C vs LA	.001	.150
Gender*C vs P	-.080	.150
Number*C vs LA	.044	.150
Number*C vs P	.056	.150
Gender*Language Dominance	-.065	.311
Number*Language Dominance	-.194	.291
Language Dominace*C vs LA	-.121	.075
Language Dominance*C vs P	-.035	.075
Gender*Number*C vs LA	.142	.300
Gender*Number C vs P	.054	.300
Gender*Number*Language Dominance	-.104	.868
Gender*Language Dominance*C vs LA	-.093	.150
Gender*Language Dominance*C vs P	.009	.150
Number*Language Dominance*C vs LA	.275•	.150
Number*Language Dominance*C vs P	-.208	.150
Gender*Number*Language Dominance*C vs LA	.076	.300
Gender*Number*Language Dominance*C vs P	-.084	.300
Random Effects	Variance	*SD*
Subject	9.522	3.086
Gender	2.982	1.727
Number	2.571	1.604
Gender*Number	24.712	4.971

* $p < .05$. ** $p < .01$. *** $p < .001$. • = marginally significant effect

At the LAN/N400 time window, another alternate model found a marginally significant interaction between gender, number, and schooling difference (see Table 7 and Figure 11), b = -.200, SE = .113, p = .084. Follow-up simple effects for the gender and number interaction at equal English and Spanish schooling show gender mismatches were significantly more negative than gender matches when number matched ($p < .05$), but significantly more positive when number mismatched ($p < .05$). Similarly, when gender mismatched number mismatches were

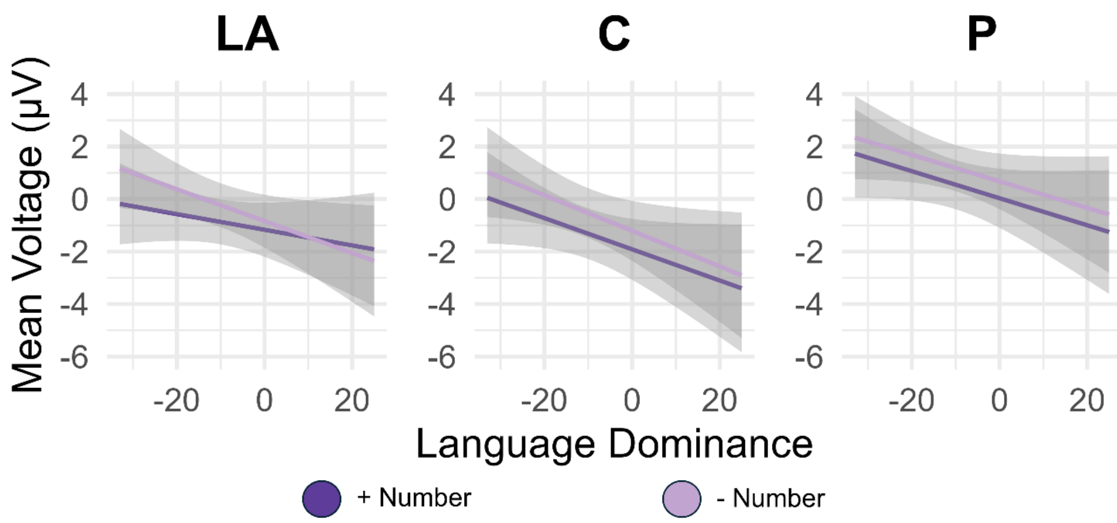

Figure 10: Number LAN by Language Dominance

significantly more positive (*p* < .01). However, the magnitude of these effects diminishes as heritage speakers have more years of English schooling relative to Spanish.

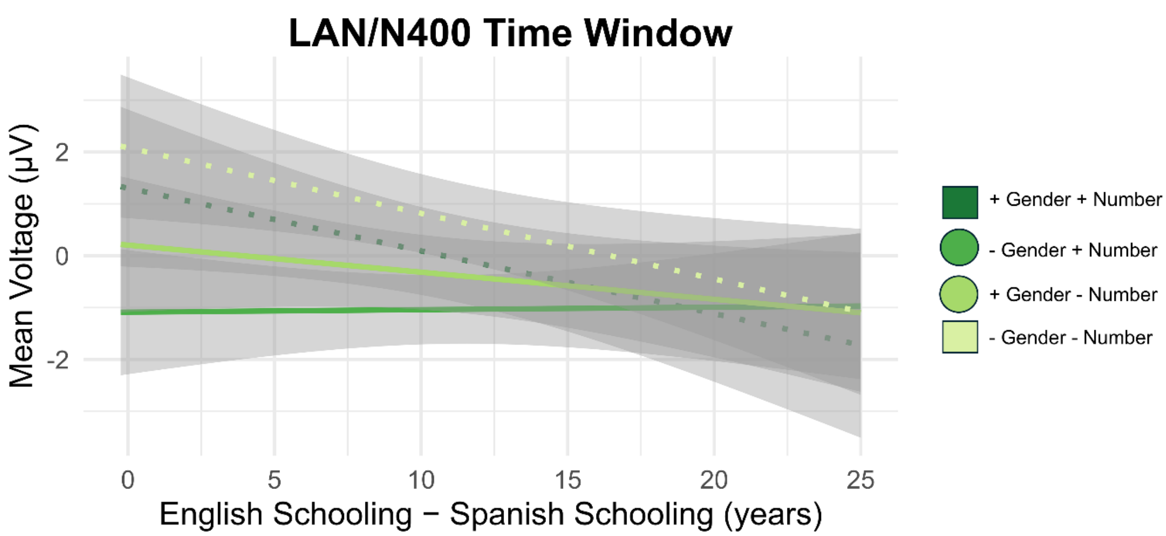

Figure 11: LAN/N400 Gender*Number Interaction by Schooling Difference

Table 7: LAN/N400 Time Window Schooling Difference Model

	Estimate	*SE*
Fixed Effects		
Intercept	.625	.966
Gender	-.257	.559
Number	1.036•	.525
C vs LA	.384**	.137
C vs P	-1.172***	.137
Schooling Difference	-.074	.072
Gender*Number	4.269**	1.501
Gender*C vs LA	-.304	.274
Gender*C vs P	.127	.274
Number*C vs LA	.175	.274
Number*C vs P	.001	.274
Gender*Schooling Difference	.026	.042
Number*Schooling Difference	-.031	.039
Schooling Difference*C vs LA	-.003	.010
Schooling Difference*C vs P	.009	.010
Gender*Number*C vs LA	-.003	.547
Gender*Number C vs P	.092	.547
Gender*Number*Schooling Difference	-.200•	.113
Gender*Schooling Difference*C vs LA	.027	.021
Gender*Schooling Difference*C vs P	-.019	.021
Number*Schooling Difference*C vs LA	-.012	.021
Number*Schooling Difference*C vs P	.005	.021
Gender*Number*Schooling Difference*C vs LA	.013	.041
Gender*Number*Schooling Difference*C vs P	-.003	.041
Random Effects	Variance	*SD*
Subject	9.877	3.143
Gender	2.942	1.715
Number	2.548	1.596
Gender*Number	22.516	4.745

* p < .05. ** p < .01. *** p < .001. • = marginally significant effect

In the P600 time window, the best fitting model (see Table 8 and Figure 12) only showed a statistically significant difference between posterior and central sites, with the former being more positive than the latter, $b = .711$, $SE = .092$, $p < .001$. Nevertheless, an alternate model (see Table 9) that included DELE scores, found a marginally significant interaction between gender and DELE, $b = .563$, SE = .312, $p = .080$. In other words, as DELE scores increase, the difference between the gender mismatch and the gender match conditions becomes more positive. Figure 13 shows this effect is driven by the gender match conditions which become more negative as DELE scores increase. Two alternate models show triple interactions between gender, number, and language use ratio (see Table 10) as well as gender, number, and schooling difference (see Table 11). The first of these triple interactions is illustrated in Figure 14. This shows that as heritage speakers use more English than Spanish on average, the voltage response

Table 8: P600 Time Window Best Fitting Model

	Estimate	*SE*
Fixed Effects		
Intercept	-1.330**	.466
Gender	.332	.325
Number	.291	.282
P vs C	.711***	.092
Gender*Number	.819	.681
Gender*P vs C	-.010	.184
Number*P vs C	-.098	.184
Gender*Number*P vs C	-.388	.368
Random Effects	Variance	*SD*
Subject	7.526	2.743
Gender	3.411	1.847
Number	2.478	1.574
Gender*Number	15.055	3.880

* p < .05. ** p < .01. *** p < .001.

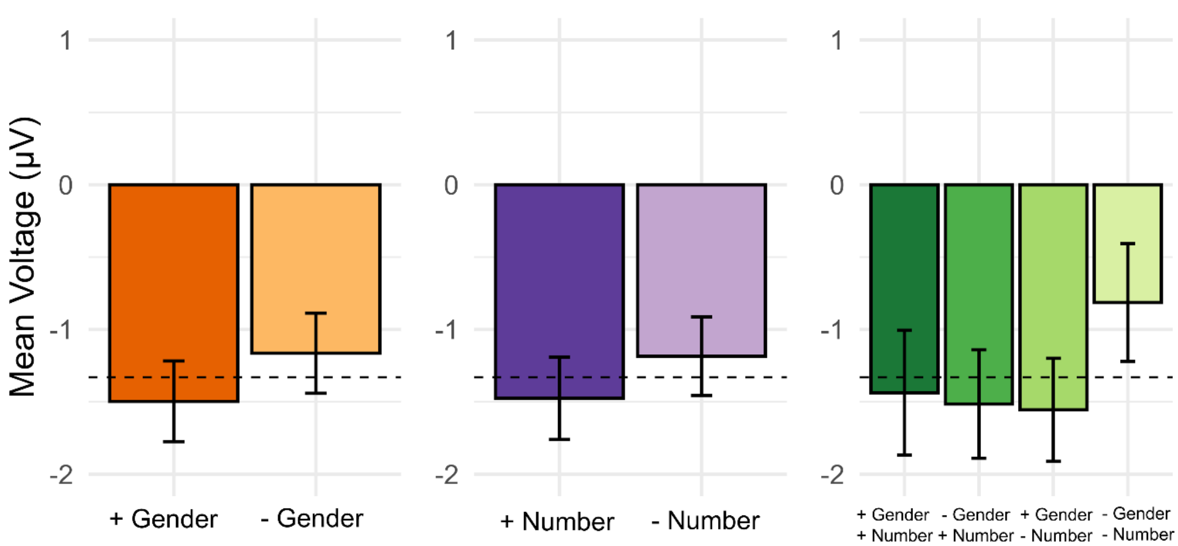

Figure 12: Gender and Number Effects at the P600 Time Window

Table 9: P600 Time Window DELE Model

	Estimate	*SE*
Fixed Effects		
Intercept	-1.330	.464
Gender	.332	.311
Number	.291	.280
P vs C	.711***	.092
DELE	-.280	.464
Gender*Number	.819	.680
Gender*P vs C	-.010	.183
Number*P vs C	-.098	.183
Gender*DELE	.563•	.312
Number*DELE	.150	.281
DELE*P vs C	.008	.092
Gender*Number*P vs C	-.388	.367
Gender*Number*DELE	.145	.682
Gender*DELE*P vs C	.077	.184
Number*DELE*P vs C	.134	.184
Gender*Number*DELE*P vs C	.199	.368
Random Effects	Variance	*SD*
Subject	7.449	2.729
Gender	3.098	1.760
Number	2.458	1.568
Gender*Number	15.045	3.879

* p < .05. ** p < .01. *** p < .001. • = marginally significant effect

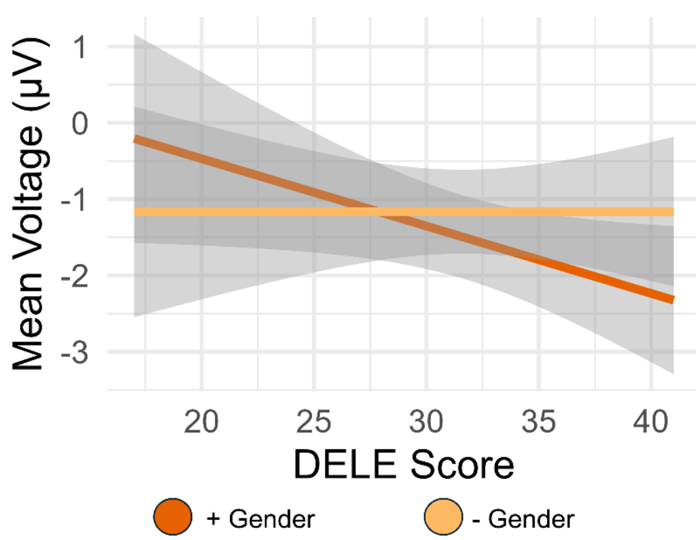

Figure 13: P600 Gender*DELE Interaction

Table 10: P600 Time Window Language Use Ratio Model

	Estimate	*SE*
Fixed Effects		
Intercept	-1.330	.466
Gender	.332	.325
Number	.291	.281
P vs C	.711***	.091
LUR	.103	.466
Gender*Number	.819	.651
Gender*P vs C	-.010	.182
Number*P vs C	-.098	.182
Gender*LUR	.104	.326
Number*LUR	.043	.282
LUR*P vs C	-.125	.091
Gender*Number*P vs C	-.388	.364
Gender*Number*LUR	-1.181•	.653
Gender*LUR*P vs C	-.189	.182
Number*LUR*P vs C	-.040	.182
Gender*Number*LUR*P vs C	.190	.365
Random Effects	Variance	*SD*
Subject	7.515	2.741
Gender	3.406	1.846
Number	2.483	1.576
Gender*Number	13.691	3.700

* p < .05. ** p < .01. *** p < .001. • = marginally significant effect

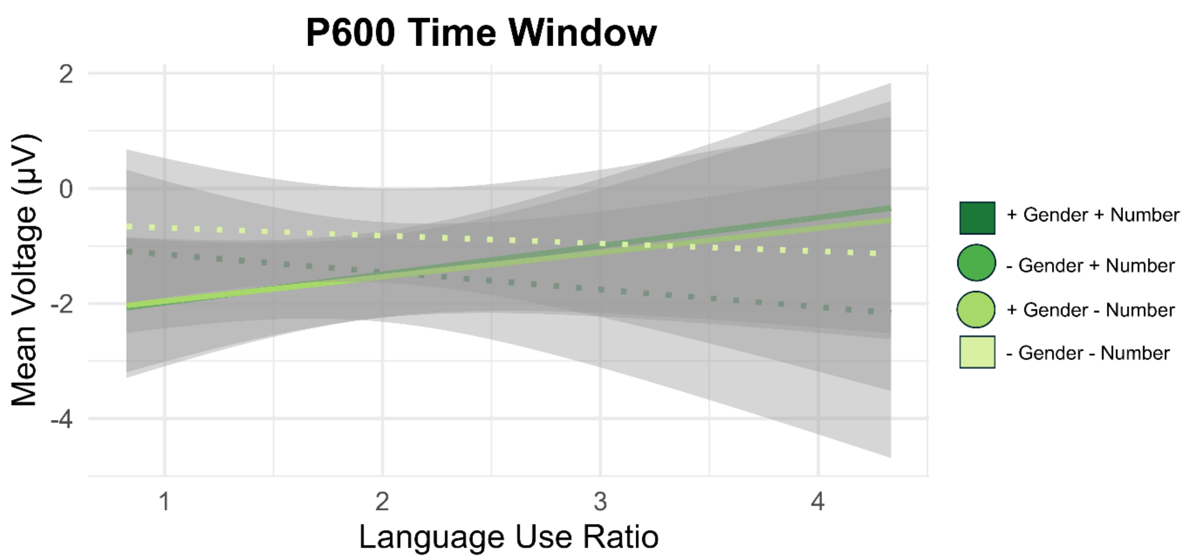

Figure 14: P600 Gender*Number Interaction by Language Use Ratio

Table 11: P600 Time Window Schooling Difference Model

	Estimate	*SE*
Fixed Effects		
Intercept	-.615	.830
Gender	.264	.589
Number	.291	.509
P vs C	.743***	.166
Schooling Difference	-.064	.062
Gender*Number	3.013*	1.149
Gender*P vs C	-.094	.333
Number*P vs C	-.204	.333
Gender*Schooling Difference	.006	.044
Number*Schooling Difference	.000	.038
Schooling Difference*P vs C	-.003	.012
Gender*Number*P vs C	-.491	.665
Gender*Number*Schooling Difference	-.197*	.086
Gender*Schooling Difference*P vs C	.008	.025
Number*Schooling Difference*P vs C	.010	.025
Gender*Number*Schooling Difference*P vs C	.009	.050
Random Effects	Variance	*SD*
Subject	7.302	2.702
Gender	3.409	1.846
Number	2.479	1.575
Gender*Number	12.946	3.598

* p < .05. ** p < .01. *** p < .001.

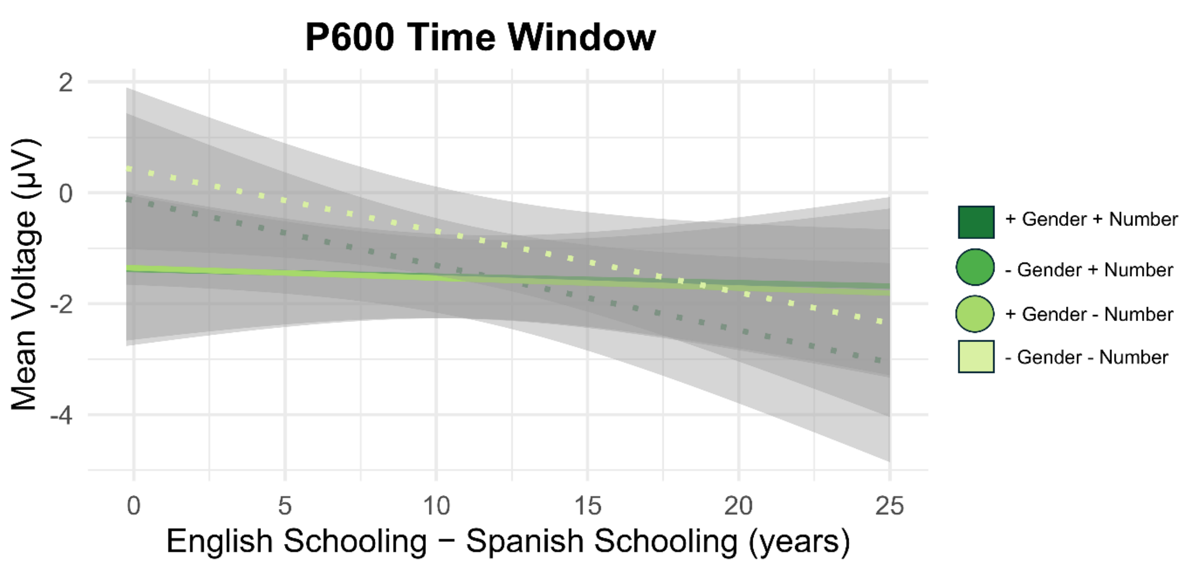

Figure 15: P600 Gender*Number Interaction by Schooling Difference

to different gender and number combinations also changes, $b = -1.181$, $SE = .653$, $p = .079$. While the figure visually shows sensitivity to pure gender mismatches and pure number mismatches increases as heritage speakers use more English than Spanish, follow-up simple analysis at average language use ratio showed the effect of gender and number did not depend on whether the other feature matched or mismatched. The second triple interaction is visually presented in Figure 15 and it shows the interaction between gender and number varies as a function of schooling difference, $b = -.197$, $SE = 0.086$, $p < .05$. Follow-up simple effects for the gender and number interaction at equal years of English and Spanish schooling, $b = 3.013$, $SE = 1.149$, $p < .05$, shows gender mismatches were significantly more positive than gender matches when number mismatched ($p < .01$), but no gender agreement effect was found when number matched. The same is true for the number agreement effect. That is, number mismatches were significantly more positive than number matches when gender mismatched ($p < .05$), but no number agreement effect was found when gender matched. Moreover, as heritage speakers had more schooling in English relative to Spanish, the magnitude if this effect decrease. Taken together, the results suggest the double violation conditions elicit a P600 when English and Spanish schooling years are equal, a sensitivity that decreases as the difference in schooling increases (i.e., more years of English relative to Spanish).

Discussion

The goals of the first study of the dissertation were to examine how Spanish heritage speakers process gender and number agreement in clitic pronouns, a syntactically complex structure, and to examine how individual differences in Spanish proficiency, language use, schooling, and language dominance affect the processing of gender and number agreement in clitic pronouns. Below, I first explain the findings from number agreement, the shared feature across languages, then the findings from gender agreement processing, a feature present only in the heritage language, and finally the findings from processing gender and number together.

Heritage Speaker's Number Processing

Behaviorally, heritage speakers were less accurate and slower to number mismatches, but as Spanish proficiency—indexed by the DELE—increased, heritage speakers went from performing below chance to performing above and beyond chance. This finding highlights the role that Spanish proficiency plays in number agreement sensitivity via metalinguistic judgments. The findings from the reaction time data showed heritage speakers were significantly slower when responding to number mismatches. Given that number is a shared agreement feature between English and Spanish, this finding is unexpected. One possible explanation is that, at the behavioral level, there is a clash between the shared number feature across languages which creates competition, thus resulting in lower accuracy and longer processing times when metalinguistic judgments are required.

Electrophysiologically, heritage speakers showed greater negativities to number matches at the LAN/N400 time window. This finding is consistent with the number main effect Rossi et al. (2014) found in their prototypical native speaker sample and suggests heritage speakers do not differ from baseline speakers in this time window. In the P600 time window, although number mismatches trended towards a P600 effect, it was neither marginally nor statistically significant. Evidence for sensitivity to number mismatches also appeared as a trend in left-anterior sites for heritage speakers' whose Spanish dominance was 2 SD above the mean. Although this trend comes from the data of two individuals, I believe it can inform our understanding of how regulation of the societal language English might be related to processing ease in the heritage language. Both individuals who produced 20 or more lexical items in

Spanish than in English: acquired English after the age of 5, were exposed to English at least 70% of the time, used more English than Spanish in their daily lives, preferred to read in English at least 80% of the time, and preferred to speak in English at least 50% of the time. In the literature, performance on verbal fluency is not only associated with lexical access, but in bilinguals it is also taken as a measure of how well they can regulate the use of their two languages. For example, Zirnstein et al. (2018) found that Mandarin-English bilinguals living in the U.S. who were better able to regulate their dominant L1 Mandarin, showed greater prediction costs when reading in their L2, English. Although this finding pertains to prediction costs in L2 learners and the current study deals with issues of agreement processing in heritage speakers, language regulation is an essential part of the bilingual experience. In this study, it seems that heritage speakers who can better regulate the interference from the societal language, as indexed by producing more Spanish than English exemplars while living and being immersed in an English-speaking context, are better able to resolve the competition that arises from having double number features across both languages. Thus, resulting in automatic detection of the number mismatch. Future studies should further explore how regulation of the societal language English might be related to morphosyntactic processing in Spanish heritage speakers.

Heritage Speaker's Gender Processing

Behaviorally, heritage speakers were less accurate to gender mismatches when compared to gender matches, but Spanish proficiency—indexed by the DELE—moderated accuracy more strongly than it did for number: as Spanish proficiency goes up, heritage speakers go from performing below chance to performing above and beyond chance. While heritage speakers responded equally fast to gender matches and mismatches, those who were more Spanish dominant in their verbal fluency responded to gender mismatches faster. This finding also fits well with the language regulation account. In other words, better regulation of the societal language English facilitates the exploitation of gender cues, the feature unique to Spanish, when performing metalinguistic judgments.

Electrophysiologically, gender mismatches did not elicit any LAN or N400 effects, only a P600 trend was observed. A marginally significant interaction did appear between DELE and gender for the P600 time window. As Spanish proficiency increases, the gender match conditions become more negative. The fact that this happens while the voltage to gender mismatches remains relatively stable, suggests processing gender mismatches in clitic pronouns depends on successful integration of gender. Such an interpretation is in line with ERP work that finds the P600 reflects difficulty with syntactic integration processes in general (Kaan et al., 2000). In this line of work, a P600 is found for grammatical sentences that are difficult to integrate (e.g., 'who'-q vs 'whether'-q). Following the logic from Kaan and colleagues, the activation of gender at the noun phrase level triggers an expected agreeing element at the clitic level. At lower levels of Spanish proficiency, integrating the gender at the clitic level requires a lot of energy and results in a P600 to gender match conditions, while at higher levels of Spanish proficiency the gender integration at the clitic level requires lower levels of activation, thus resulting in a P600 to gender mismatches. This interpretation patterns well with what was observed at the behavioral level with accuracy. At lower levels of proficiency, heritage speakers perform at chance in gender match conditions as a result of their difficulty integrating gender at the clitic level, but at higher levels of proficiency heritage speakers perform above and beyond chance in gender match trials as a result of successful gender integration. While accuracy across gender conditions remains below 75% despite high levels of Spanish proficiency, I believe this reflects how metalinguistic judgments are not always indicative of actual language competence. Instead, metalinguistic judgments might be influenced by perceived competence in the heritage language, something heritage speakers tend to underestimate quite a lot (e.g., Vargas Fuentes & Torres, 2026).

Heritage Speakers' Gender and Number Processing

Behaviorally, gender and number seem to be processed together. In accuracy, the gender and number interaction shows accuracy is highest when both features match. When both features mismatch, although still below chance, accuracy is the highest when compared to the conditions where only one feature mismatched. Furthermore, visual inspection of how Spanish proficiency affects the processing of gender and number together (see Figure 6) suggests Spanish proficiency does not influence judgment accuracy when both features match, instead all the mismatch conditions seem to improve with higher Spanish accuracy. In reaction time, the gender and number interaction shows heritage speakers were fastest when both features matched. When both features mismatched, the number main effect was reduced and heritage speakers responded faster than when only number mismatched. Taken together, the behavioral results support the representational economy claim put forward by Scontras et al. (2018).

Electrophysiologically, this study also finds evidence that heritage speakers process gender and number together. At the 350ms – 550ms time window, the effect of gender agreement depends on number agreement, such that when both features mismatch the strongest positivity is observed. The positive voltage to the double violation might in fact reflect the beginning of a P600 effect that begins earlier and continues well into 750ms. The fact that the double violation remains the most positive in the 550ms – 750ms seems to support such a notion. Potential explanations as to what factors contribute to the processing of gender and number together come from the alternate models. In both time windows the processing of both features together was associated with the difference in schooling years, such that heritage speakers with equal years of schooling in both languages showed greater positivity to the double violation. Although informative, this trend should be taken with caution since heritage speakers in this study were asked separately for each language to 'list the number of years and/or months spent in a school and/or working environment where _____ is spoken.' Thus, from the data collected it is not possible to differentiate whether the years reflect educational experience in Spanish or if Spanish was spoken only among peers in a school or work environment. Furthermore, the question did not allow for differentiation between Spanish instruction as a foreign language vs Spanish instruction in dual language immersion schools nor Spanish instructions received in weekly Saturday or Sunday schools. Despite this limitation, I believe examining different literacy experiences will be key in understanding the extent to which heritage speakers can process gender and number together.

Summary

In summary, the findings from Study 1 support the notion that heritage speakers process gender and number together. The implications of such a finding is further discussed in the General Discussion (see Chapter VII).

CHAPTER VI: *Les latines*, *les perres*, *y les cuartes*: The Role of the Animacy Hierarchy in the Comprehension and Production of Spanish Gender-Inclusive Morpheme –e

This chapter presents the second study of the dissertation, which includes two experiments that examine the comprehension (Experiment 1) and production (Experiment 2) of gender inclusive –e across the animacy hierarchy. Below, I first re-introduce the concept of grammatical gender in Spanish, followed by the concept of animacy and how it relates to gender inclusive language. Afterwards, research on gender-inclusive language is presented followed by the goals and predictions for the second study.

Grammatical Gender in Spanish

Grammatical "genders are classes of nouns reflected in the behavior of associated words" (Hockett, 1958, p. 231). For example, consider the following noun phrases in English and Spanish.

a) the red book → ***el** libr**o** roj**o***
b) the red table → ***la** mes**a** roj**a***

In English, the noun constituents (i.e., determiner and adjective) do not change according to nouns, while in Spanish they do. When noun constituents change according to the noun class, the language has grammatical gender. In Spanish, there are two classes that denote grammatical gender, a 'masculine' class which is typically marked with the morpheme –o (e.g., *libro*), and a 'feminine' class which is typically marked with the morpheme –a (e.g., *mesa*). According to Teschner & Russell, (1984), this pattern is consistent 99.9% of the time for nouns that end with –o and 96.3% of the time for nouns that end with –a, a pattern that native Spanish speakers exploit to facilitate predictive processing (see Chapter V for supporting evidence). Functionally, the two classes are different, with the masculine serving as the default and unmarked gender and the feminine serving as the exclusive marked gender. This asymmetry has been related to differential processing (see Beatty-Martínez & Dussias, 2019 and Beatty-Martínez et al., 2021 for a review). Although the terms 'masculine' and 'feminine' might seem like they refer to semantic gender, this is not the case for every noun. Figure 16 illustrates the relationship between grammatical gender, social gender, and animacy. While all nouns have an abstract grammatical gender, not all of them have a real-world semantic gender. Semantic gender is restricted to animate nouns. Even if some animate nouns lack semantic sex like *el sujeto* 'the$_{\text{MASC}}$ subject$_{\text{MASC}}$' and *la ballena* 'the$_{\text{FEM}}$ whale$_{\text{FEM}}$', the fact that they still possess grammatical gender highlights the distinction between abstract versus semantic representations of gender. The fact that semantic gender only affects animate nouns reflects the animacy hierarchy (Dahl, 2000b; Silverstein, 1976) a concept that is explained in the next section.

Animacy Hierarchy

	Example	Translation	Grammatical Gender	Biological/Social Gender
Animate: Humans	*el maestro* (m) *la maestra* (f)	the teacher the female teacher	Yes	Yes
	el sujeto (m) *la víctima* (f)	the subject the victim	Yes	No
Animate: Animals	*el perro* (m) *la perra* (f)	the dog the female dog	Yes	Yes
	el castor (m) *la ballena* (f)	the beaver the whale	Yes	No
Inanimate	*el libro* (m) *la mesa* (f)	the book the table	Yes	No

Abstract Semantic

Figure 16: Grammatical Gender, Biological/Social Gender and the Animacy Hierarchy

Animacy

To understand the concept of animacy, an ontological[9] definition is necessary. According to Dahl (2008), animacy is an ontological category that possesses two ontological types: one with animate entities and the other with inanimate entities. In general, animate entities have the capacity to perceive and act upon the environment, while inanimate entities do not. The differences between these entities may be reflected in grammatical rule systems such as grammatical gender, thematic role assignment, and case marking. In the realm of grammatical gender, the overlap between grammatical and semantic gender only occurs with animate beings. In thematic role assignment, the roles of 'agent' and 'experiencer' can only be occupied by animate beings in most languages. In case marking, accusative marking is restricted to animate beings in most languages, while overt ergative marking may be sometimes restricted to inanimate beings. The differential treatment between these categories is often described as a preferential organization of entities according to an animacy hierarchy, human > animal > inanimate (Silverstein, 1976; Dahl, 2000b). This hierarchy serves as a flexible continuum, where humans are at the top, animals are in the middle, and inanimate beings are on the lower end, with abstract concepts ranking the lowest (Dixon, 1979; Aissen, 2003).

Language Research on the Animacy Hierarchy

Cross-linguistic work shows the animacy hierarchy has consistent effects in linear ordering. For example, in Navajo (Hale, 1972), Lakhota (Van Valin & LaPolla, 1997), and Mayan (Jakeltek: Craig, 1977; Mam-Mayan: Minkoff, 2000) animate nouns must grammatically precede inanimate nouns when using transitive constructions. In these languages a sentence such as 'The rock hit the man' is ungrammatical and must be expressed passively to convey the same

[9] Ontology is the branch of Metaphysics that studies the existence and nature of fundamental types of entities (Quine, 1948)

meaning ‘The man was hit by the rock.’ Minkoff’s work (2000) provides further nuance by showing this effect is based on relative animacy. For example, in Mam-Mayan, ‘The dog bit the man’ is ungrammatical because humans outrank animals, but ‘The rock hit the window’ is grammatical because both nouns are inanimate. English has a similar restriction with ditransitive verbs, where a sentence may be considered ungrammatical when the indirect object is inanimate and the direct object is animate ‘The president sent England the ambassador’, while the prepositional form is typically considered more acceptable ‘The president sent the ambassador to England.’ Even in cases where the language does not have syntactic constraints on animacy ordering, there is a tendency to place animate before inanimate nouns, whether that be animate nouns preceding inanimate nouns in relative clauses (German: Kempen & Harbusch, 2004), or to place animate nouns heavily in subject positions and inanimate nouns in object positions (Swedish and Spanish: Dahl, 2000a; English: Dahl, 2000a; Snider & Zaenen, 2006; Norwegian: Øvrelid, 2004).

In the realm of production, there is a bias to use active voice with animate beings and passive voice with inanimates. Studies by Bock and colleagues (Bock, 1986; Bock & Loebell, 1990; Bock et al., 1992) show that participants employ the passive voice roughly up to 74% of the time when asked to describe pictures that contain an inanimate actor and an animate undergoer. The passive voice was practically never used when the actor was animate. Studies using delayed recall tasks show participants transform the original sentences to place animate arguments before inanimate arguments (Greek: Feleki & Braningan, 1997; English: McDonald et al., 1993; Japanese: Tanaka et al., 2005). This animacy production bias appears to be learned at an early age (see Paczynski & Kuperberg, 2011 for a review).

In comprehension, animacy has been shown to influence language through its association with thematic roles, but the exact stage at which it happens has been debated (Paczynski & Kuperberg, 2011). One of the first studies to document differences in processing related to animacy and thematic roles was Frazier and Rayner (1982). They examined animacy effects on the processing of reduced (e.g., 'The witness examined by the lawyer…') and unreduced (e.g., 'The witness that was examined by the lawyer') relative clauses. Frazier and Rayner found the processing cost of reduced relative clauses was smaller when the initial argument was inanimate (e.g., 'The evidence examined by the lawyer'). Because inanimate nouns are more often assigned to patient roles, Frazier and Rayner conclude this relationship makes the reassignment of the subject noun to the patient role easier. This finding was replicated in other studies (Ferreira & Clifton, 1986; Rayner et al., 1983). Other studies aimed at examining the time course processing of this effect found animacy effects occur during early stages of processing and the immediate availability of this information helps readers in thematic role reanalysis (MacDonald et al., 1994; Trueswell et al., 1994).

At the electrophysiological level, a smaller N400 response has been found for inanimate subject nouns compared to animate subject nouns, a finding that is interpreted as animate subjects being more easily assigned the role of agent (e.g., Kuperberg et al., 2003). This animacy effect seems to be strongest for readers who are more proficient readers and have higher working memory spans (Nakano et al., 2010). A semantic P600 has also been found in animacy-based selection restriction violations between verbs and arguments (see Kuperberg, 2007 for a review). To shed light on this semantic P600, Paczynski & Kuperberg (2011) conducted two ERP experiments contrasting competing accounts of this effect. Their findings show the semantic

P600 is driven primarily by propositional impossibility rather than thematic role reversibility or semantic attraction (Kim & Osterhout, 2005).

Finally, regarding grammatical gender processing and animacy two studies stand out (Sá-Leite et al., 2021; De La Garza Espinosa, 2022). Sá-Leite et al. (2021) examined grammatical gender processing during language production by means of a picture-word interference task in European Portuguese speakers. They manipulated the percentage of animate referents. A gender congruency facilitation effect was found only for inanimate nouns. However, when any amount of animate nouns was present, animacy overrode the grammatical gender interference. De La Garza Espinosa (2022) examined the semantic load of grammatical gender during processing. Participants were exposed to a noun manipulated for animacy which served as a prime. Afterwards they were exposed to a pronoun and had to decide whether it was masculine or feminine. A gender congruency effect was found, and this effect was strongest for animate nouns with biological gender. De La Garza Espinoza thus concludes grammatical gender carries more semantic load when nouns are biologically gendered.

Gender-Inclusive Language

Now that the previous section has discussed the concept of animacy and how it affects comprehension and production of nouns in different languages, we can turn our attention towards the current study's object of interest: Spanish gender-inclusive language. The phenomenon of gender-inclusive language is also an example of how grammatical gender interacts with the animacy hierarchy. Gender-inclusive language, also referred to as gender-fair language (Sczesny et al., 2016), emerged as a way to combat the asymmetric use of masculine generics across different languages (Hellinger and Bußmann, 2001) by replacing masculine forms (e.g., 'policeman') with gender unmarked forms (e.g., 'police officer'). Its goal is to reduce gender stereotypes and social discrimination. By default, this type of language only applies to human referents and serves as a perfect example of the 'personhood' concept Dahl (2008) considers quintessential for a definition of animate beings. As such, the use of gender-inclusive language follows the animacy hierarchy in different languages, but it is especially salient in grammatical gender languages like Spanish. Unlike genderless languages, like Finnish and Mandarin, where gender is neither marked in nouns nor pronouns, natural gender languages (i.e., languages where gender is marked in some suffixes and in pronouns) and grammatical gender languages must employ a gender-fair strategy when trying to avoid masculine generics. According to Sczesny et al. (2016), neutralization and feminization are two common gender-fair strategies. While the former uses gender-indefinite nouns over gender-marked terms and is a common strategy in natural gender languages like English, the latter explicitly represents women in language and is used alongside the former as a gender-fair strategy in grammatical gender languages. As such, a natural gender language like English may employ a neutralization strategy by using nouns with the '-person' suffix (e.g., 'chairperson') in lieu of the '-man' suffix (e.g., 'chairman'), while a grammatical gender language like Spanish may use epicenes[10] as a form of neutralization (e.g., *la ciudadanía* 'the citizens') or heteronymous word pairs (also called doublets) as a form of feminization (e.g., *niños y niñas* 'the male children and the female children'). In any case, the gender-fair strategy is employed for human referents only.

While both neutralization and feminization strategies may aid in achieving equal gender representations for men and women, they do not provide an option for someone who identifies outside of this binary. In response to this, LGBTQIA+[11] communities have proposed gender-

[10] Epicenes are animate nouns that employ only one grammatical gender to designate beings regardless of gender.
[11] The acronym stands for lesbian, gay, bisexual, transgender, questioning, intersex, and asexual. The plus sign denotes community members who relate to other forms of gender diverse identities and sexualities.

inclusive morphemes to represent a gender-neutral category in otherwise masculine and feminine grammatical gender systems. While this 'third' category exists alongside the traditional masculine/feminine grammatical gender, the innovative gender-inclusive morphemes only apply to nouns in which the grammatical gender also signals semantic gender. Thus gender-inclusive morphemes possess a dual purpose to serve: 1) as gender-neutral morphemes to masculine generics, 2) as markers of a non-binary identity.

Gender-Inclusive Morphemes

Gender-inclusive morphemes vary among languages. For example, French uses the *point médian* 'interpunct' (e.g., Pozniak et al., 2024), Slovene uses the underscore (e.g., Popič & Gorjanc, 2018), and German uses the * (e.g., Körner et al., 2022). In Spanish, several morphological alternatives[12] have been proposed, but –x (e.g., *Latinx*) and –e (e.g., *Latine*) are among the most commonly used (Papadopoulos, 2022). The –x because of its association with the unknown when the letter first entered the Spanish language (Salinas & Lozano, 2021), and the –e because it does not reliably signal masculine or feminine grammatical gender. For example, *la carne* 'the$_{\text{FEM}}$ meat$_{\text{FEM}}$' is feminine and *el valle* 'the$_{\text{MASC}}$ valley$_{\text{MASC}}$' is masculine. Similarly animate nouns that end with –e such as *estudiante* 'student' do not differ morphologically when they refer to a person's gender. Although scholars have pointed out the fact that these innovative gender-inclusive morphemes do not seek to abolish grammatical gender categories of other animate and inanimate nouns, but rather pursue gender neutrality as their goal (Vidal-Ortiz & Martínez, 2018), no study, to my knowledge has directly tested how the animacy hierarchy affects the processing of nouns with –e. Below, I review the emerging literature on innovative gender-inclusive morphemes. While there has been quite a lot of

[12] See Salinas and Lozano (2021) for other morphological alternatives.

discussion surrounding the use of gender-inclusive language in Spanish and its controversies (e.g. see González-Rivera, 2025, Muelas-Gil, 2026, Román Irizarry, 2021, and Zunino & Stetie, 2023 for reviews on this), research on the topic ranges from corpus work to psycholinguistic research measuring online processing, to studies examining spoken production.

Corpus Studies on Gender-Inclusive Morphemes in Spanish

Among the existing corpus work, six studies stand out (Díaz Colmenares, 2023; Kalinowski, 2020; Núñez-Román et al., 2023; Ryan et al., 2026; Samples, 2024; Slemp et al. 2020). Kalinowski (2020) examined the use of different forms of gender-inclusive language amongst Twitter[13] users in Argentina from 2007-2020. They compared different forms of feminization (e.g., *amigos y amigas*, *amig@s*, *amigos/as*, 'male friends and female friends') against the use of innovative morphological neutralizations –x and –e (e.g., *amigxs*, *amigues* 'friends'), and found that in the last two years, the latter forms were more frequently used than the former. A detailed account of the twenty most frequently occurring masculine generic forms in the corpus by inclusive strategy (see Figure 17) shows all forms were used with human referents. Slemp et al. (2020) examined written gender inclusive forms in a YouTube corpus from their previous study (Slemp et al., 2019: further explained in the production section ahead) and Twitter from 2012-2019. The heteronymous word pairs were used more in YouTube, while the –e was used more in Twitter. Slemp et al. (2020) also report the use of inclusive language in Twitter was accompanied by negative reactions from the public, while in YouTube the use of inclusive language was accompanied with more positive reactions. The authors attributed this difference to the presence of the *Real Academia Española* (RAE) 'Royal Spanish Academy', the institution responsible for setting prescriptive Spanish grammar rules and which has repeatedly

[13] Officially known as 'X' since July 2023.

declared gender-inclusive language strategies unnecessary (see Román Irizarry et al., 2025), on Twitter, but not on YouTube.

Díaz Colmenares (2023) examined gender-inclusive language strategies in Venezuelan Twitter users and found the use of -@ was the most common one. Similarly, Núñez-Román et al., (2023) examined the different gender-inclusive language strategies teachers in Spain used on Twitter. While the use of collective nouns was the most used strategy (e.g., *los niños y las niñas*

	Desdoblamiento barra	Desdoblamiento y/o	@	x	e
1	amigo (9856)	todo (7970)	todo (11608)	el (15460)	todo (9405)
2	el (7396)	el (3807)	amigo (9074)	todo (11431)	chico (7426)
3	todo (3337)	amigo (2998)	el (5229)	amigo (10147)	amigo (6108)
4	novio (3041)	chico (1788)	chico (1676)	chico (4462)	hijo (1583)
5	hijo (1558)	vecino (1781)	novio (1321)	hijo (2291)	nosotros (995)
6	chico (1205)	trabajador (668)	bienvenido (1266)	uno (2120)	muchacho (617)
7	boludo (858)	nene (662)	hijo (1010)	otro (1351)	otro (377)
8	uno (699)	pibe (524)	boludo (701)	novio (1315)	diputado (186)
9	hermano (683)	hermano (286)	solo (535)	nosotros (1259)	novio (175)
10	otro (632)	hijo (276)	alguno (528)	alguno (1245)	alumno (161)
11	alguno (593)	alumno (238)	otro (524)	boludo (1115)	uno (153)
12	pibe (550)	novio (236)	puto (500)	pibe (1060)	querido (129)
13	solo (543)	argentino (199)	agradecer (462)	lo (969)	bienvenido (123)
14	lindo (528)	neuquino (164)	afortunado (461)	trabajador (837)	sobrino (81)
15	lo (473)	solo (146)	hermano (439)	hermano (814)	egresar (64)
16	vecino (385)	abuelo (144)	vecino (410)	solo (794)	mogolico (39)
17	ese (348)	tio (144)	nosotros (379)	mucho (675)	jubilado (35)
18	profesor (341)	mucho (140)	loco (372)	lindo (625)	mismo (35)
19	sobrino (331)	maestro (126)	mucho (348)	vecino (602)	interesado (30)
20	nene (324)	riojano (125)	lindo (347)	loco (540)	indignar (29)

Figure 17: Twenty most frequently occurring masculine forms by gender-inclusive strategy (Kalinowski, 2020)

‘the boys and the girls’), among the innovative morphemes (what Núñez-Román et al. refer to as ‘textisms’) –@ was the most preferred followed by –x. Closer inspection of the counts used with –x showed it was exclusively used with human referents. Another study by Samples (2024) analyzed the forms of inclusive language used in the word *nosotros* ‘we’ amongst Twitter users in Madrid between 2010-2018. They found no inclusive forms were used in 2010, but by 2018 the –@ and the –x were the most used while –e was the least used.

While the corpus studies reviewed so far suggest the use of –x and –e is increasing after 2018, a more recent study shows this trend might be declining. In a recent study, Ryan et al. (2026) examined the use of inclusive language in tweets and newspapers from twenty different countries across the Spanish speaking world. For the Twitter data, they searched for singular and plural instances of –x and –e in the following community terms: *todos* ‘all’, *chicos* ‘guys/boys’, *amigos* ‘friends’, *algunos* ‘some.’ They also searched for identity terms related to the lemma Latin_. Regarding community terms, the total number of words with –x and –e rose from 2018 to 2020, but declined afterwards (see Figure 18A). In contrast, the results for the term *lenguaje inclusivo* ‘inclusive language’ in Spanish speaking newspapers was found to be increasing over time (see Figure 18B).

The discrepancy between social media usage of inclusive forms and the mention of inclusive language in newspapers was interpreted by authors as related to “each country’s unique gender agenda and the ongoing public debate regarding equality” (p. 962). Looking at the lemma Latin_ revealed an increased use of the term *Latinx* over the term *Latine* in the United States across time, but not in Argentina. Furthermore, the use of *Latinx* and *Latine* was predominantly classified by Twitter as coming from tweets written in English.

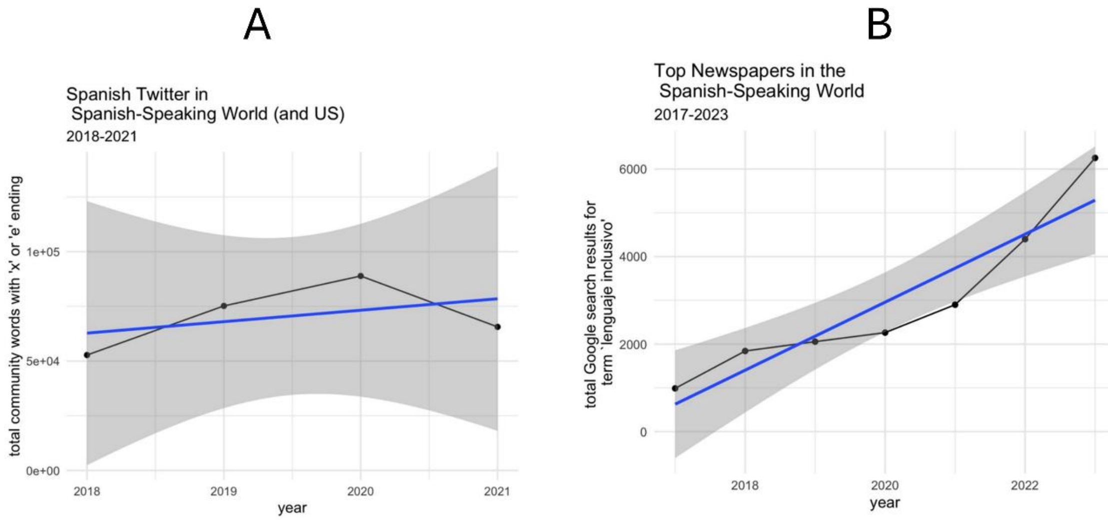

Figure 18: Gender-Inclusive Spanish Trends in Twitter and Newspapers from the Spanish-Speaking World (Ryan et al., 2026)

Although the corpus studies reviewed support the notion that the use of gender-inclusive language follows the animacy hierarchy, one final caveat is worth mentioning. Some minor exceptions have been pointed out by scholars before (e.g., Gil, 2020) and are commonly used as evidence against the rule-based nature of gender-inclusive language use. For example, Gil (2020) presents instances where first-person present indicative plural Spanish verbs are conjugated with –x and –e instead of –o, where –o is part of the stem *-mos* that signals first person plural in lieu of grammatical gender. An example of this is *somos* 'We are' which has appeared conjugated as *somxs* or *somes*. While Gil (2020) is not incorrect in signaling the use of gender-inclusive morphemes should not apply in verbs because Spanish does not mark grammatical gender in them and the forms are already neutral, I argue the use of verbs like *somxs* does not represent an animacy hierarchy violation. That is because the use of the first-person plural encodes an egophoric reference (Dahl, 2008) within the human hierarchy by including the self in the expression. Nevertheless, while instances of inanimate objects being gendered with inclusive morphemes can be frequently seen online as a form of satire (see Figure 19), I argue this use is restricted to satirical contexts where users violate the animacy hierarchy on purpose (see general discussion in Chapter VII).

Psycholinguistic Studies on Processing Spanish Gender-Inclusive Morphemes

Psycholinguistic research on gender-inclusive language has gained increasing attention across different languages in recent years (Cervone et al., 2025) and Spanish is no exception. In Spanish, psycholinguistic research has focused on examining the processing of gender-inclusive morphemes –x and –e against the masculine generic –o. Most of this work has been conducted at

Figure 19: Satirical use of –e in the inanimate noun phrase 'the screws'
Note. Source https://fity.club/lists/suggestions/memes-del-lenguaje/

the noun level (Durand-López et al., 2025; Stetie & Zunino, 2022; 2024; Zarwanitzer & Gelormini-Lezama, 2023;) and in pronouns with gender-neutral antecedents (Román Irizarry et al., 2025; Román Irizarry & Guzzardo Tamargo, 2025). The current findings show mixed results, with some studies finding greater processing costs for gender-inclusive morphemes and others not finding a difference at all. Stetie and Zunino (2022) examined how –x and –e were processed in stereotypical role nouns with high male stereotypicality (e.g., 'the carpenters') and low male stereotypicality (e.g., 'the teachers'). After reading the sentence, participants were asked to identify the possible referents evoked by the noun phrase. They found that nouns with masculine –o had significantly longer RTs at both the sentence level and question response level when compared to nouns with –x and –e. Regarding gender representations, –x (e.g., *lxs carpinterxs*) and –e (e.g., *les carpinteres*) evoked more gender-neutral representations. In a follow-up study, Stetie and Zunino (2024) examined the relationship between –x and –e and the processing of female stereotyped role names by means of a self-paced reading task. They found that the use of –e evoked longer RTs at the first spillover region and attributed it to the low cooccurrence of female stereotyped nouns and –e. Zarwanitzer and Gelormini-Lezama (2023) also examined the processing of gender-inclusive morphemes –x and –e against masculine generic –o in nouns and found longer RTs to sentences with gender-inclusive morphemes. Something to note is that the nouns included in their study did not control for the noun's stereotype (see the study materials in Zarwanitzer, 2019), but Stetie and Zunino (2022) did and still found there were no differences in sentence RTs between the masculine and the gender-inclusive morphemes.

Durand-López et al. (2025) examined processing of gender-inclusive morpheme –e at the adjective level by means of a self-paced reading task. Their materials included a carrier sentence, followed by the mention of four names, 2 masculine and 2 feminine, the verb *están* 'They are', and an adjective manipulated for gender agreement: *Mencionan que Juan, Silvia, Pedro y María están contentos/contentes en su trabajo.* 'They mention that Juan, Silvia, Pedro, and María are happy at their workplace.' They found longer RTs to adjectives with –e, both in the adjective region and the first spillover region. Moreover, a small subset of individuals who reported using gender-inclusive language from ‘frequently’ to ‘as much as possible’ showed no RTs differences between adjectives with masculine –o and inclusive –e. Taken, together the findings from Durand-López et al. (2025) support the claim initially put forward by Román Irizarry et al. (2025), who suggested the differences in findings between Stetie and Zunino (2022) and Zarwanitzer and Gelormini-Lezama (2023) where due to familiarity with gender-inclusive forms. That is, 75% of the participants in Stetie and Zunino reported being inclusive language users, while Zarwanitzer and Gelormini-Lezama do not report this data.

Research using gender-inclusive morphemes in pronouns that refer to gender-neutral antecedents offer additional insights into the time course processing of linguistic forms with –x and –e as well as identify other extralinguistic factors that moderate the processing of gender-inclusive morphemes. Román Irizarry and Guzzardo Tamargo (2025) examined how pronouns with –x (e.g., *todxs* ‘all’) were processed in comparison to pronouns with masculine –o (e.g., *todos*) and pronouns with feminine –a (e.g., *todas*) when gender-neutral nouns like *estudiante* ‘student’ were used by means of the eye-tracking technique. They also examined the effect of extralinguistic factors such as gender identity, sexual orientation, experience with gender studies, and attitudes towards gender-inclusive language. At the group level, they found pronouns with –x incurred greater processing costs than pronouns with –o and –a during early processing measures, but during late stages of processing it was pronouns with –a which incurred the greatest processing costs. This latter finding was interpreted as 1) evidence that the processing of –x is not insurmountable and can be overcome during late processing stages, and 2) the feminine plural is harder to integrate than the inclusive –x in generic contexts. The authors also made a distinction between online processing and comprehension measured via accuracy, which did not differ across the morphemes. Regarding extralinguistic factors, those who identified as LGBTQIA+ and those who identified as non-binary trended towards processing –x faster than –a across reading measurements.

In a follow-up study, Román Irizarry et al. (2025) examined 1) whether the processing of pronouns with gender-inclusive morphemes –x and –e in gender-neutral antecedents could be moderated by cognitive factors such as working memory and gender/sex diversity beliefs, and 2) whether the processing of pronouns with gender-inclusive morphemes differed from processing grammatical gender violations at the pronoun level by means of two self-paced reading tasks. The authors replicated the eye-tracking findings from Román Irizarry and Guzzardo Tamargo with self-paced reading. That is, pronouns with –x and –e temporarily disrupt processing but do not affect overall comprehension. Notably, the processing of pronouns with gender-inclusive morphemes was significantly moderated by gender/sex diversity beliefs: traditional views on gender increased RTs on pronouns with –x and –e, while stronger affirmation of diverse gender identities reduced RTs to pronouns with –x and –e on the first spillover region. A marginally significant effect was also found between higher working memory and reduced RTs on the second spillover region of pronouns with –x and –e. In contrast, the processing of grammatical gender violations was not moderated by working memory nor gender/sex diversity beliefs.

Production Studies on Spanish Gender-Inclusive Morphemes

When compared to the emerging literature on processing Spanish gender-inclusive morphemes, very few studies have examined gender-inclusive morphemes in spoken production (Slemp, 2020; Slemp et al., 2019). Slemp et al. (2019) examined the production of different Spanish gender-inclusive language forms in YouTube videos. Using 23 keywords that included demonyms with –@, –x, and –e (e.g., *mexican@s*, *mexicanxs*, *mexicanes*, ‘Mexicans’), they analyzed thirty YouTube videos with the goal of comparing written versus spoken forms of inclusive language. In writing, the most common gender-inclusive morphemes were –x, –@, and –e. In speech, however, heteronymous word pairs or doublets (e.g., *mexicanos y mexicanas* 'male Mexicans and female Mexicans') were the most common form of gender-inclusive language. Notably, all three gender-inclusive morphemes were pronounced in a way that matched these doublets. The second most common form of spoken gender-inclusive language was the use of [e] to pronounce gender-inclusive morphemes –x and –e. In another study, Slemp (2020) recruited Spanish speakers from Facebook and Twitter to inquire via survey which forms of gender-inclusive language they used when speaking. The most commonly reported form used in spoken language was –e , which came from Argentinian (n = 8) and Colombian (n = 3) Spanish speakers, followed by the heteronymous word pairs, which were spread more evenly across speakers from different parts of the Spanish speaking world. Two of the surveyed participants came back for interviews. One identified as genderfluid and came from Spain (P1), while the other was a cisgender woman from Argentina (P2). P1 expressed using indirect forms of non-binary language such as '*las personas*' because it was easier to use in spontaneous production, while P2 used more heteronymous word pairs and expressed using –e in less formal contexts. Something to note in both of Slemp's studies is that the production of gender-inclusive language forms occurred alongside the masculine generic, showing how both forms can coexist among inclusive language users

The Current Study

Now that evidence has been presented regarding animacy effects on comprehension and production as well as the link between gender-inclusive language and the animacy hierarchy, the current study seeks to test the following research questions. First, to what extent do Spanish heritage speakers, who may vary greatly in their exposure to gender-inclusive language, comprehend and produce gender-inclusive –e according to the animacy hierarchy? Many of the previous studies included speakers who reported using at least some form of inclusive language. However, more research needs to be conducted on native Spanish speakers with limited exposure to gender-inclusive language. Heritage speakers, whose Spanish is acquired natively at home as the first language and may have little to no knowledge regarding gender-inclusive language, can aid in this process. A second goal is to continue exploring the role of individual differences, specifically cognitive control and gender/sex diversity beliefs, on the processing –e. Is it the case the moderating effects are related to the processing of newer morphemes in general or do they aid specifically in the integration of –e in correct contexts (i.e., human referents)? For heritage speakers, do individual differences in Spanish proficiency and language dominance play a role in processing nouns with –e across the animacy hierarchy? In production, which individual differences are related to the production of –e across the animacy hierarchy? For this purpose, the current study has two experiments. Experiment 1 assesses the comprehension of –e across the animacy hierarchy by means of a self-paced reading task, while Experiment 2 explores the production of inclusive –e in oral narratives.

Based on previous research using the self-paced reading technique, I predict Spanish heritage speakers will have longer RTs to nouns with inclusive –e when compared to nouns with –o, but this effect will be moderated by the animacy hierarchy. In other words, processing costs should be smaller for humans, and larger for animals and objects. Regarding cognitive control, based on previous work finding inhibitory control can mitigate prediction costs (Beatty-Martínez et al., 2021; Zirnstein et al., 2018), better inhibitory control is expected to mitigate the prediction errors associated with processing –e in cases where the animacy hierarchy is violated (i.e., animals and nouns). Moreover, positive gender/sex diversity beliefs are expected to mitigate processing costs associated with processing –e in human referents. For Spanish proficiency and language dominance, higher scores are expected to reduce processing costs in general. Finally, for the production data, masculine and feminine grammatical gender human referents are expected to coexist alongside inclusive forms. Specifically, it is predicted that inclusive nouns with –e will only be produced in regards to human referents.

Methods: Experiment 1

Participants

Fifty heritage speakers of Spanish who were undergraduate students from Southern California universities were recruited and compensated for participation in accordance to IRB approval. To participate in this study, participants had to: be 18 years or older, have grown up speaking Spanish at home in the U.S., have corrected-to-normal vision, have no serious eye or neurological conditions, feel comfortable reading in Spanish, and have access to a computer with an integrated keyboard. Out of the 50 participants, six were excluded for data analysis. The details are explained further below in the data pre-processing section. The remaining 44 participants had an average age of 20.9 years old (*SD* = 3.3). Regarding gender identity[14], 39 self-identified with the feminine gender, while 5 self-identified with the masculine gender. Most of the participants (*n* = 37) reported Mexico as their Hispanic heritage country, followed by Guatemala (*n* = 3) and El Salvador (*n* = 1). Additionally, two participants reported having two Hispanic heritage countries: Mexico and Honduras, and Mexico and Colombia. Only one participant did not report their Hispanic heritage country. The full language history, self-rated proficiency, and gender-inclusive Spanish use of the analyzed sample can be found in Table 12.

[14] This study used gender self-identification in the same way as Román Irizarry et al. (2025). More details are described in this first paper.

Materials

Bilingual Language Profile

The Bilingual Language Profile (BLP; Birdsong et al., 2012) is a language questionnaire that places bilinguals on a language dominance continuum. This language dominance measure is a multidimensional construct that is calculated based on language history, language use, language proficiency, and language attitudes assessed via 19 questions in each language spoken by the bilingual (Birdsong, 2016; Olson, 2023). In this study, the BLP was administered in Spanish, therefore a positive number indicates Spanish dominant, while a negative number indicates English dominant, and a score near zero indicates a more balanced bilingual. On average, the heritage speakers in this sample were English dominant (M = -22.5, SD = 38.2).

Elicited Imitation Task

The Spanish version of the elicited imitation task (EIT) created by Ortega et al. (2002) was administered as an objective Spanish proficiency measure. This was the same version of the Spanish EIT that was used in the EEG clitic pronoun study. For scoring, see the Elicited Imitation Task subsection from Chapter V's Methods section. On average, the heritage speakers in this study performed at ceiling (M = 5.6, SD = .3). This indicates that although the sample was overall English dominant, proficiency in the native heritage language is very high.

Contextual and Individual Linguistic Diversity Questionnaire

The Contextual and Individual Linguistic Diversity Questionnaire (CILD-Q; Wigdorowitz et al., 2020) measures linguistic diversity both within the individual and in contextual exposure (i.e., where participants have lived for the longest period of time). It is comprised of three scales: multilingualism in context, which measures contextual use and societal practice of multiple languages aside from the dominant language in a given social

Table 12: Language History, Self-rated Proficiency, and Gender-Inclusive Spanish Use from the Participants Included in the Animacy and Grammatical Gender Comprehension Experiment

	Mean (SD)
Language history	
Age[a] of acquisition Spanish	.0 (.0)
Age of acquisition English	3.5 (3.2)
Years of study in Spanish	5.4 (5.3)
Years of study in English	15.2 (4.0)
Language use[b] in Spanish	18.3 (8.4)
Language use in English	35.5 (8.4)
Language dominance[c]	-22.5 (38.2)
Elicited Imitation Task – Spanish	5.6 (.3)
Self-rated proficiency[d]	
Speaking in Spanish	4.8 (.9)
Understanding in Spanish	5.4 (.8)
Reading in Spanish	4.8 (1.1)
Writing in Spanish	4.2 (1.4)
Speaking in English	5.4 (.9)
Understanding in English	5.7 (.5)
Reading in English	5.6 (.8)
Writing in English	5.5 (.8)
Gender-inclusive -e frequency use[e]	
Humans	2.2 (1.4)
Animals	1.7 (1.2)
Objects	1.6 (1.5)
CILD-Q[f]	
Multilingual Context	3.8 (0.5)
Multilingual Practice	3.9 (0.4)
Linguistic Diversity Promotion	3.9 (0.5)
Number of participants	$n = 44$

[a] Measured in years.
[b] Based on the average percentage of time during the week participants use Spanish and English with friends and family as well as at school and/or work. A score of 50 indicates participants used only one language a 100% of the time, while a score of 25 indicates participants use each language 50% of the time.
[c] A positive number indicates Spanish dominant, while a negative number indicates English dominant.
[d] Ratings range from 0 = 'not well at all' to 6 = 'very well'
[e] Ratings range from 0 = 'I don't use it at all' to 7 = 'I use it all the time.' A one-way ANOVA showed no differences between frequency use among the 3 categories.
[f] Scored on a five-point Likert scale ranging from 1 = 'strongly disagree' to 5 = 'strongly agree'

context (i.e., English in the United States), multilingualism in practice, which measures both direct and indirect linguistic exchanges and conversational exposure (e.g., the speaker mixing two languages, others mixing two languages), and linguistic diversity promotion, which measures societal and governmental endorsement and promotion of linguistic diversity. All scales are scored on a five-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). Although not included as variables in the main analysis, I report them here with the goal of providing a measure of linguistic diversity that can be used to compare against findings from studies in more or less linguistically diverse contents (see Chapter VII).

Gender/Sex Diversity Beliefs Scale

The Gender/Sex Diversity Beliefs scale (GSDB; Schudson & van Anders, 2021) measures beliefs surrounding the ontology of gender/sex by including both social constructionist (i.e., gender/sex is diverse and socioculturally contingent) and essentialist beliefs (i.e., gender/sex is a singular reality). It contains five subscales: Affirmation (of diverse gender identities), Gender Normativity (traditional views on gender), Uniformity (how similar people of the same gender are), Surgery (legitimacy of transgender surgeries), and Upbringing (how upbringing influences gender identity). All subscales used a 7-point Likert scale. In this study, I excluded the Uniformity subscale because the creators of the scale found it was the only factor that did not correlate with transgender prejudice (Román Irizarry et al., 2025; Schudson & van Anders, 2021). For the Affirmation and Upbringing subscales, scores closer to 7 indicate positive gender/sex diversity beliefs, while for the Gender Normativity and Surgery subscales a score closer to 7 indicates negative gender/sex diversity beliefs. In general, the participants in this study had overall positive gender/sex diversity beliefs (see Table 13).

Table 13: GSDB and AX-CPT Descriptive Statistics from Study 2, Experiment 1

	Mean (SD)
GSDB Subscale	
Affirmation	5.3 (1.2)
Gender Normativity [a]	1.8 (.8)
Surgery [a]	2.5 (1.3)
Upbringing	4.3 (1.8)
AX-CPT Measure	
AX error rate	.05 (.05)
AY error rate	.11 (.13)
BX error rate	.18 (.30)
BY error rate	.20 (.32)
AX Reaction Time (ms)	468.0 (106.0)
AY Reaction Time (ms)	598.0 (105.0)
BX Reaction Time (ms)	421.0 (132.0)
BY Reaction Time (ms)	433.0 (123.0)

Note. GSDB = Gender/sex diversity beliefs scale. AX-CPT = AX Continuous Performance Task.
[a] Indicates the subscale was reverse scored.

Gender-inclusive -e Frequency Use

Because the main goal of this study was to examine the role of the animacy hierarchy in processing nouns with inclusive morpheme –e, participants were explicitly asked via a 7-point Likert scale how often they used gender-inclusive language when referring to humans, animals, and objects. The 7-point Likert scale, ranged from 1, *No lo uso en absoluto* ‘I don’t use it at all’, to 7, *Lo uso todo el tiempo* ‘I use it all the time.’ For each category, they were given an explicit example (e.g., the use of *les latines* instead of *los latinos*). The examples were added to get a better estimate of participant’s use of gender-inclusive language, since previous work has shown Spanish speakers can have different conceptualizations of what ‘inclusive language’ means (see the discussion in Román Irizarry et al. 2025). The exact instructions and questions presented to participants are available in Appendix B. In general, the participants in this study reported using gender-inclusive Spanish very little, and differences between animacy categories were not significant.

AX-Continuous Performance Task

The AX- Continuous Performance Task (AX-CPT) without distractors (Gullifer et al., 2018; Ophir et al., 2009), shows a sequence of letters that form cue-probe pairs: AX, AY, BX, or BY. AX cue-probe pairs make up 70% of the trials, while each of the other conditions make up 10% of trials. Participants pressed "yes" only if they saw the cue "A" followed by the probe "X", and "no" when they saw any other cue-probe sequence. Because AX trials make up most of the trials, they establish a strong tendency to press "yes." This strong tendency creates an inappropriate expectancy bias for AY trials. Therefore, participants who rely greatly on cue information are likely to show increased error rates and slower reaction times in AY trials. This study follows the approach from previous research (Beatty-Martínez et al., 2021; Zirnstein et al., 2018) and uses AY reaction time as a measure of inhibitory control where longer RTs indicate greater difficulty with successful inhibition (i.e., worse inhibitory control) and shorter RTs indicate less difficulty with successful inhibition (i.e., better inhibitory control). The descriptive statistics for the AX-CPT can be found in Table 13.

Self-paced Reading Task

The self-paced reading task employed a non-cumulative word-by-word paradigm (Keating & Jegersky, 2015). A 3x2 animacy (i.e., humans, animals, and objects) by grammatical gender[15] (i.e., masculine, inclusive) manipulation was used to create the experimental items for the self-paced reading task (see Table 14). This resulted in a total of 48 experimental sentences, eight per condition (Jegersky & VanPatten, 2013). Each experimental sentence began with a

[15] While both the human and animal nouns selected admit masculine and feminine grammatical gender forms, the grammatical gender of object nouns is inherently fixed as either masculine or feminine. Because gender-inclusive morphemes are primarily used as alternatives to the masculine generic, only objects with masculine grammatical gender were chosen to avoid introducing object nouns with inherent feminine gender, which would have introduced an uncontrolled variable into the design.

prepositional phrase that established the context in which the subject noun phrase was situated (e.g., *En Estados Unidos…* 'In the United States'). Immediately after, the subject noun phrase appeared in one of the six conditions of interest (e.g., *los latinos*…. 'the$_{\text{MASC}}$ Latinos$_{\text{MASC}}$'), with the determiner and noun always agreeing in gender and number, which was consistently marked as plural.

Table 14: Example of the Experimental Sentences Used in the Self-paced Reading Task

Animacy	Grammatical Gender	
	Masculine	Inclusive
Humans	En Estados Unidos, los latinos representan un 18% de la población.[a]	En Estados Unidos, les latines representan un 18% de la población.
Animals	En ese albergue, los perros reciben mucho cariño y paseos diarios.[b]	En ese albergue, les perres reciben mucho cariño y paseos diarios.
Objects	En esta casa, los cuartos tienen mucho espacio.[c]	En esta casa, les cuartes tienen mucho espacio.

Note. The noun phrase appears underlined for illustration purposes. In the experiment they were not underlined. Geographical regions with compound names (i.e., Estados Unidos, Puerto Rico, República Dominicana) were presented as a single word instead of two.
[a] 'In the United States, Latinos represent 18% of the population.'
[b] 'In that shelter, the dogs receive a lot of affection and daily walks.'
[c] 'In this house, the rooms have a lot of space.'

Each animacy category contained a total of 16 nouns. To avoid gender stereotype processing effects associated with role names (see Stetie & Zunino, 2022 for a review), demonyms from the Spanish-speaking world were selected as the nouns for the human condition, with geographical regions serving as their contextual setting. In contrast, the animal and object nouns required the addition of a determiner between the preposition and the noun to sound more natural (e.g., *En esa cabaña…* 'In that cabin…') to sound more natural. This addition resulted in the critical region for human nouns appearing at word position 4, while the critical region for animal and object nouns appeared at word position 5. Furthermore, 48 distractor sentences were constructed to expose participants to structural variety beyond the experimental sentences. These sentences also contained humans, animals, and objects. All sentences had an average length of 9.2 words (SD = 1.4, [7, 13]) and were co-created by the first author and three undergraduate heritage speakers of Spanish to ensure accuracy and to verify that the selected nouns selected would be familiar to the study population. To ensure participants only read one version of each experimental sentence, experimental sentences were counterbalanced and pseudorandomized into two lists.

Procedure

All tasks were conducted in Spanish on the online behavioral research platform FindingFive (FindingFive Team, 2019). Each participant was monitored on Zoom to ensure the quality of online data collection. Participants were not required to share their screen, rather the Zoom room served as a virtual lab space where they could complete the experimental tasks at their own pace and ask clarifying questions if needed. Once they provided their informed consent, participants completed the BLP and CILD-Q followed by the self-paced reading task. In the self-paced reading task, participants read each sentence carefully by pressing the spacebar to reveal the next word and conceal the previous one. At the end of the sentence, participants were asked a yes/no comprehension question. After the self-paced reading task, participants completed the AX-CPT followed by the Spanish EIT and three other production tasks (further described in the Experiment 2 Methods section). In the final part, they first completed the GSDB scale and then answered the gender-inclusive –e frequency questions. Participants were offered breaks after the self-paced reading task and after the AX-CPT.

Data Pre-processing

To be included in the analysis, participants had to follow instructions, complete all of the parts in the study and score well enough in the AX-CPT to be able to calculate their RTs and error rate by each condition. This resulted in excluding six participants. For the final 44 participants, reaction times (RTs) were extracted from the noun with the purpose of examining early processing of gender-inclusive –e across the animacy hierarchy. Processing costs were operationalized as longer RTs in comparison to the baseline conditions. Only correct trials were included in the analysis. To control for outliers, RTs < 200 ms and > 3000 ms were excluded (Keating & Jegersky, 2015). This affected 3.2% of the data and affected all animacy conditions equally, while for the grammatical gender conditions slightly more inclusive conditions (2.2%) than masculine conditions (1%) were dropped. For this reason, when identifying relative outliers for windsorizing, nouns were grouped by grammatical gender categories. Then, RTs ± 2.5 SDs above the mean of their grammatical gender condition were replaced with their respective cutoff values (Keating & Jegersky, 2015). This affected 3.4% of the remaining data and was evenly distributed across grammatical gender conditions (masculine = 1.5%, inclusive = 1.8%). Descriptive statistics for the RTs by animacy and grammatical gender are shown in Table 15 and Figure 20. Finally, to prepare the RT data for statistical modeling, the *boxcox* function from the MASS R-package was used (Ripley et al., 2013) to inspect whether transformation was necessary to meet normality and homoscedasticity assumptions. The optimal transformation was achieved using the log scale, so a log10 transformation was applied (Winter, 2019).

Statistical Modeling

All the statistical analysis were conducted in R (Version 4.6.0; R Core Team, 2026) using the *lme4* package for mixed-effects modeling (Bates et al., 2015). The animacy factor was coded using repeated contrasts (Schad et al., 2020), which enabled a humans vs animals contrast as well as an objects vs humans contrast. The grammatical gender factor was sum coded as a 2-level factor where inclusive nouns were assigned a -0.5 value and masculine nouns were assigned a

+0.5 value. All continuous predictors were scaled. First, to answer research question 1, a model with maximal random effects structures for Subjects, a random intercept by item, and an Animacy*Gender interaction as fixed effects was built using maximum likelihood (model e).

Table 15: Average Reaction Time by Animacy and Grammatical Gender at the Noun Level

Animacy	**Grammatical Gender**		
	Masculine	Inclusive	
Humans	639. 7 (310.1)	792.3 (453.9)	715.4 (395.3)
Animals	565.3 (253.2)	779.5 (482.0)	668.0 (395.0)
Objects	629.0 (293.3)	878.9 (482.0)	751.6 (416.1)
	612.0 (288.6)	817.4 (473.8)	

Note. Reaction times are in milliseconds (ms). Standard deviations are presented in parentheses. For animacy, marginal means for each level appear at the end of each row. For grammatical gender, marginal means appear at the bottom of each column.

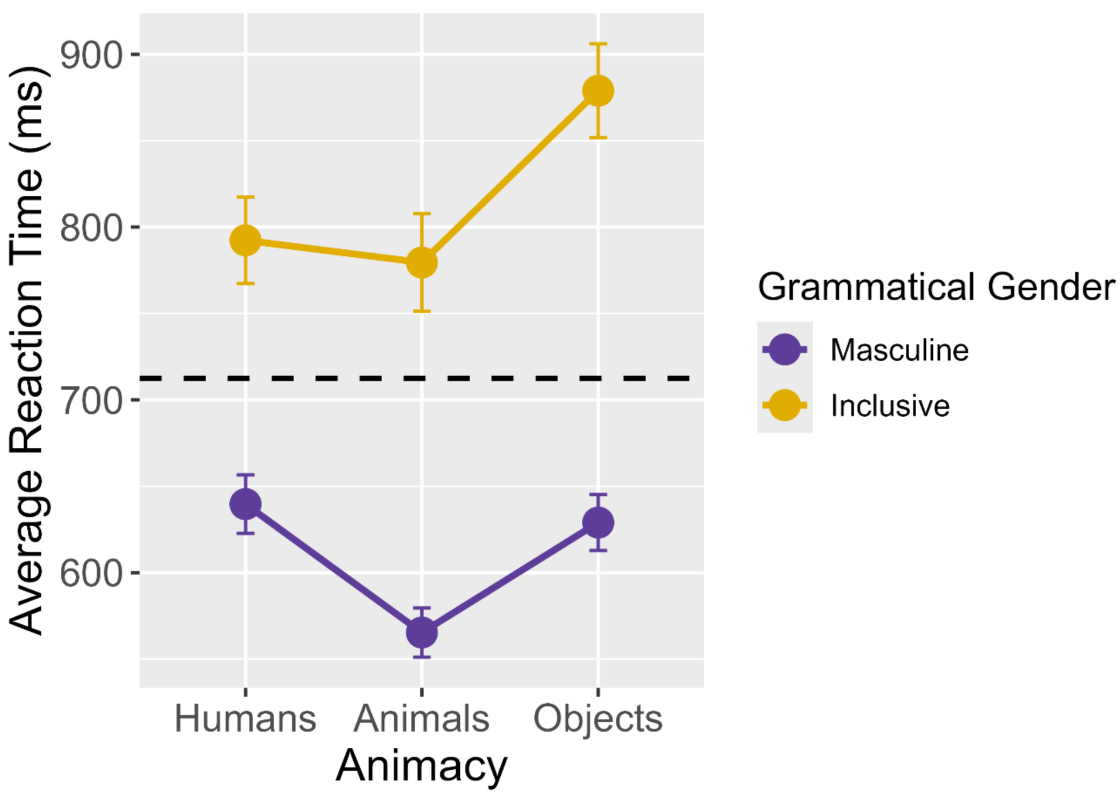

Figure 20: Average Reaction Time by Animacy and Grammatical Gender at the Noun Level

e) Log10RT ~ Animacy*Gender + (1 + Animacy*Gender | Subject) + (1 | Item)

If the model failed to converge, control = lmerControl(calc.derivs = FALSE) and optimizer = "bobyqa" was added to the model. If it still did not converge, the random effect that contributed the least variance was removed and if the model still did not converge, one random slope at a time was removed until convergence was achieved (Matuschek et al., 2017). This process resulted in a reduced maximal model (model f) which served as the baseline model.

f) Log10RT ~ Animacy*Gender + (1 + Gender | Subject) + (1 | Item)

To examine whether Spanish proficiency, language dominance, cognitive control, and gender/sex diversity beliefs moderate the processing of the animacy and gender interaction, each individual differences (ID) variable was included in an interaction with the two main variables of interest (model g). Correlations between each gender/sex diversity beliefs subscale and reaction times were conducted for every animacy and gender combination to determine which subscales would be included in the statistical models.

g) Log10RT ~ Animacy*Gender*ID + (1 + Gender | Subject) + (1 | Item)

To determine the best fitting model, the ANOVA function from the *lmerTest* package was used (Kuznetsova et al., 2017). All the p values were also computed using this package.

Results: Experiment 1

Baseline Model

The results from the baseline model are presented in Table 16. There was a main effect of grammatical gender, for which nouns with –e resulted in statistically longer reading times than nouns with –o, $b = -.109$, $SE = .012$, $p < .001$. There was also a statistically significant interaction between grammatical gender and the objects vs humans contrast, $b = -.053$, $SE = .018$, $p < .01$. In other words, the difference between masculine and inclusive nouns was greater for objects (*estimate* = -.13, *SE* = .02) than it was for humans (*estimate* = -.08, *SE* = .02).

Similarly, there was a marginally significant interaction between grammatical gender and the humans vs animals contrast, $b = .034$, $SE = .019$, $p = .069$. This suggests the difference between masculine and inclusive nouns was greater for animals (*estimate* = -.11, *SE* = .02) than it was for humans (*estimate* = -.08, *SE* = .02). In summary, the results show the difference between masculine and inclusive nouns is smallest for humans, marginally followed by animals, and greatest for objects. These results support the hypothesis that processing nouns with –e depends on the animacy hierarchy.

Alternate Models

Before running the alternate models with the individual differences of interest, noun lemma frequency was added to the baseline model to explore why the masculine human nouns trended towards the longest RTs for the animacy condition. For this purpose, lemma frequencies were extracted for each noun from the EsPal Latin American token subtitle database (Duchon et al., 2013). All lemma frequencies are available in Appendix B Table B1. Once extracted, lemma frequency was added as a control variable to test the robustness of the baseline model, but it did not improve overall model fit nor did it change the significance of the results (see Appendix B Table B2 for all model fit indices).

Although none of the alternate models significantly improved overall model fit when compared to the baseline model, some marginal and statistically significant relationships were identified. Starting with language dominance (see Table 17), a marginally significant interaction was found with grammatical gender. Specifically, as Spanish dominance increases the difference between masculine and inclusive nouns becomes greater, $b = -.022$, $SE = .012$, $p = .080$, (see Figure 21). Regarding cognitive control (see Table 18 and Figure 22), a statistically significant main effect was found for AY reaction time, $b = .036$, $SE = .016$, $p < .05$. In other words, as AY

Table 16: Noun Level Reaction Time Regression: Animacy*Grammatical Gender

	Estimate	*SE*
Fixed Effects		
Intercept	2.804***	.019
Humans vs Animals	.030	.022
Objects vs Humans	.019	.022
–o vs –e	-.109***	.012
Humans vs Animals*–o vs –e	.034•	.019
Objects vs Humans*–o vs –e	-.053**	.018
Random Effects	Variance	*SD*
Subject	.012	.111
Item	.003	.057
–o vs –e	.004	.064

* $p < .05$. ** $p < .01$. *** $p < .001$. • = marginally significant effect

reaction time increased (i.e., worse inhibitory control) there was an overall increase in noun reaction time.

Significant negative correlations were found between reaction times and gender/sex diversity beliefs (see Table 19). According to the criteria established by Gignac and Szodorai (2016), these effects ranged from small (< .10) to typical (<.20) and most of them affected the reaction times to masculine nouns regardless of animacy. Because significant correlations appeared for the Surgery and Upbringing subscales, these two were chosen for the alternate models. However, only Upbringing showed a marginally significant main effect, $b = -.028$, $SE = .016$, $p = .095$ (see Table 20 and Figure 23). In other words, as beliefs that gender is influenced by upbringing increase, overall noun reaction time decreases.

Table 17: Noun Level Reaction Time Regression: Animacy*Grammatical Gender*Language Dominance

	Estimate	*SE*
Fixed Effects		
Intercept	2.804***	.019
Humans vs Animals	.030	.022
Objects vs Humans	.019	.022
–o vs –e	-.109***	.012
Language Dominance	.000	.017
Humans vs Animals*–o vs –e	.034•	.019
Objects vs Humans*–o vs –e	-.053**	.018
Humans vs Animals*Language Dominance	-.007	.009
Objects vs Humans*Language Dominance	.009	.009
–o vs –e*Language Dominance	-.022•	.012
Humans vs Animals*–o vs –e*Language Dominance	-.002	.019
Objects vs Humans*–o vs –e*Language Dominance	.014	.018
Random Effects	Variance	*SD*
Subject	.012	.111
Item	.003	.057
–o vs –e	.004	.061

* p < .05. ** p < .01. *** p < .001. • = marginally significant effect

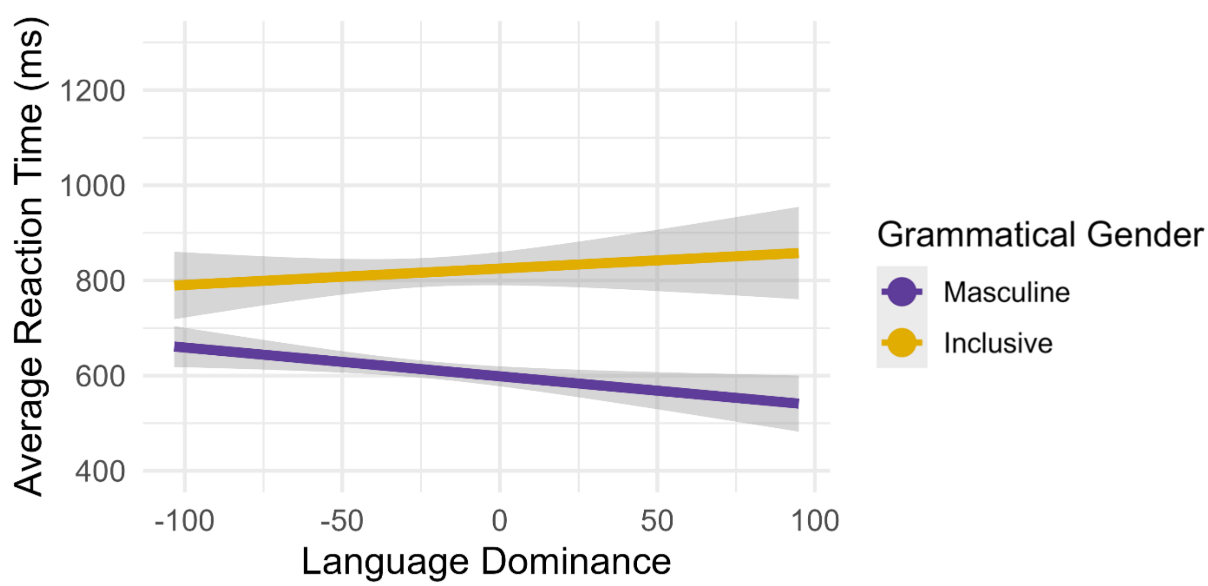

Figure 21: Grammatical Gender*Language Dominance Interaction

Table 18: Noun Level Reaction Time Regression: Animacy*Grammatical Gender*AY Reaction Time

	Estimate	*SE*
Fixed Effects		
Intercept	2.804***	.018
Humans vs Animals	.030	.022
Objects vs Humans	.019	.022
–o vs –e	-.109***	.012
AY Reaction Time	.036*	.016
Humans vs Animals*–o vs –e	.034•	.019
Objects vs Humans*–o vs –e	-.053**	.018
Humans vs Animals*AY Reaction Time	.004	.009
Objects vs Humans*AY Reaction Time	.004	.009
–o vs –e*AY Reaction Time	-.006	.012
Humans vs Animals*–o vs –e*AY Reaction Time	-.009	.019
Objects vs Humans*–o vs –e*AY Reaction Time	-.003	.019
Random Effects	Variance	*SD*
Subject	.011	.104
Item	.003	.057
–o vs –e	.004	.064

* $p < .05$. ** $p < .01$. *** $p < .001$. • = marginally significant effect

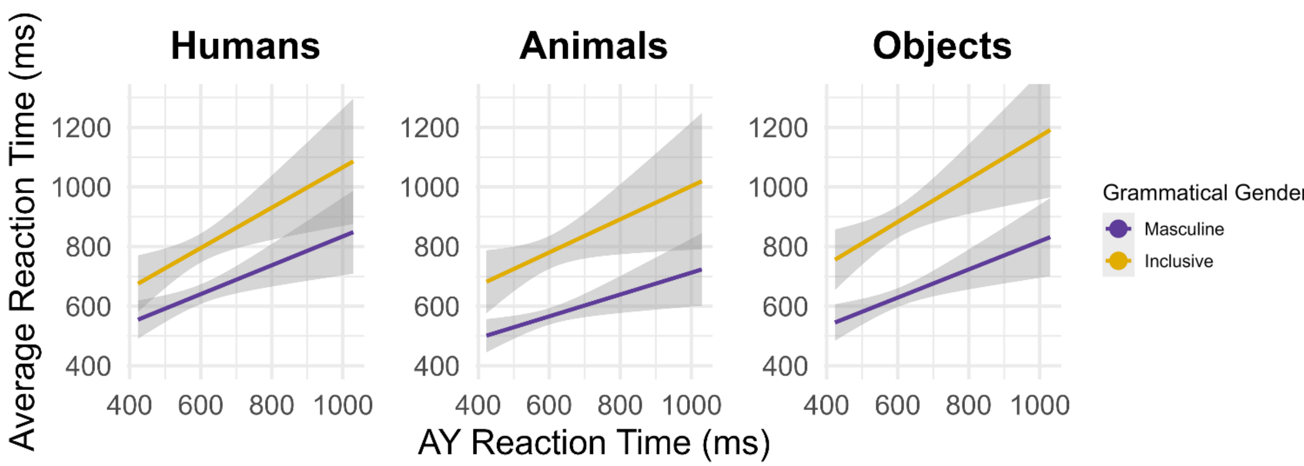

Figure 22: Inhibitory Control Main Effect

Table 19: Correlations Between Reaction Times and GSDB Subscale by Animacy*Grammatical Gender

		Grammatical Gender	
Animacy	GSDB Subscale	Masculine	Inclusive
Humans	Affirmation	.04	.07
	Gender Normativity	-.09	-.02
	Surgery	-.12*	-.06
	Upbringing	-.13*	-.10
	n	335	330
Animals	Affirmation	.00	.07
	Gender Normativity	.00	-.02
	Surgery	-.11	-.07
	Upbringing	-.14*	-.05
	n	317	292
Objects	Affirmation	.07	-.02
	Gender Normativity	-.02	.08
	Surgery	-.14**	-.07
	Upbringing	-.18**	-.12*
	n	327	315

$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

Table 20: Noun Level Reaction Time Regression: Animacy*Grammatical Gender*Upbringing

	Estimate	*SE*
Fixed Effects		
Intercept	2.804***	.019
Humans vs Animals	.030	.022
Objects vs Humans	.019	.022
–o vs –e	-.109***	.012
Upbringing	-.028•	.016
Humans vs Animals*–o vs –e	.034•	.019
Objects vs Humans*–o vs –e	-.054**	.018
Humans vs Animals*Upbringing	-.006	.009
Objects vs Humans*Upbringing	-.009	.009
–o vs –e*Upbringing	-.009	.012
Humans vs Animals*–o vs –e*Upbringing	.012	.019
Objects vs Humans*–o vs –e*Upbringing	.012	.018
Random Effects	Variance	*SD*
Subject	.011	.107
Item	.003	.057
–o vs –e	.004	.064

* p < .05. ** p < .01. *** p < .001. • = marginally significant effect

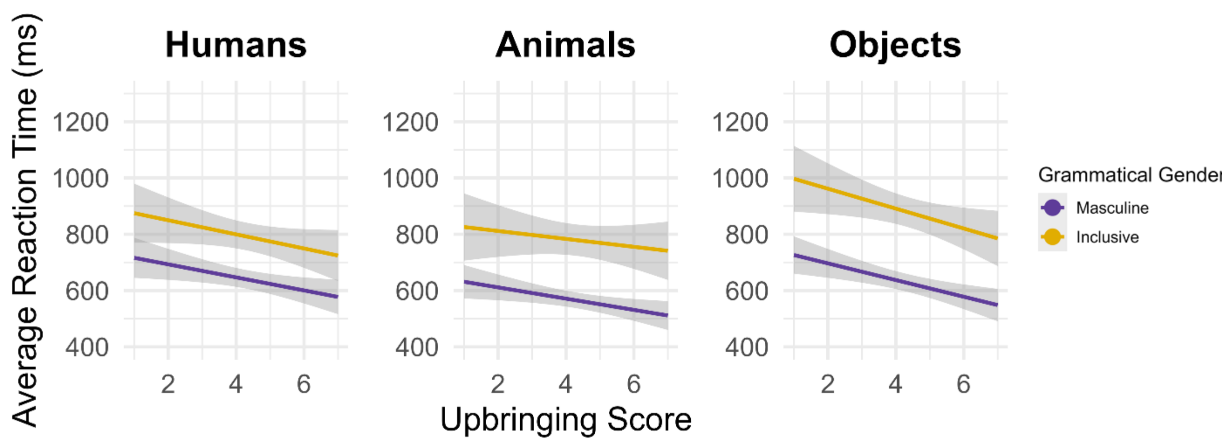

Figure 23: Upbringing Main Effect

Discussion: Experiment 1

The purpose of Experiment 1 was to assess how Spanish heritage speakers process nouns with gender-inclusive –e. Overall, nouns with inclusive –e resulted in longer RTs when compared to nouns with –o, a finding that aligns with prior psycholinguistic studies at the word level (Durand-López et al., 2025; Román Irizarry et al., 2025; Román Irizarry & Guzzardo Tamargo, 2025). This finding suggests that heritage speakers do not qualitatively differ from prototypical native speakers in their processing of gendered forms. That is, forms with the masculine generic are processed the fastest, while forms with inclusive –e are processed the slowest. They also serve to highlight that the heritage language experience in itself is not an excuse to exclude Spanish heritage speakers from studies examining the processing of gender-inclusive language in native Spanish speakers.

A new finding from this study is that the processing cost of nouns with –e was moderated by the animacy hierarchy. In other words, the processing cost of –e was smallest for humans, followed by animals, and greatest for objects. Strictly speaking, even if the marginally significant difference between animals and humans is disregarded, the difference between masculine and inclusive nouns is smaller for animate nouns than it is for inanimate objects. Two potential explanations could be responsible for the marginally significant effect. First, the sample size might be underpowered to detect the interaction. Second, it might be possible, that when participants use inclusive language very little, as in this sample, exposing them to a great amount of nouns with –e activates the association between grammatical gender and semantic gender for animate referents. It would be interesting to test this hypothesis with different groups of Spanish speakers who vary in their frequency use of gender-inclusive language.

A second goal of Experiment 1 was to test the extent to which individual differences moderate the relationship with inclusive –e and the animacy hierarchy. While none of the individual differences moderated the aforementioned interaction, findings from the alternate models provide further context. First, a statistically significant main effect was found for inhibitory control. One potential explanation for this finding could be that those who can inhibit faster an incorrect response, are also able to read faster in the heritage language, a task that might take heritage speakers a bit longer given the average years of schooling was 5.4 (*SD* = 5.3) and the average self-rated reading proficiency ranged from 'moderate' to 'well' (*M* = 4.8, *SD* = 1.1). Second, marginally significant effects were found for language dominance and the upbringing subscale. The first of these effects, seems to suggest that as heritage speakers become more dominant in their heritage language, the processing of masculine nouns becomes more efficient and the processing of inclusive nouns with –e increases. One potential explanation is that, at higher levels of English dominance, the effect of the English language, which lacks grammatical gender, reduces sensitivity to the masculine generic and therefore leads to smaller processing costs between masculine and inclusive nouns. Rather than framing this desensitization as something negative, it highlights the flexibility of the bilingual language system. Finally, regarding the marginally significant main effect for upbringing scores, no coherent explanation comes to mind as to why this positive gender/sex diversity belief reduced reading times across all conditions. In summary, while the alternate models identify inhibitory control and language dominance as important variables for future work to explore, they did not result in a better fit than the baseline model which supports the notion that heritage speakers process nouns with –e in accordance with the animacy hierarchy.

Methods: Experiment 2

Participants

The second experiment recruited the same participants used in the first experiment. This time, 49 Spanish heritage speakers were included in the analysis. Five participants who were excluded from Experiment 1 (see Experiment 1 Data Pre-processing section for exclusion criteria) but who were included in the second experiment self-identified as follows: 4 identified with the feminine gender, while 1 identified with the masculine gender. Regarding Hispanic heritage country, 3 reported Mexico, 1 reported Argentina, and 1 reported Colombia. The average age of the 49 participants together was 20.8 (*SD* = 3.2). Overall they were highly proficient in Spanish while also being dominant in English. Their full language history, self-rated proficiency, and gender-inclusive Spanish use can be found in Table 21.

Materials

Oral Narrative Tasks

To study Spanish heritage speaker's spontaneous production of gender-inclusive morpheme –e in different contexts, three oral narrative tasks were created. The first two oral narratives asked participants to tell a story about two different events. The events were 'A day with my family' (referred to below as 'family'), and an 'An event from the Latine Resource Center of your university' (referred to below as 'college'). Participants had the option to choose whether the story was based on real life events or fictional ones. Before revealing the name of each event, a fixation cross appeared in the center of the screen for 500 ms. Once the name of the event was revealed, participants had 1 minute to mentally plan out their narrative. After one minute passed, participants heard a tone that signaled it was time to start telling their story. They were given three minutes to do so. The order of the events was counterbalanced across

Table 21: Language History, Self-rated Proficiency, and Gender-Inclusive Spanish Use from the Participants Included in the Animacy and Grammatical Gender Production Experiment

	Mean (SD)
Language history	
Age[a] of acquisition Spanish	.4 (2.9)
Age of acquisition English	3.8 (3.9)
Years of study in Spanish	5.1 (5.1)
Years of study in English	15.1 (4.2)
Language use[b] in Spanish	18.6 (8.5)
Language use in English	35.2 (8.5)
Language dominance[c]	-22.8 (39.7)
Elicited Imitation Task – Spanish	5.6 (.5)
Self-rated proficiency[d]	
Speaking in Spanish	4.7 (1.0)
Understanding in Spanish	5.3 (.9)
Reading in Spanish	4.8 (1.1)
Writing in Spanish	4.2 (1.4)
Speaking in English	5.4 (.9)
Understanding in English	5.7 (.5)
Reading in English	5.6 (.8)
Writing in English	5.5 (.7)
Gender-inclusive -e frequency use[e]	
Humans	2.3 (1.4)
Animals	1.7 (1.3)
Objects	1.8 (1.7)
CILD-Q[f]	
Multilingual Context	3.8 (0.5)
Multilingual Practice	3.9 (0.5)
Linguistic Diversity Promotion	3.8 (0.5)
Number of participants	*n* = 49

Note. Age of acquisition statistics might be misleading due to one participant misunderstanding instructions. See Experiment 2 data analysis section for the justification of why this participant was retained in the final production analysis.

[a] Measured in years.

[b] Based on the average percentage of time during the week participants use Spanish and English with friends and family as well as at school and/or work. A score of 50 indicates participants used only one language a 100% of the time, while a score of 25 indicates participants use each language 50% of the time.

[c] A positive number indicates Spanish dominant, while a negative number indicates English dominant.

[d] Ratings range from 0 = 'not well at all' to 6 = 'very well'

[e] Ratings range from 0 = 'I don't use it at all' to 7 = 'I use it all the time.' A one-way ANOVA showed no differences between frequency use among the 3 categories.

[f] Scored on a five-point Likert scale ranging from 1 = 'strongly disagree' to 5 = 'strongly agree'

The third oral narrative task included the retelling of a short story seeded with instances of gender-inclusive morpheme –e. The purpose of this seeded story was to examine whether

Spanish heritage speakers would produce gender-inclusive –e with only human referents when primed to do so, while animals and objects maintained their respective masculine and feminine grammatical gender. The seeded story was created with the help from Zot GPT, an AI-powered chat designed for the University of California, Irvine community. The first prompt was entered in Spanish and asked ‘Can you make up a short story with inclusive language?’ Once the story was created and human referents were correctly gendered with morpheme –e, a follow-up question was asked in Spanish ‘Can you incorporate animals into the story as well?’ Thus, the short story *El festival del pueblo* ‘The town festival’ was created (see Appendix C). *El festival del pueblo* (referred to below as ‘festival’) contains 220 words and 15 instances of gender-inclusive –e. It tells the story about Max, a non-binary youth who had the idea of creating a festival for everyone in their town, including the animals. The text appeared left aligned in size 24 with FindingFive’s standard font, a sans serif font. Only the title of the story was bolded. Participants were instructed to read carefully the short story a maximum of two times and that they would have to retell it afterwards with as much detail as possible in the span of three minutes. No annotations were allowed for any of the narratives.

Procedure

All oral narrative tasks were conducted in Spanish on the online behavioral research platform FindingFive (FindingFive Team, 2019). They took place right after the Spanish EIT, just before the GSDB scale and the questions on gender-inclusive –e frequency use. While participants planned their narratives, a 1-minute countdown appeared on the screen. Similarly, while they completed their recordings a 3-minute countdown appeared on the screen. Participants had the option to end the recording as soon as they were done telling their story.

Data Preparation

To be included in the analysis, participants had to have data from at least 1 of the three production tasks and not have any missing gender-inclusive –e frequency use or GSDB data. Out of the 50 participants recruited for this study, only one did not complete these last two measures so they were excluded from analysis. The remaining 49 participants produced a total of 144 recordings. Only three audio files were missing: two from the college context and one from the festival. These missing audios all came from different participants who had technical difficulties. The remaining audios were transcribed with the help of an undergraduate Spanish heritage speaker. The resulting mini corpus contained a total of 23,639 words: 8,940 from the family context, 6,695 from the college context, 8,004 from the festival context.

Because the overall purpose of the study was to examine how the animacy hierarchy relates to the production of inclusive –e at the noun phrase level, noun phrases from each animacy category under study were extracted and classified according to their grammatical gender. This was done individually for each context. Additionally, in the festival context, mention of human and animal nouns counted if they appeared in the original story without a determiner. Objects[16] were counted if they appeared within the noun phrase. After reviewing the transcriptions, the presence of gender plural subject pronouns was found (i.e., *nosotros* 'we', *ellos* 'they') when referring to animals or objects. These were also included in the counts. For objects and animals, the grammatical gender of the noun was the one that counted, regardless of whether the participant made an agreement error, which tended to occur when the gender ending was not transparent (e.g., *la tren* instead of *el tren*, *la festival* instead of *el festival*). To count a noun as inclusive, the morpheme of the noun needed to change from –o or –a to –e (e.g., *los*

[16] Objects include the broader category of inanimate nouns, which also includes abstract and temporal concepts.

latines). Furthermore, when complex noun phrases were encountered each individual noun phrase counted. For example, *el Día de los Muertos* 'Day of the Dead' counted as two masculine gender instances, one for the 'object' category and the other towards the 'human' category. When participants repeated the same noun phrase, it was counted separately if it was done with emphasis, but only once if it constituted stuttering.

Furthermore, to understand in which contexts and type of word gender-inclusive –e appeared, instances where it was used as an inclusive morpheme were counted separately.

Results: Experiment 2

Figure 24 shows the total count of noun phrases and gendered subject pronouns together by context, animacy, and grammatical gender. Starting with context, the family narrative produced the largest number of counts, followed by the festival narrative, and the college narrative. When examining animacy, objects produced the largest number of counts, followed by humans, and animals. Regarding grammatical gender, masculine and feminine grammatical gender were approximately equally distributed for the objects category, with slightly more masculine than feminine objects in the festival context, and slightly more feminine objects in the college context. These minor differences are most likely due to the different contexts, while 'college' in Spanish is feminine (i.e., *la universidad*), the words 'festival' and 'town' are both masculine in Spanish (i.e., *el festival*, *el pueblo*). When referring to humans and animals, the masculine grammatical gender produced the largest number of counts when compared to the feminine and inclusive categories. Such an asymmetric account shows Spanish heritage speakers also rely on the masculine gender as the generic. The exception was the family context, where the masculine and feminine grammatical gender were evenly distributed. In this context, the feminine count is influenced by the mention of female family members (e.g., mothers, sisters) as well as the noun phrase ‘my family’ which is feminine in Spanish (i.e., *mi familia*). Although inclusive –e counts produced the least number of grammatical gender counts, they only occurred with human referents in the college and festival context. For the college context, the noun phrases occurred after the mention of a masculine noun: *mis compañeros o compañeres* ‘my classmates’ (participant 1), *los hispánicos o latines* ‘the Hispanics or Latines’ (participant 12). In the festival context, one participant mentioned the noun phrase *les niñes* ‘the children’ (participant 1), three mentioned the noun *niñes* (participants 37, 44, 45), and one participant mentioned *les comerciantes* ‘the merchants’ (participant 40).

Table 22 shows the total number of tokens with gender-inclusive –e produced across the three narrative tasks. Looking at word class, eight tokens were produced as adjectives, five tokens were produced as nouns, five tokens were produced as determiners, and one token was produced together as a determiner and noun pair. Within adjectives, the use of *latine* was the most common, including one variant that pronounced Latinx in Spanish as *latineks*. Within nouns, the use of *niñes* was the most common. Among the determiners, the plural inclusive article *les* was more common than the singular inclusive article *le*. All gender-inclusive –e tokens were produced in the college and festival contexts by twelve different speakers. However, neither their language history, nor their gender/sex diversity beliefs nor their gender-inclusive –e use differed descriptively from the mean.

Figure 24: Counts of Noun Phrases and Gendered Subject Pronouns by Context, Animacy, and Grammatical Gender

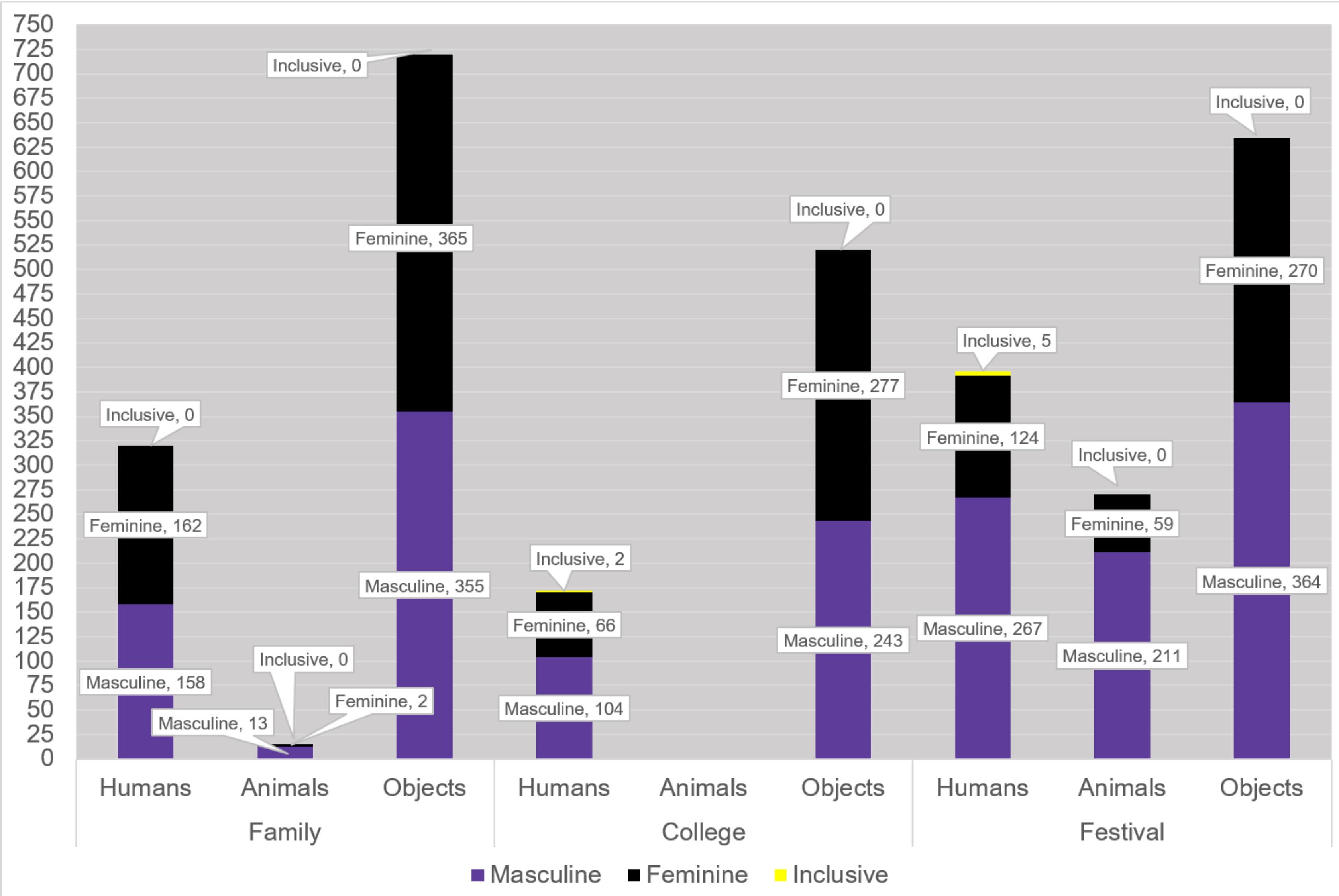

Table 22: Gender-inclusive –e Tokens Across Narrative Tasks

Token	Count	Word Class	Context
1. *el latine* 'Resource Center'	4	adjective	college
2. *el latineks* 'Resource Center'	1	adjective	college
3. *los grupos umm… latines*	1	adjective	college
4. *los estudiantes latines*	1	adjective	college
5. *los hispánicos o latines*	1	noun	college
6. *mis compañeros o compañeres*	1	noun	college
7. *niñes*	3	noun	festival
8. *les niñes*	1	determiner + noun	festival
9. *un joven llamade Max*	1	adjective	festival
10. *le pueblo completa* [a]	1	determiner	festival
11. *les profesor* [b]	1	determiner	festival
12. *les comerciantes*	1	determiner	festival
13. *les de colares* [c]	1	determiner	festival
14. *les estrellas* [d]	1	determiner	festival
Total	19		

Note. Tokens 9-14 come from the same speaker. See discussion section for more details on this case.

[a] Target structure = *el pueblo completo* 'the whole town'

[b] Target structure = *les profesores* 'the professors'

[c] Target structure = *las de colores* (*banderas*) 'the colored ones' (flags)

[d] Target structure = las estrellas 'the stars'

Discussion: Experiment 2

The purpose of Experiment 2 was to explore Spanish heritage speaker's spontaneous production of gender-inclusive –e against the spontaneous production of the masculine and feminine grammatical genders across the animacy hierarchy in three different contexts. Out of the three contexts, the family context produced the largest number of counts across animacy and grammatical gender followed by the festival context. This reflects Spanish heritage speaker's familiarity with spoken language in the family context, a hallmark of the heritage language experience. It also shows heritage speakers in general described the festival short story with great detail. In contrast, college produced the fewest number of counts possibly due to their college experience being in English, which might have limited their vocabulary as well as some participants saying they had not visited the Latine Resource Center at their university.

Regarding animacy, no specific differences were hypothesized in the number of counts. However, the ‘object’ category naturally occurred more frequently in oral narratives because speakers must describe an event including details such as date and time, the action that took place, and the emotions and feelings involved. In contrast, when telling oral narratives, human referents were produced much more frequently than animal referents.

In Spanish heritage speaker’s vocabulary, the grammatical gender of objects was approximately evenly distributed among the masculine and feminine gender. A caveat of this finding is that some noun phrases had gender agreement errors between the determiner and the noun, especially when the noun possessed a non-transparent ending (e.g., *un vez* instead of *una vez* ‘one time’). These agreement errors were not reflected in the counts because the main goal of Experiment 2 was to observe whether inclusive –e would be used according to the animacy hierarchy, and not to assess knowledge of the grammatical gender of object nouns. Nevertheless, a follow-up analysis would most likely reveal the masculine determiner is preferred over the feminine when nouns possess non-transparent endings, a pattern attested in Spanish-English codeswitching corpora (e.g., Parafita Couto & Gullberg, 2017). On the other hand, for humans and animals there was a strong masculine generic bias. This pattern is in line with normative gender rules and shows Spanish heritage speakers, like prototypical native speakers, also default to the masculine gender when referring to animate beings.

When examining the inclusive –e counts, they (*n* = 12) strongly refer to human referents (11 out of 19), and not objects nor animals, thus supporting the relationship between the animacy hierarchy and inclusive –e. Even the five instances were heritage speakers described and event from their Latinx/Latine Resource Center, the inclusive –e appears when the adjective Latine is said in Spanish (i.e., *el latine*). Although, *el latine* describes the inanimate referent 'resource center', heritage speakers do not fully break the animacy hierarchy to extend –e to other adjectives whose stem refers to inanimate objects. For example, the word 'resource' serves as an adjective in the noun phrase 'resource center' that modifies the word 'center', both words that refer to inanimate realities in the real world. But heritage speakers do not say *le centre de recurses latines* instead, they say *el latine* 'Resource Center' where the inclusive morpheme –e is preserved in the adjective *latine*, which describes a group of people from Latin American descent. Thus, it seems that adjectives that describe associations with human referents, such as demonyms, preserve their human animacy hierarchy and import it to other cases where they describe inanimate realities that are associated with human referents (e.g., *el movimiento chicane/latine* see Appendix C Figure C1 for an example). This direct importation evidences the natural bidirectional influence between a bilingual's two languages, where English's lack of grammatical gender allows heritage speakers to directly import *Latine* into Spanish. Testing this observed relationship between demonyms and inclusive –e in a broader sample of Spanish-English bilinguals, would allow researchers to better disentangle whether this is characteristic of the heritage language experience or whether it is just a general nuance of the animacy hierarchy that is present in highly proficient Spanish-English bilinguals. Nevertheless, due to the reasons explained above, all inclusive –e instances produced in the oral narratives applied in cases where the referent was human or the word inflected represent association with humans.

One small exception to the general trend described above appeared for one participant, who produced six different instances of inclusive –e. Out of these six instances, three were used with human referents and three were used with inanimate referents. The human instances included two counts that followed the usage pattern produced by the other participants (i.e., *un joven llamade Max*, *les comerciantes*) and one instance that contained a number agreement error (i.e., *les profesor* instead of *les profesores*). The remaining instances used the inclusive determiner in its singular (i.e., *le*) or plural (i.e., *les*) form when referring to inanimate entities, with two instances containing gender agreement errors (i.e., *le pueblo completa* instead of *el pueblo completo*) or phonological errors (i.e., *les de colares* instead of *las de colores*). This participant scored an average of 2.6 on the Spanish EIT, which places them five standard deviations below the average score. In addition, they had a very hard time following the spoken instructions on Zoom and entered '20+' as the age of acquiring both languages, despite confirming via email exchange, written in English, they were eligible to participate in the study. In the self-paced reading study, they also scored the lowest when compared to the rest of the sample included in the comprehension analysis. Despite these issues, they were retained in the production experiment because they were the only participant who 1) repeatedly used inclusive determiners and 2) used these inclusive determiners with inanimate nouns. Despite the participant's low proficiency in the heritage language, inclusive –e was used in cases where its usage applied but also in cases where its usage rule did not apply. Taken together, the results suggest extracting the animacy hierarchy of –e depends on Spanish proficiency.

One final caveat when interpreting these results is that the few instances were inclusive –e were produced contained a sort of 'prime', with the festival context including the largest number of seeded instances and the college event including the actual name of a resource center for students from Latin American descent. Thus, the claim is not that Spanish heritage speakers will always default to inclusive –e over –o, but rather encountering inclusive –e seems to bolster its production in accordance to the animacy hierarchy. Testing this hypothesis in a future psycholinguistic experiment would be the next step.

Summary

In summary, the findings from Study 2 support the notion that Spanish heritage speakers can comprehend and produce gender-inclusive –e across the animacy hierarchy. The implications of such findings are further discussed in the General Discussion (see Chapter VII).

CHAPTER VII: GENERAL DISCUSSION

The overall goal of this dissertation was to examine morphosyntactic variation in Spanish heritage speakers. For this purpose, the processing of traditionally complex structures and complex innovations was assessed using different measures of online processing as a way to tap into heritage language competence. In this chapter, I use the evidence from both studies to frame Spanish heritage speaker's morphosyntactic variation as a result of the flexibility and openness of the bilingual language system. This account places the focus on understanding morphosyntactic variation contextually, while moving away from long held deficit-based perspectives on heritage language acquisition (e.g., Rothman et al., 2023). Below, the discussion is divided into what the findings from the complex structures and the complex innovations tell us about the broader picture.

Knowledge of Complex Structures: Clitic Pronouns

Both the behavioral and electrophysiological data from the clitic pronoun processing study showed strong evidence that Spanish heritage speakers process gender and number together. Such a finding extends the representational economy claim put forth by Scontras and colleagues (2018) to electrophysiological data. According to Scontras et al. (2018), heritage speakers shrink their feature inventory, so it includes fewer nodes, a process that is possible due to the feature categories being close to each other on the nominal scale. They also argue that the processing of gender and number together is not in itself a reason to doubt Spanish heritage speakers' native speaker status. Instead, the fact that the bundling of the two features proceeds according to Spanish's structural constraints highlights the native knowledge of the heritage language. To their argument, I would also add that the bundling of the two features together further supports the notion of cross-linguistic interference documented in bilingual speakers (see

Chapter I). In a way, it seems like the bundling process is influenced by the features of the societal language English, which marks number agreement in nouns but lacks grammatical gender altogether. While such an interpretation is coherent for the heritage bilinguals, it does not explain why the prototypical native speakers (who were international students in the US) or the highly advanced L2 Spanish bilinguals from Rossi et al. (2014) processed gender and number independently. A possible explanation lies in the differences the three group of bilinguals have with Spanish instruction, which is explicitly formal and more uniform for the L2 group and prototypical native speaker group, and greatly varied for the heritage speakers. Therefore, it is quite possible that the processing of gender and number together arises when there is very little to no formal Spanish schooling.

While the findings from the clitic pronoun study align with Scontras et al., (2018) they differ from previous work that finds heritage speakers process gender and number independently (Fuchs et al., 2026). There could be several explanations for these observed differences. First, the observed differences between this study and Fuchs et al.'s might lie in the structure that is examined, which has been the noun phrase, and the task employed which is predictive processing. Because predictive processing implies actively looking for a target noun after hearing instructions (e.g., *Encuentra el…*/*Encuentra la…* 'Find the'), it could be that such a process bolsters the gender and number representations, whereas a more 'passive' task like RSVP might not. Although Study 1 employed the event-related potential technique which reflects real time processing with an incredibly fast temporal resolution, it could be that the bundling of the two features happens when reading is involved (Scontras and colleagues also assessed their heritage speakers via reading), a skill that was developed less in the current sample (i.e., less years of Spanish schooling, strong preference for reading in English, lower self-rated

Spanish reading ability). This account would need to be followed up with an EEG study that presents sentences orally. Such a study could shed light on why potential literacy effects might still happen at high levels of Spanish proficiency.

Second, it might the case that the processing of gender and number arises in cases of increased syntactic complexity. Scontras et al. (2018) examined gender and number agreement in embedded clauses: *El niño considera el artículo en el periódico terriblemente aburrido…*'The boy considers the news item in the newspaper terribly boring…' (underline words marked agreement), while the current study examined clitic pronouns. While clitics have been found to be fairly robust in heritage speech (e.g., Polinski, 2018), they represent a structure that is acquired relatively late regardless of monolingual or bilingual upbringing (e.g., Domínguez, 2003). To test whether sensitivity of gender and number depends syntactic complexity, the same group of heritage speakers would need to be tested in their online processing of gender and number across different types of structures that vary in their syntactic complexity.

Third, it might be the case that the observed differences occurred due to the fact that both gender (i.e., masculine and feminine) and number (i.e., singular and plural) categories were collapsed together in order to facilitate a comparison with Rossi et al., (2014). Given how both heritage speakers and prototypical native speakers have been found to be sensitive to the asymmetrical distributions of each feature, it might the case that one of the levels of each feature might be driving the effect of bundling gender and number together.

In summary, despite the limitations and considerations expressed above, the findings regarding the bundling of gender and number together when reading syntactically complex structures in the less literate language show how the bilingual language system can adapt to maximize real-time processing efficiency. Such a change is consistent with the idea that heritage grammars are quite flexible and open to restructuring. This idea of 'oppenness' is explored in the next section.

Knowledge of Complex Innovations: Gender-Inclusive –e Across the Animacy Hierarchy

The major findings from both experiments in Study 2 is that, despite using very little the gender-inclusive –e morpheme, Spanish heritage speakers comprehend and produce gender-inclusive morpheme –e according to the animacy hierarchy. Even though only a small subset of the participants produced –e when referring to human referents in 'primed' contexts, the fact the sample as a whole was able to comprehend forms they do not produce is remarkable. Moreover, the greater processing costs for object nouns with –e, support the notion that this type of animacy category constitutes a violation for the use of gender-inclusive language. In production, only the participant with the lowest Spanish proficiency produced inclusive determiners with objects, thus highlighting the role proficiency plays in extracting the animacy hierarchy of gender-inclusive language use. It also supports the notion that when inanimate objects appear with –e, the violation is done on purpose to achieve a satirical effect.

These findings are in line with psycholinguistic models that link comprehension and production processes (e.g., Dell & Chang, 2014; MacDonald, 2013). According to MacDonald's (2013) Production-Distribution-Comprehension (PDC) account, utterance planning in production (P) is complex and speakers implicitly follow certain biases to reduce burdens and improve fluency. Such biases are repeated over many utterances and individuals, and a distribution (D) is created based on language use. Language perceivers then implicitly learn the statistical regularities from their linguistic input and use it to guide their comprehension (C). In the case of grammatical gender and animacy, Spanish heritage speakers, like other native speakers, implicitly learn the masculine generic rule as the default during production and use this knowledge to facilitate its processing during comprehension. In contrast, the gender-inclusive –e is less frequently produced in their statistical landscape which creates greater processing costs at the noun level. However, the reduced processing cost observed for human referents shows Spanish heritage speakers have access to the statistical regularity that governs the relationship between the animacy hierarchy and the use of gender-inclusive language, particularly the morpheme –e, and can use it to guide their processing during comprehension. Similarly, Dell and Chang's (2014) P-chain framework provides a complimentary explanation to the results. According to this framework, "**P**rocessing involves **P**rediction, **P**rediction is **P**roduction, **P**rediction leads to **P**rediction error, **P**rediction error creates **P**riming, **P**riming is im**P**licit learning, im**P**licit learning is the mechanism for acquisition/adaptation of **P**rocessing, **P**rediction and **P**roduction, and **P**roduction provides the input for training **P**rocessing" (p. 2). Crucial for my interpretation is the fact that prediction ability allows for seamless transfer towards production, which is seen in the comprehension and production data both supporting the use of gender-inclusive –e in accordance with the animacy hierarchy. While the fact that instances with –e were produced the least in the natural production tasks might seem inconsistent with both the PDC approach and the P-chain, I argue it is not. Slemp's (2020) interviews with gender-inclusive language users show its use is highly contextual and more present in informal contexts. Therefore, it might be the case that Spanish heritage speakers also employ gender-inclusive language contextually. For this purpose, future work on the topic could employ tools from Personal Social Network Science to better understand when and with whom Spanish speakers use different forms of inclusive language (e.g., Cuarteto et al., 2023).

Taken together, the results underscore the fact that heritage language status in itself is not a reason to exclude heritage speakers from research on the acquisition of gender-inclusive forms in Spanish. This supports the justification put forth by Román Irizarry et al. (2025) who included both prototypical native speakers of Spanish along with Spanish heritage speakers in their sample and controlled for language dominance effects. Furthermore, work on heritage speakers may be used as a model for understanding how innovative forms may be acquired under linguistically diverse contexts. Although not examined as a variable of interest, the linguistic diversity measures collected in this study were quite high and comparable with Wigdorowitz et al.'s (2020) South African sample. In the future, they may be used to test the extent to which environmental factors related to linguistic diversity may shape the acquisition of innovative forms across different monolingual and bi/multilingual communities (e.g., Bice & Kroll, 2019). While not acknowledged enough in psycholinguistic research, the Spanish speaking world in Latin America and Spain coexists alongside a variety of other minority languages including indigenous ones (e.g., Barros & Orlandi, 2026; Morales-Gálvez & Pérez-Lozano, 2024).

One final note worth mentioning concerns the role of psychosocial and cognitive factors in the processing of innovative gender-inclusive language forms. Román Irizarry et al. (2025) found gender normativity scores moderated the processing of pronouns with –x and –e, but the current study did not. One important distinction between the two studies is that the heritage speakers from the previous study had slightly more traditional views on gender and their scores were spread out wider (M = 2.1, SD = 1.5) when compared to the heritage speakers included in the comprehension experiment (M = 1.8, SD = .8). To fully address the effect of gender/sex diversity beliefs on the processing of Spanish gender-inclusive forms, a broader sample with representative scores across the full spectrum of gender/sex diversity beliefs is needed as well as a broader sample with different attitudes towards the use of gender-inclusive language (Román Irizarry & Guzzardo Tamargo, 2025). Finally, while cognitive control only resulted in a main effect via better inhibitory control facilitating general processing in the heritage language, future work should examine how this effect varies across other types of native speakers and the extent to which the moderation of cognitive control mitigating prediction costs might be related to being a better reader in the language under study. While the heritage speakers in Study 2 were highly proficient in Spanish, their years of formal Spanish schooling were very little when compared to their formal English schooling.

Conclusion

In conclusion, the findings from the current dissertation support the notion that heritage grammars are flexible and open to restructuring and this openness also seems to aid in the comprehension and production of innovative forms. Research on heritage languages is much needed, as they represent the most common form of bilingualism in the world, but are the least studied (e.g., American Academy of Arts and Sciences, 2017; González Melón et al., 2026). The findings from this dissertation call upon researchers and practitioners to move away from deficit-based perspectives to a more strength-based approach that focuses on understanding the benefits of having a more open language system.

REFERENCES

Aissen, J. (2003). Differential object marking: Iconicity vs. economy. *Natural Language and Linguistic Theory*, *21*, 435-483.

Alemán Bañón, J., Fiorentino, R., & Gabriele, A. (2012). The processing of number and gender agreement in Spanish: An event-related potential investigation of the effects of structural distance. *Brain Research*, *1456*, 49–63. https://doi.org/10.1016/j.brainres.2012.03.057

Alemán Bañón, J., Miller, D., & Rothman, J. (2017). Morphological variability in second language learners: An examination of electrophysiological and production data. *Journal of Experimental Psychology: Learning Memory and Cognition*, *43*(10), 1509–1536. https://doi.org/10.1037/xlm0000394

Alemán Bañón, J., & Rothman, J. (2016). The role of morphological markedness in the processing of number and gender agreement in Spanish: an event-related potential investigation. *Language, Cognition and Neuroscience*, *31*(10), 1273–1298. https://doi.org/10.1080/23273798.2016.1218032

American Academy of Arts and Sciences. (2017). *America's languages: Investing in language Education for the 21st century.*

Babou, C. & Loporcaro, M. (2016). Noun classes and grammatical gender in Wolof. *Journal of African Languages and Linguistics*, *37*(1), 1-57. https://doi.org/10.1515/jall-2016-0001

Barber, H., & Carreiras, M. (2003). Integrating gender and number information in Spanish word pairs: An ERP study. *Cortex*, *39*, 465–482.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Barros, M. C. D., & Orlandi, E. P. (2026). *Políticas del lenguaje en América Latina*. R. E. Hamel (Ed.). De Gruyter.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1),1–48. http://dx.doi.org/10.18637/jss.v067.i01

Beatty-Martínez, A. L., & Dussias, P. E. (2017). Bilingual experience shapes language processing: Evidence from codeswitching. *Journal of Memory and Language*, *95*, 173–189. https://doi.org/10.1016/j.jml.2017.04.002

Beatty-Martínez, A. L., Bruni, M. R., Bajo, M. T., & Dussias, P. E. (2021). Brain potentials reveal differential processing of masculine and feminine grammatical gender in native Spanish speakers. *Psychophysiology*, *58*(3). https://doi.org/10.1111/psyp.13737

Beatty-Martínez, A. L., & Dussias, P. E. (2019). Revisiting masculine and feminine grammatical gender in Spanish: Linguistic, psycholinguistic, and neurolinguistic evidence. *Frontiers in Psychology*, *10*(751), 1–11. https://doi.org/10.3389/fpsyg.2019.00751

Beatty-Martínez, A. L., Navarro-Torres, C. A., Dussias, P. E., Bajo, M. T., Guzzardo Tamargo, R. E., & Kroll, J. F. (2020). Interactional Context Mediates the Consequences of Bilingualism for Language and Cognition. *Journal of Experimental Psychology: Learning Memory and Cognition*. https://doi.org/10.1037/xlm0000770

Benmamoun, E., A. Albirini, S. Montrul, and E. Saadah. 2014. Arabic plurals and root and pattern morphology in Palestinian and Egyptian heritage speakers. *Linguistic Approaches to Bilingualism* 4, 89–123.

Bianchi, J. (2013). Gender in Italian-German Bilinguals: A comparison with German L2 Learners of Italian. *Bilingualism: Language and Cognition*, *16*(3), 538–557.

Bice, K., & Kroll, J. F. (2015). Native language change during early stages of second language learning. *NeuroReport*, *26*, 966-971. https://doi.org/10.1097/WNR.0000000000000453

Bice, K., & Kroll, J. F. (2021). Grammatical processing in two languages: How individual differences in language experience and cognitive abilities shape comprehension in heritage bilinguals. *Journal of Neurolinguistics*, *58*. https://doi.org/10.1016/j.jneuroling.2020.100963

Birdsong, D. (1992). Ultimate attainment in second language acquisition. *Language, 68*(4), 706–755. https://doi.org/10.1353/lan.1992.0035

Birdsong, D. (2016). Dominance in bilingualism: Foundations of measurement, with insights from the study of handedness. In C. Silva-Corvalán & J. Treffers-Daller (Eds.), *Language dominance in bilinguals: Issues of measurement and operationalization* (pp. 85–105). Cambridge University Press. https://doi.org/10.1017/CBO9781107375345.005

Birdsong, D., Gertken, L. M., & Amengual, M. (2012). Bilingual language profile: An easy-to-use instrument to assess bilingualism. *COERLL, University of Texas at Austin*.

Bley-Vroman, R., & Chaudron, C. (1994). Elicited imitation as a measure of second-language competence. In E. Tarone, S. Gass, & A. Cohen (Eds.), Research methodology in second-language acquisition (pp. 245–261). Hillsdale, NJ: Erlbaum.

Bock, J. K. (1986). Meaning, sound, and syntax: Lexical priming in sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*, 575-586.

Bock, J. K., & Loebell, H. (1990). Framing Sentences. *Cognition*, *35*,1-39.

Bock, J. K., Loebell, H., & Morey, R. (1992). From conceptual roles to structural relations: Bridging the syntactic cleft. *Psychological Review*, *99*, 150-171.

Bonfiglio, T. P. (2010). *Mother tongues and nations: The invention of the native speaker.* Berlin: Walter de Gruyter.

Botezatu, M. R., Kroll, J. F., Trachsel, M. I., & Guo, T. (2022). Second language immersion impacts native language lexical production and comprehension. *Linguistic Approaches to Bilingualism*, *12*(3), 347–376. https://doi.org/10.1075/lab.19059.bot

Boudewyn, M. A., Long, D. L., & Swaab, T. Y. (2015). Graded expectations: Predictive processing and the adjustment of expectations during spoken language comprehension. *Cognitive, Affective and Behavioral Neuroscience*, *15*(3), 607–624. https://doi.org/10.3758/s13415-015-0340-0

Boumans, L. 2006. The attributive possessive in Moroccan Arabic spoken by young bilinguals in the Netherlands and their peers in Morocco. *Bilingualism: Language and Cognition*, *9*, 213–31.

Braver, T. S. (2012). The variable nature of cognitive control: A dual-mechanisms framework. *Trends in Cognitive Sciences*, *16*(2), 106. https://doi.org/10.1016/J.TICS.2011.12.010

Braver, T. S., Barch, D. M., Keys, B. A., Carter, C. S., Cohen, J. D., Kaye, J. A., Janowsky, J. S., Taylor, S. F., Yesavage, J. A., Mumenthaler, M. S., Jagust, W. J., & Reed, B. R. (2001). Context processing in older adults: Evidence for a theory relating cognitive control to neurobiology in healthy aging. *Journal of Experimental Psychology: General, 130*(4), 746–763. https://doi.org/10.1037/0096-3445.130.4.746

Brice, H., Frost, S. J., Bick, A. S., Molfese, P. J., Rueckl, J. G., Pugh, K. R., & Frost, R. (2021). Tracking second language immersion across time: Evidence from a bi-directional longitudinal cross-linguistic fMRI study. *Neuropsychologia*, *154*. https://doi.org/10.1016/j.neuropsychologia.2021.107796

Bybee, J. (1985). Morphology: A Study of the Relation between Meaning and Form. *Typological Studies in Language, 99*. Amsterdam: John Benjamins.

Caffarra, S., & Barber, H. A. (2015). Does the ending matter? The role of gender-to-ending consistency in sentence reading. *Brain Research*, *1605*(1), 83–92. https://doi.org/10.1016/j.brainres.2015.02.018

Caffarra, S., Molinaro, N., Davidson, D., & Carreiras, M. (2015). Second language syntactic processing revealed through event-related potentials: An empirical review. *Neuroscience & Biobehavioral Reviews*, *51*, 1–17. https://doi.org/10.1016/j.neubiorev.2015.01.010

Caldwell-Harris, C. L., & MacWhinney, B. (2023). Age effects in second language acquisition: Expanding the emergentist account. *Brain and Language*, *241*. https://doi.org/10.1016/j.bandl.2023.105269

Carminati, M.N. (2005) Processing reflexes of hierarchy (Person > Number > Gender) and implications for linguistic theory. *Lingua*, *115*(3), 259-285.

Castilla-Earls, A., Ronderos, J., & Fitton, L. (2023). Spanish bilingual morphosyntactic development in bilingual children with and without developmental language disorder: Articles, clitics, verbs, and the subjunctive mood. *Journal of Speech, Language, and Hearing Research*, *66*(12), 4678–4698. https://doi.org/10.1044/2023_JSLHR-23-00091

Cervone, C., Lewendon, J., & Maass, A. (2025). The Gendered Language (R)Evolution: Introduction to the Special Topic. *Social Psychological Bulletin*, *20*. https://doi.org/10.32872/spb.16491

Chang, C. B. (2016). Bilingual perceptual benefits of experience with a heritage language. *Bilingualism: Language and Cognition*, *19*(4), 791–809. https://doi.org/10.1017/S1366728914000261

Cheng, L. S. P., Burgess, D., Vernooij, N., Solís-Barroso, C., McDermott, A., & Namboodiripad, S. (2021). The Problematic Concept of Native Speaker in Psycholinguistics: Replacing Vague and Harmful Terminology With Inclusive and Accurate Measures. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.715843

Chipere, N. (2001). Native speaker variations in syntactic competence: Implications for first language teaching. *Language Awareness*, *10*, p. 107–124.

Chomsky, N. (1965). *Aspects of the theory of syntax.* Cambridge, MA: MIT Press.

Conway, A., Jarrold, C., Kane, M., Miyake, A., & Towse, J. (Eds.). (2007). *Variation in working memory*. Oxford University Press.

Coseriu, E. (1981). Los conceptos de «dialecto», «nivel» y «estilo de lengua» y el sentido propio de la dialectología. *Lingüística Española Actual*, *2*, 1–32.

Craig, C. (1977). *The structure of Jaceltec*. Austin/London: Texas Press.

Cuartero, M., Rossi, E., Navarro, E., & Pascual y Cabo, D. (2023). Mind the net! Unpacking the contributions of social network science for heritage and Bilingualism research. *Research Methods in Applied Linguistics*, *2*(1). https://doi.org/10.1016/j.rmal.2023.100041

Dąbrowska, E. (1997). The LAD goes to school: A cautionary tale for nativists. *Linguistics*, *35*, p. 735–766.

Dąbrowska, E. (2008a). The later development of an early-emerging system: The curious case of the Polish genitive. *Linguistics*, *46*, p. 629–650.

Dąbrowska, E. (2008b). The effects of frequency and neighbourhood density on adult speakers' productivity with Polish case inflections: An empirical test of usage-based approaches to morphology. *Journal of Memory and Language*, *58*, p. 931–951.

Dąbrowska, E. (2012). Different speakers, different grammars. *Linguistic Approaches to Bilingualism*, *2*(3), 219–253. https://doi.org/10.1075/lab.2.3.01dab

Dąbrowska, E., & Street, J. (2006). Individual differences in language attainment: Comprehension of passive sentences by native and non-native English speakers. *Language Sciences*, *28*, p. 604–615.

Dahl, Ö. (2000a). Egophoricity in discourse and syntax. *Functions of Language, 7*, 33–77.

Dahl, Ö. (2000b). Animacy and the notion of semantic gender. In B. Unterbeck, M. Rissanen, T. Nevalainen, & M. Saari (Eds.), *Gender in grammar and cognition: Vol. 1. Approaches to gender* (pp. 99–115). Mouton de Gruyter.

Dahl, Ö. (2008). Animacy and egophoricity: Grammar, ontology and phylogeny. *Lingua*, *118*(2), 141–150. https://doi.org/10.1016/j.lingua.2007.02.008

Davies, M. (1995). Analyzing syntactic variation with computer-based corpora. *Hispania*, *78*, 360–403. https://doi.org/10.2307/345438

De la Garza Espinosa, V. P. (2022). *Gender congruency effect across the animacy hierarchy and its neural correlates* [Master's thesis, University of Padua & University of Granada]. University of Padua Thesis Repository. https://hdl.handle.net/20.500.12608/42462

Dell, G. S., & Chang, F. (2014). The p-chain: Relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1634). https://doi.org/10.1098/rstb.2012.0394

De Vincenzi, M., Job, R., Di Matteo, R., Angrilli, A., Penolazzi, B., Ciccarelli, L., & Vespignani, F. (2003). Differences in the perception and time course of syntactic and semantic violations. *Brain and language*, *85*(2), 280-296.

Díaz Colmenares, Y. (2023). Estimad@s amigxs: Una mirada a los usos inclusivos en Twitter de Venezuela [Dear$_{\text{inc@}}$ friends$_{\text{inclX}}$: A look at inclusive uses on Twitter in Venezuela]. In E. Pato (Ed.), *Estudios sobre el español de Venezuela* (pp. 373-403). Vervuert Iberoamericana. https://doi.org/10.31819/9783968694542

Dijkstra, T., Van Jaarsveld, H., & Ten Brinke, S. (1998). Interlingual homograph recognition: Effects of task demands and language intermixing. *Bilingualism: Language and Cognition*, *1*(1), 51-66.

Dixon, R. M. W. (1979). Ergativity. *Language*, *55*(1), 59-138.

Domínguez, L. (2003). Interpreting reference in the early acquisition of Spanish clitics. In S. Montrul and F. Ordóñez, Francisco (eds.), Linguistic Theory and Language Development in Hispanic Languages: Papers from the 5th Hispanic Linguistics Symposium and the 4th Conference on the Acquisition of Spanish and Portuguese. Somerville, USA, Cascadilla Press, 212-228.

Donchin, E. Ritter, W., & McCallum, W.C. (1978). Cognitive psychophysiology: The endogenous components of the ERP. In E. Callaway, P. Tueting, & S.H. Koslow (Eds.), *Event-Related Brain Potentials in Man* (pp. 349-441). New York: Academic Press.

Duchon, A., Perea, M., Sebastián-Gallés, N., Martí, A., & Carreiras, M. (2013). EsPal: One-stop shopping for Spanish word properties. *Behavior research methods*, *45*(4), 1246-1258.

Durand-López, E. M., & Dickinson, K. V. (2025). Gender-inclusive morphology in Spanish: learnability, processing costs, and use. *Studies in Hispanic and Lusophone Linguistics*, *18*(2), 391–424. https://doi.org/10.1515/shll-2025-2013

Dussias, P. E., & Sagarra, N. (2007). The effect of exposure on syntactic parsing in Spanish–English bilinguals. *Bilingualism: Language and cognition*, *10*(1), 101-116.

Eckert, P. (1989). The whole woman: Sex and gender differences in variation. *Language Variation and Change*, *1*(3), 245–267.

Feleki, E., & Branigan, H. (1997). *Animacy effects in Greek sentence production: Evidence for single stage syntactic processing?* Poster presented at the Third Conference of Architectures and Mechanisms of Language Processing (AMLaP-97), Edinburgh, UK.

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology*, *47*, p. 164–203.

Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language*, *25*, 348-368.

Fiebach, C. J., Schlesewsky, M., & Friederici, A. D. (2002). Separating syntactic memory costs and syntactic integration costs during parsing: The processing of German WH-questions. *Journal of Memory and Language*, *47*(2), 250-272.

Foote, R. (2011). Integrated knowledge of agreement in early and late English-Spanish bilinguals. *Applied Psycholinguistics*, *32*(1), 187–220. https://doi.org/10.1017/S0142716410000342

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, *14*, 178-210.

Fuchs, Z. (2021). Facilitative use of grammatical gender in Heritage Spanish. *Linguistic Approaches to Bilingualism*, *12*(6), 845–871. https://doi.org/10.1075/lab.20024.fuc

Fuchs, Z., Eldem-Tunç, E., Pham, L., Mermelstein, L., & Runova, A. (2026). When gender meets number: facilitative processing of one vs. two features on Spanish definite articles. *Frontiers in Language Sciences*, *5*, https://doi.org/10.3389/flang.2026.1760372

Fuchs, Z., Polinsky, M., & Scontras, G. (2015). The differential representation of number and gender in Spanish. The linguistic review, 32(4), 703-737.

Gaillard, S., & Tremblay, A. (2016). Linguistic Proficiency Assessment in Second Language Acquisition Research: The Elicited Imitation Task. *Language Learning*, *66*(2), 419–447. https://doi.org/10.1111/lang.12157

Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences*, *102*, 74–78. https://doi.org/10.1016/j.paid.2016.06.069

Gil, J. M. (2020). Las paradojas excluyentes del" lenguaje inclusivo": sobre el uso planificado del morfema flexivo-e. *Revista Española De Lingüística*, 65–84. https://doi.org/10.31810/RSEL.50.1

González Melón, E., Buyse, K., & Álvarez Mella, H. (2026). Multilingual Spanish heritage speakers in Europe: sociolinguist dynamics and educational perspectives. *Journal of Spanish Language Teaching*. https://doi.org/10.1080/23247797.2026.2677952

González Rivera, M. (2025). Perspectivas sobre el sexismo lingüístico y el lenguaje inclusivo en el contexto puertorriqueño. *Anuario de Letras. Lingüística y Filología*, *13*(2), 189–250. https://doi.org/10.19130/iifl.adel.2025.13.2/01x225s4w96

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition*, *1*(2), 67-81.

Greenberg, J. (1963). *Universals of language*. Cambridge: MIT Press.

Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology*, *25*(5), 515–530. https://doi.org/10.1080/20445911.2013.796377

Grosjean, F. (2024). The Statistics of Bilingualism. In *On Bilinguals and Bilingualism* (pp. 138–147). Cambridge University Press. https://doi.org/10.1017/9781009210409.010

Guasti, M. T. (1993). Verb syntax in Italian child grammar: Finite and nonfinite verbs. *Language acquisition*, *3*(1), 1-40.

Gullifer, J. W., Chai, X. J., Whitford, V., Pivneva, I., Baum, S., Klein, D., & Titone, D. (2018). Bilingual experience and resting-state brain connectivity: Impacts of L2 age of acquisition and social diversity of language use on control networks. *Neuropsychologia*, *117*, 123-134. https://doi.org/10.1016/j.neuropsychologia.2018.04.037

Gullifer, J. W., & Titone, D. (2020). Characterizing the social diversity of bilingualism using language entropy. *Bilingualism: Language and Cognition*, *23*, 283-294. https://doi.org/10.1017/S1366728919000026

Hale, K. (1972). A note on subject-object inversion in Navajo. In B. Kachru, R. Lees, Y. Malkiel, A. Pietrangeli, & S. Saporta (Eds.), *Issues in linguistics: Papers in honor of Henry and Renée Kahane* (pp. 300-309). Urbana: University of Illinois Press.

Hao, J., Rossi, E., Nakamura, M., Luque, A., & Rothman, J. (2025). Individual differences matter in heritage language bilingual processing. *Studies in Second Language Acquisition*, 1–20. https://doi.org/10.1017/S0272263125101149

Hartshorne, J. K., Tenenbaum, J. B., & Pinker, S. (2018). A critical period for second language acquisition: Evidence from 2/3 million English speakers. *Cognition*, *177*, 263–277. https://doi.org/10.1016/J.COGNITION.2018.04.007

Hellinger, M., & Bußmann, H. (Eds.). (2001). *Gender across languages: The linguistic representation of women and men*, *Vol.* 1. John Benjamins Publishing Company. https://doi.org/10.1075/impact.9

Hernandez, A. E., Bodet, J. P., Gehm, K., & Shen, S. (2021). What does a critical period for second language acquisition mean?: Reflections on Hartshorne et al. (2018). *Cognition*, *206*, 104478. https://doi.org/10.1016/J.COGNITION.2020.104478

Hernandez Santacruz, J., Torres, J. and Kroll, J.F. (2026). Heritage speakers reveal the dynamics of bilingual language regulation. *Bilingualism: Language and Cognition*, 29,78–91. https://doi.org/ 10.1017/S136672892500020

Hockett, C. F. (1958). A course in modern linguistics. (New York, NY: Macmillan).

Jegerski, J., & Fernández Cuenca, S. (2025). Variability in the Online Processing of Subject–Verb Number Agreement in Spanish as a Heritage Language: The Role of Lexical Frequency. *Languages*, *10*(9). https://doi.org/10.3390/languages10090211

Jegerski, J., & Sekerina, I. A. (2021). The Psycholinguistics of Heritage Languages. In *The Cambridge Handbook of Heritage Languages and Linguistics* (pp. 449–470). Cambridge University Press. https://doi.org/10.1017/9781108766340.021

Jegerski, J., & VanPatten, B. (2013). Self-paced reading. In *Research methods in second language psycholinguistics* (pp. 20–49). Taylor & Francis Group. http://ebookcentral.proquest.com/lib/uci/detail.action?docID=1582610

Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes*, *15*(2), 159–201. https://doi.org/10.1080/016909600386084

Kaiser, G. A. (1994). More about INFL-ection and agreement: The acquisition of clitic pronouns in French. In J. Meisel (Ed.), *Bilingual first language acquisition: French and German grammatical development* (pp. 131–160). John Benjamins.

Kalinowski, S. (2020). Lenguaje inclusivo en usuarios de Twitter en Argentina: un estudio de corpus. *Cuarenta Naipes Revista de Cultura y Literatura*, *3*, 233-259. http://fh.mdp.edu.ar/revistas/index.php/cuarentanaipes/article/view/4888

Keating, G. D. (2022). The Effect of Age of Onset of Bilingualism on Gender Agreement Processing in Spanish as a Heritage Language. *Language Learning*, *72*(4), 1170–1208. https://doi.org/10.1111/lang.12510

Keating, G. D., & Jegerski, J. (2015). Experimental designs in sentence processing research. *Studies in Second Language Acquisition*, *37*(1), 1–32. https://doi.org/10.1017/S0272263114000187

Kempen, G., & Harbusch, K. (2004). A corpus study into word order variation in German subordinate clauses: Animacy affects linearization independently of grammatical function assignment. In T. Pechmann & C. Habel (Eds.), *Multidisciplinary approaches to language production* (pp. 173-181). Berlin: Mouton de Gruyter.

Kim, J. Y. (2020). Discrepancy between heritage speakers' use of suprasegmental cues in the perception and production of Spanish lexical stress. *Bilingualism*, *23*(2), 233–250. https://doi.org/10.1017/S1366728918001220

Kim, A. E., Oines, L., & Miyake, A. (2018). Individual Differences in Verbal Working Memory Underlie a Tradeoff Between Semantic and Structural Processing Difficulty During Language Comprehension: An ERP Investigation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(3), 406–420. https://doi.org/https://doi.org/10.1037/xlm0000457

Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language*, *52*, 205-225.

Kluender, R., & Kutas, M. (1993). Bridging the gap: Evidence from ERPs on the processing of unbounded dependencies. *Journal of cognitive neuroscience*, *5*(2), 196-214.

Körner, A., Abraham, B., Rummer, R., & Strack, F. (2022). Gender representations elicited by the gender star form. *Journal of Language and Social Psychology, 41*(5), 553–571. https://doi.org/10.1177/0261927X221080181

Kostromitina, M., & Plonsky, L. (2022). Elicited imitation tasks as a measure of L2 proficiency: A meta-analysis. *Studies in Second Language Acquisition*, *44*(3), 886-911.

Kroll, J. F., & Dussias, P. E. (2024). Beyond Transfer: Language Processing in Bilinguals is Shaped by Competition and Regulation. *Language Teaching Research Quarterly*, *44*, 55–70. https://doi.org/10.32038/ltrq.2024.44.07

Kroll, J. F., Dussias, P. E., Bice, K., & Perrotti, L. (2015). Bilingualism, Mind, and Brain. *Annual Review of Linguistics*, *1*(1), 377–394. https://doi.org/10.1146/annurev-linguist-030514-124937

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connection between bilingual memory representations. *Journal of Memory and Language*, *33*, 149-174. https://doi.org/ 10.1006/jmla.1994.1008

Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research (Special Issue)*, *1146*,23-49.

Kuperberg, G. R., Sitnikova, T., Caplan, D., & Holcomb, P. J. (2003). Electrophysiological distinctions in processing conceptual relationships within simple sentences. *Cognitive Brain Research*, *17*, 117-129.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*(August), 621–647. https://doi.org/10.1146/annurev.psych.093008.131123

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13). https://doi.org/10.18637/jss.v082.i13

Lew-Williams, C. (2017). Specific Referential Contexts Shape Efficiency in Second Language Processing: Three Eye-Tracking Experiments with 6- and 10-Year-Old Children in Spanish Immersion Schools. *Annual Review of Applied Linguistics*, *37*, 128–147. https://doi.org/10.1017/S0267190517000101

Lew-Williams, C., & Fernald, A. (2007). Young Children Learning Spanish Make Rapid Use of Grammatical Gender in Spoken Word Recognition. *Psychological Science*, *18*(3), 193–198. https://doi.org/10.1111/J.1467-9280.2007.01871.X

Lew-Williams, C., & Fernald, A. (2010). Real-time processing of gender-marked articles by native and non-native Spanish speakers. *Journal of Memory and Language*, *63*(4), 447–464. https://doi.org/10.1016/j.jml.2010.07.003

Libben, M. R., & Titone, D. A. (2009). Bilingual lexical access in context: evidence from eye movements during reading. *Journal of Experimental Psychology: Learning, memory, and cognition*, *35*(2), 381.

Linck, J. A., Kroll, J. F., & Sunderman, G. (2009). Losing access to the native language while immersed in a second language: evidence for the role of inhibition in second-language learning. *Psychological Science*, *20*, 1507-1515. https://doi.org/ 10.1111/j.1467-9280.2009.02480.x

Luck, S. J. (2014). *An introduction to the event-related potential technique*. MIT press.

Luque, A., Rossi, E., Kubota, M., Nakamura, M., Rosales, C., López-Rojas, C., Rodina, Y., & Rothman, J. (2023). Morphological transparency and markedness matter in heritage speaker gender processing: an EEG study. *Frontiers in Psychology*, *14*. https://doi.org/10.3389/fpsyg.2023.1114464

MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00226

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, *101*, 676-703.

McDonald, J. L., Bock, K., & Kelly, M. (1993). Word and word order: Semantic, phonological and metrical determinants of serial position. *Cognitive Psychology*, *25*, 188-230.

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing Language Profiles in Bilinguals and Multilinguals. *Journal of Speech, Language, and Hearing Research*, *50*(4), 940–967. https://doi.org/10.1044/1092-4388(2007/067)

Marian, V., & Spivey, M. (2003). Competing activation in bilingual language processing: Within-and between-language competition. *Bilingualism: Language and cognition*, *6*(2), 97-115.

Martohardjono, G., Phillips, I., Madsen, C. N., and Schwartz, R. (2017). "Cross-linguistic influence in bilingual processing: an ERP study," in *Proceedings of the 41st Annual BOSTON University Conference on Language Development*. eds. M. LaMendola and J. Scott (Somerville, MA: Cascadilla Press), 452–465.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. Journal of Memory and Language, 94, 305–315. https://doi.org/10.1016/j.jml.2017.01.001

Meuter, R. F., & Allport, A. (1999). Bilingual language switching in naming: Asymmetrical costs of language selection. *Journal of memory and language*, *40*(1), 25-40.

Ming, T., and Tao, H. 2008. Developing a Chinese heritage language corpus: issues and a preliminary report. In A. W. He and Yun Xiao (eds.), *Chinese as a Heritage Language: Fostering Rooted World Citizenry*, 167–88. Honolulu: University of Hawaii Mānoa.

Minkoff, S. (2000). Animacy hierarchies and sentence processing. In A. Carnie & E. Guilfoyle (Eds.), *The syntax of verb initial languages* (pp. 201 212). Oxford, UK: Oxford University Press.

Misra, M., Guo, T., Bobb, S. C., & Kroll, J. F. (2012). When bilinguals choose a single word to speak: Electrophysiological evidence for inhibition of the native language. *Journal of Memory and Language*, *67*, 224-237.

Molinaro, N., Barber, H. A., & Carreiras, M. (2011). Grammatical agreement processing in reading: ERP findings and future directions. *Cortex*, *47*(8), 908–930. https://doi.org/10.1016/j.cortex.2011.02.019

Montrul, S. (2010). How similar are L2 learners and heritage speakers? Spanish clitics and word order. Applied Psycholinguistics, 31, 167–207. https://doi.org/10.1017/S014271640999021

Montrul, S. (2022). *Native Speakers, Interrupted: Differential Object Marking and Language Change in Heritage Languages*. Cambridge: Cambridge University Press.

Montrul, S., & Bateman, N. (2020a). Vulnerability and stability of Differential Object Marking in Romanian heritage speakers. *Glossa: A Journal of General Linguistics*, 5(1): 119.

Montrul, S., & Bateman, N. (2020b). Differential Object Marking in Romanian as a heritage language. In A. Mardale & S. Montrul (eds.), *The acquisition of Differential Object Marking,* pp. 261–282. Amsterdam: John Benjamins.

Montrul, S., Bhatt, R., & Girju, R. (2015). Differential Object Marking in Spanish, Hindi, and Romanian as heritage languages. *Language*, *91*(3), 564–610.

Montrul, S., & Sánchez-Walker, N. (2013). Differential Object Marking in child and adult Spanish heritage speakers. *Language Acquisition*, *20*(2), 109–132.

Morales-Gálvez, S., & Pérez-Lozano, L. (2024). Contra la presunción monolingüe: Una defensa normativa del multilingüismo en el caso español. Revista Española de Ciencia Política, 66, 69-90. https://doi.org/10.21308/recp.66.03

Muelas-Gil, M. (2026). Gender-inclusive language in Spanish: Where it is now and where it is headed. In F. Pfalzgraf (Ed.), *Gender-inclusive language: Findings from 14 languages and open research questions* (pp. 285–314). De Gruyter Mouton. https://doi.org/10.1515/9783111701608-012

Münte, T. F., Matzke, M., & Johannes, S. (1997). Brain activity associated with syntactic incongruencies in words and pseudo-words. *Journal of cognitive neuroscience*, *9*(3), 318-329.

Nakano, H., Saron, C., & Swaab, T. Y. (2010). Speech and span: Working memory capacity impacts the use of animacy but not ofworld knowledge during spoken sentence comprehension. *Journal of Cognitive Neuroscience*, *22*, 2886-2898.

Núñez-Román, F., Gómez-Camacho, A., & Hunt-Gómez, C. I. (2023). Gender-Inclusive Textisms: How Spanish-speaking educational communities promote linguistic innovations on twitter. *Linguistics and Education*, *76*. https://doi.org/10.1016/j.linged.2023.101191

O'Grady, W., Kwak, H. Y., Lee, O. S., & Lee, M. (2011). An emergentist perspective on heritage language acquisition. *Studies in Second Language Acquisition*, *33*(2), 223-245.

Olson, D. J. (2023). Measuring bilingual language dominance: An examination of the reliability of the Bilingual Language Profile. *Language Testing*. https://doi.org/10.1177/02655322221139162

Ophir, E., Nass, C., & Wagner, A. D. (2009). Cognitive control in media multitaskers. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(37), 15583–15587. https://doi.org/10.1073/PNAS.0903620106

Ortega, L., Iwashita, N., Norris, J. M., & Rabie, S. (2002). An investigation of elicited imitation tasks in crosslinguistic SLA research [Paper presentation]. *The 20th Second Language Research Forum*, Toronto, Ontario, Canada.

Osterhout, L., Bersick, M., & Mclaughlin, J. (1997). Brain potentials reflect violations of gender stereotypes. *Memory & Cognition*, *25*(3), 273–285.

Øvrelid, L., (2004). Disambiguation of grammatical functions in Norwegian: Modeling variation in word order interpretations conditioned by animacy and definiteness. In F. Karlsson (Ed.), *Proceedings of the 20th Scandinavian Conference of Linguistics, Helsinki, January 7-9, 2004.* Department of General Linguistics, University of Helsinki.

Paczynski, M., & Kuperberg, G. R. (2011). Electrophysiological evidence for use of the animacy hierarchy, but not thematic role assignment, during verb-argument processing. *Language* *and Cognitive Processes*, *26*(9), 1402–1456. https://doi.org/10.1080/01690965.2011.580143

Pakulak, E., & Neville, H. J. (2010). Proficiency Differences in Syntactic Processing of Monolingual Native Speakers Indexed by Event-related Potentials. *Journal of Cognitive Neuroscience*, *22*(12), 2728–2744. https://doi.org/10.1162/jocn.2009.21393

Papadopoulos, B. (2022). A brief history of gender-inclusive Spanish. *Deportate, esuli, profughe*, *48*(1), 31-48.

Parafita Couto, M.C., & Gullberg, M. (2017). Code-switching within the noun phrase: Evidence from three corpora. *International Journal of Bilingualism*, *23*, 695-714. https://doi.org/10.1177/1367006917729543

Pérez-Leroux, A. T., Cuza, A., & Thomas, D. (2011). Clitic placement in Spanish-English bilingual children. *Bilingualism: Language and Cognition*, *14*, 221–232. https://doi.org/10.1017/S1366728910000234

Phillips, C., Kazanina, N., & Abada, S. H. (2005). ERP effects of the processing of syntactic long-distance dependencies. *Cognitive Brain Research*, *22*(3), 407–428. https://doi.org/10.1016/j.cogbrainres.2004.09.012

Polinsky, M. (2018). *Heritage languages and their speakers*. Cambridge: Cambridge University Press.

Popič, D., & Gorjanc, V. (2018). Challenges of adopting gender-inclusive language in Slovene. *Suvremena lingvistika*, *44*(86), 329–350. https://doi.org/10.22210/suvlin.2018.086.07

Pozniak, C., Corbeau, E., & Burnett, H. (2024). Contextual dilution in French gender inclusive writing: An experimental investigation. *Journal of French Language Studies*. https://doi.org/10.1017/S0959269523000236

Prior, A., Degani, T., Awawdy, S., Yassin, R., & Korem, N. (2017). Is susceptibility to cross-language interference domain specific?. *Cognition*, *165*, 10-25.

Quine, W. V. (1948). On what there is. *The Review of Metaphysics, 2*, 21–38.

R Core Team. (2026). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Rayner, K., Carlson, M., & Frazier, L. (1983). The interaction of syntax and semantics during sentence processing: Eye movements in the analysis of semantically biased sentences. *Journal of Verbal Learning and Verbal Behavior*, *22*, 358-374.

Requena, P., & Dracos, M. (2018). Impermeability of L1 syntax: Spanish variable clitic placement in bilingual children. In A. B. Bertolini & M. J. Kaplan (Eds.), *Proceedings of the 42nd annual Boston University Conference on Language Development*, 2 (pp. 644–658). Cascadilla Press.

Ripley, B., Venables, B., Bates, D. M., Hornik, K., Gebhardt, A., Firth, D., & Ripley, M. B. (2013). Package 'mass'. *Cran r*, *538*(113-120), 822.

Román Irizarry, A. (2021). ¿Todxs, todes, o todos? Actitudes de la juventud puertorriqueña hacia los morfemas de lenguaje inclusivo. In B. Llenín Figueroa (Ed.), *Proceedings of the VIII Coloquio ¿Del otro la'o?: Arte y activismo cuir en el Puerto Rico contemporáneo* (pp. 282 – 310). Editora Educación Emergente. https://www.researchgate.net/publication/361980392_Todxs_todes_o_todos_Actitudes

Román Irizarry, A., Beatty-Martínez, A. L., Torres, J., & Kroll, J. F. (2025). "Todes" and "Todxs", linguistic innovations or grammatical gender violations? *Cognition*, *257*, 1–25. https://doi.org/10.1016/j.cognition.2025.106061

Román Irizarry, A., & Guzzardo Tamargo, R. E. (2025). Processing of pronouns with gender-inclusive –x in Spanish: An eye-tracking study. *Social Psychological Bulletin*, *20*. https://doi.org/10.32872/spb.13611

Rosa, J., & Flores, N. (2017). Unsettling race and language: Toward a raciolinguistic perspective. Language in Society, 46(5), 621–647.

Rossi, E., Kroll, J. F., & Dussias, P. E. (2014). Clitic pronouns reveal the time course of processing gender and number in a second language. *Neuropsychologia*, *62*(1), 11–25. https://doi.org/10.1016/j.neuropsychologia.2014.07.002

Rothman, J., & Treffers-Daller, J. (2014). A prolegomenon to the construct of the native speaker: Heritage speaker Bilinguals are Natives Too! *Applied Linguistics*, *35*(1), 93–98. https://doi.org/10.1093/applin/amt049

Rothman, J., Bayram, F., Deluca, V., di Pisa, G., Duñabeitia, J. A., Gharibi, K., Hao, J., Kolb, N., Kubota, M., Kupisch, T., Laméris, T., Luque, A., van Osch, B., Pereira Soares, S. M., Prystauka, Y., Tat, D., Tomić, A., Voits, T., & Wulff, S. (2023). Monolingual comparative normativity in bilingualism research is out of control: Arguments and alternatives. *Applied Psycholinguistics*, *44*(3), 316–329. https://doi.org/10.1017/S0142716422000315

Ryan, N. C., Stoudt, S., & Vecchione, F. (2026). Spatial and temporal trends in the use of gender inclusive language: a study of Spanish-language Twitter and news media. *Digital Scholarship in the Humanities*, *41*(2), 950–968. https://doi.org/10.1093/llc/fqaf156

Sá-Leite, A. R., Haro, J., Comesaña, M., & Fraga, I. (2021). Of Beavers and Tables: The Role of Animacy in the Processing of Grammatical Gender Within a Picture-Word Interference Task. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.661175

Salinas, C., Jr., & Lozano, A. (2021). The history and evolution of the term Latinx. In E. G. Murillo, D. Delgado Bernal, S. Morales, L. Urrieta, E. Ruiz Bybee, J. Sánchez Muñoz, V.B. Saenz, D. Villanueva, M. Machado-Casas, & K. Espinoza (Eds.), *Handbook of Latinos and education* (2nd ed., pp. 249–263). Routledge. https://doi.org/10.4324/9780429292026

Samples, C. E. (2024). Gender-inclusive Language in Twitter. *Hispania*, *107*(1), 139–160. https://doi.org/10.1353/hpn.2024.a921466

Sagarra, N., & Rodriguez, N. (2022). Subject–Verb Number Agreement in Bilingual Processing: (Lack of) Age of Acquisition and Proficiency Effects. *Languages*, *7*(1). https://doi.org/10.3390/languages7010015

Sánchez–Alonso, S., Martinez–Ferreriro, S., & Bastiaanse, R. (2011, October). Clitics in Spanish agrammatic aphasia: A study of the production of unaccusative, reflexive and object clitics. In *Discourse Anaphora and Anaphor Resolution Colloquium* (pp. 184-197). Berlin, Heidelberg: Springer Berlin Heidelberg.

Satterfield, T. (2021). Dual competence in dual language learners. In T. Gupton & E. Gielau (Eds.), East and west of the Pentacrest: Linguistic studies in honor of Paula Kempchinsky (pp. 189–214). John Benjamins Publishing Company. https://doi.org/10.1075/ihll.33.09sat

Schad, D. J., Vasishth, S., Hohenstein, S., & Kliegl, R. (2020). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language*, *110*. https://doi.org/10.1016/j.jml.2019.104038

Schudson, Z. C., & van Anders, S. M. (2021). Gender/sex diversity beliefs: Scale construction, validation, and links to prejudice. *Group Processes and Intergroup Relations*, *00*(0), 1–26. https://doi.org/10.1177/1368430220987595

Schwartz, A. I., & Kroll, J. F. (2006). Bilingual lexical activation in sentence context. *Journal of Memory and Language*, *55*(2), 197-212. https://doi.org/10.1016/j.jml.2006.03.004

Schwartz, A. I., Kroll, J. F., & Diaz, M. (2007). Reading words in Spanish and English: Mapping orthography to phonology in two languages. *Language and Cognitive processes*, *22*(1), 106-129.

Scontras, G., Polinsky, M., & Fuchs, Z. (2018). In support of representational economy: Agreement in heritage Spanish. *Glossa*, *3*(1). https://doi.org/10.5334/gjgl.164

Sczesny, S., Formanowicz, M., & Moser, F. (2016). Can Gender-Fair Language Reduce Gender Stereotyping and Discrimination? In *Frontiers in Psychology* (Vol. 7). Frontiers Media S.A. https://doi.org/10.3389/fpsyg.2016.00025

Shin, N., Requena, P., & Kemp, A. (2017). Bilingual and monolingual children's patterns of syntactic variation: Variable clitic placement in Spanish. In A. Auza Benavides & R. G. Schwartz (Eds.), Language disorders and development in bilingual children (pp. 63–88). Springer.

Silva-Corvalán, C. (1994). *Language Contact and Change: Spanish in Los Angeles*. Oxford: Oxford University Press.

Silverstein, M., (1976). Hierarchy of features and ergativity. In: Dixon, R.M.W. (Ed.), *Grammatical Categories in Australian Languages*. Australian Institute of Aboriginal Studies, Canberra, pp. 112–171.

Sinnott, S. & Smith, E. (2007). ¿Subir o no subir? A look at clitic climbing in Spanish (Unpublished manuscript). The Ohio State University and Pikeville College.

Slemp, K. (2020). *Latino, Latina, Latin@, Latine, and Latinx: Gender inclusive oral expression in Spanish* [Master's thesis, The University of Western Ontario].

Slemp, K., Black, M., & Cortiana, G. (2020). Reactions to Gender-Inclusive Language in Spanish on Twitter and You-Tube. *Actes du Congrès Annuel de l 'Association Canadienne de Lingüistique*, 1-13.

Slemp, K., Heap, D., & Díaz, Y. (2019). Writing vs. Speech: How do Spanish online videos pronounce gender inclusive language?. *Actes du Congrès Annuel de l'Association Canadienne de Lingüistique*, 1-14.

Snider, N., & Zaenen, A. (2006). Animacy and syntactic structure: Fronted NPs in English. In M. Butt, M. Dalrymple, & T. H. King (Eds.), *Intelligent linguistic architectures: Variations on themes by Ronald M. Kaplan*. Stanford: CSLI Publications.

Stetie, N. A., & Zunino, G. M. (2022). Non-binary language in Spanish? Comprehension of non-binary morphological forms: a psycholinguistic study. *Glossa: A Journal of General Linguistics*, *7*(1), 1–38. https://doi.org/10.16995/glossa.6144

Stetie, N. A., & Zunino, G. M. (2024). Do gender stereotypes bias the processing of morphological innovations? The case of gender-inclusive language in Spanish. *Psychology of Language and Communication*, *28*(1), 446–469. https://doi.org/10.58734/plc-2024-0016

Street, J., & Dąbrowska, E., 2010. More individual differences in language attainment: How much do adult native speakers of English know about passives and quantifiers? *Lingua*, 120, p. 2080–2094.

Tanaka, M., Branigan, H. P., & Pickering, M. J. (2005). *The role of animacy in Japanese sentence production.* Paper presented at eighteenth annual CUNY conference, Tucson, AZ.

Teschner, R. V., & Russell, W. M. (1984). The gender patterns of Spanish nouns: An inverse dictionary-based analysis. *Hispanic Linguistics*, *1*(1), 115–132.

Thierry, G., & Wu, Y. J. (2007). Brain potentials reveal unconscious translation during foreign language comprehension. *Proceedings of National Academy of Sciences*, *104*, 12530-12535. https://doi.org/10.1073/pnas.0609927104

Trueswell, J., Tanenhaus, M., & Garnsey, S. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language*, *33*, 285-318.

Valdés Kroff, J. R., Dussias, P. E., Gerfen, C., Perrotti, L., & Bajo, M.T. (2017). Experience with code-switching modulates the use of grammatical gender during sentence processing. *Linguistic Approaches to Bilingualism*, *7,* 163-198.

Vargas Fuentes, N. A., & Torres, J. R. (2026). La inseguridad lingüística en hablantes de herencia Explorando los factores de proficiencia y dominancia. *Estudios de Linguistica Del Espanol*, *51*(1), 120–147. https://doi.org/10.36950/elies.2026.51.1.7

Van Valin, R., & LaPolla, R. (1997). *Syntax: Structure, meaning and function*. Cambridge, UK: Cambridge University Press.

Vidal-Ortiz, S., & Martínez, J. (2018). Latinx thoughts: Latinidad with an X. *Latino Studies*, *16*, 384–395. https://doi.org/10.1057/s41276-018-0137-8

Wei, L., & Lee, S. (2001). L1 development in an L2 environment: The use of Cantonese classifiers and quantifiers by young British-born Chinese in Tyneside. *International Journal of Bilingual Education and Bilingualism*, *4*(6), 359-382.

Wechsler, S. (2009). Agreement features. *Language and linguistics compass*, *3*(1), 384-405.

Wigdorowitz, M., Pérez, A. I., & Tsimpli, I. M. (2020). A holistic measure of contextual and individual linguistic diversity. *International Journal of Multilingualism*. https://doi.org/10.1080/14790718.2020.1835921

Winter, B. (2019). *Statistics for linguists: An introduction using R*. Routledge. https://doi.org/10.4324/9781315165547

Zarwanitzer, A. (2019). *Procesamiento de oraciones con lenguaje inclusivo* [Processing sentences with inclusive language, Bachelor's thesis]. Universidad de San Andrés, Argentina. https://repositorio.udesa.edu.ar/jspui/handle/10

Zarwanitzer, A., & Gelormini-Lezama, C. (2023). Tiempos de lectura de oraciones con lenguaje inclusivo en español: un estudio psicolingüístico. *Vertex Revista Argentina de Psiquiatría*, *34*(159), 40–46. https://doi.org/10.53680/vertex.v34i159.366

Zhang, H., Diaz, M. T., Guo, T., & Kroll, J. F. (2021). Language immersion and language training: Two paths to enhanced language regulation and cognitive control. *Brain and Language*, *223*, 105043. https://doi.org/10.1016/j.bandl.2021.105043

Zirnstein, M., van Hell, J. G., & Kroll, J. F. (2018). Cognitive control ability mediates prediction costs in monolinguals and bilinguals. *Cognition*, *176*, 87–106. https://doi.org/10.1016/j.cognition.2018.03.001

Zunino, G. M., & Stetie, N. A. (2023). Procesamiento de formas no binarias en distintas lenguas: evidencias, tensiones y desafíos. *Revista Comunicación,Cultura y Política*, *14*.

APPENDIX A: Clitic Pronoun Study Supplementary Materials

Table A1: Model Fit Indices for the Clitic Pronoun Study Behavioral Data

Model	Parameters	AIC	BIC	Log likelihood	Deviance	Chi-square	df	*p* value
Accuracy								
1a. Accuracy ~ Gender*Number + (1 + Gender*Number | Subject) + (1 | Item)	15	5225.1	5323.8	-2597.6	5195.1			
1b. Accuracy ~ Gender*Number*DELE + (1 + Gender*Number | Subject) + (1 | Item)	19	5220.6	5345.6	-2591.3	5182.6	12.5	4	<.05*
1c. Accuracy ~ Gender*Number*EIT Spanish + (1 + Gender*Number | Subject) + (1 | Item)	19	5228.0	5352.9	-2595.0	5190.0	5.2	4	.27
1d. Accuracy ~ Gender*Number*LUR + (1 + Gender*Number | Subject) + (1 | Item)	19	5223.2	5348.2	-2592.6	5187.4	9.9	4	<.05*
1e. Accuracy ~ Gender*Number*SCD + (1 + Gender*Number | Subject)	18	5223.4	5341.7	-2593.7	5219.9	7.7	3	.05
1f. Accuracy ~ Gender*Number*LD + (1 + Gender*Number | Subject)	19	5225.3	5350.3	-2593.7	5187.3	7.8	4	.10
Reaction Time								
2a. log10RT ~ Gender*Number + (1 + Gender*Number | Subject) + (1 | Item)	16	1244.3	1349.5	-606.2	1212.3			
2b. log10RT ~ Gender*Number*DELE + (1 + Gender*Number | Subject) + (1 | Item)	20	1243.3	1374.8	-601.7	1203.3	9.0	4	.06
2c. log10RT ~ Gender*Number*EIT Spanish + (1 + Gender*Number | Subject) + (1 | Item)	20	1245.5	1377.1	-602.8	1205.5	6.8	4	.15
2d. log10RT ~ Gender*Number*LUR + (1 + Gender*Number | Subject)	19	1267.7	1392.7	-614.9	1229.7	0	3	1
2e. log10RT ~ Gender*Number*SCD + (1 + Gender*Number | Subject) + (1 | Item)	20	1247.9	1379.4	-604.0	1207.9	4.4	4	.35
2f. log10RT ~ Gender*Number*LD + (1 + Gender| Subject) + (1 + Gender| Number) + (1 | Item)	16	1242.4	1347.6	-605.2	1210.4	1.9	0	

Note. *p* value shows the results of the likelihood-ratio test through the ANOVA function from the lmerTest R-package (Kuznetsova et al., 2017). DELE = *Diploma de Español como Lengua Extranjera*; LUR = Language Use Ratio; SCD = Spanish – English schooling difference; LD = Language Dominance.

Table A2: Model Fit Indices for the Clitic Pronoun Study EEG Data

Model	Parameters	AIC	BIC	Log likelihood	Deviance	Chi-square	df	*p* value
350 ms – 550 ms (LAN/N400)								
3a. Voltage ~ Gender*Number*Electrode + (1 + Gender*Number | Subject)	23	1711.9	1804.8	-833.0	1665.9			
3b. Voltage ~ Gender*Number*Electrode*DELE + (1 + Gender*Number | Subject)	35	1722.0	1863.4	-826.0	1652.0	13.9	12	.30
3c. Voltage ~ Gender*Number*Electrode*EIT Spanish + (1 + Gender*Number | Subject)	35	1727.9	1869.3	-828.9	1657.9	8.1	12	.78
3d. Voltage ~ Gender*Number*Electrode*LUR + (1 + Gender*Number | Subject)	35	1706.5	1847.9	-818.3	1636.5	29.4	12	<.01**
3e. Voltage ~ Gender*Number*Electrode*SCD + (1 + Gender*Number | Subject)	35	1729.0	1870.4	-829.5	1659.0	7.0	12	.86
3f. Voltage ~ Gender*Number*Electrode*LD + (1 + Gender*Number | Subject)	35	1724.0	1865.5	-827.0	1654.0	11.9	12	.45
550 ms – 750 ms (P600)								
4a. Voltage ~ Gender*Number*Electrode + (1 + Gender*Number | Subject)	19	1093.8	1162.9	-527.9	1055.8			
4b. Voltage ~ Gender*Number*Electrode*DELE + (1 + Gender*Number | Subject)	27	1102.0	1200.1	-524.0	1048.0	7.9	8	.45
4c. Voltage ~ Gender*Number*Electrode*EIT Spanish + (1 + Gender*Number | Subject)	27	1105.2	1203.3	-525.6	1051.2	4.7	8	.79
4d. Voltage ~ Gender*Number*Electrode*LUR + (1 + Gender*Number | Subject)	27	1102.4	1200.5	-542.2	1048.4	7.4	8	.49
4e. Voltage ~ Gender*Number*Electrode*SCD + (1 + Gender*Number | Subject)	27	1103.6	1201.8	-524.8	1049.6	6.2	8	.62
4f. Voltage ~ Gender*Number*Electrode*LD + (1 + Gender*Number | Subject)	27	1102.3	1200.4	-524.1	1048.3	7.6	8	.47

Note. *p* value shows the results of the likelihood-ratio test through the ANOVA function from the lmerTest R-package (Kuznetsova et al., 2017). Electrode = C vs LA, C vs P in the LAN/N400 time window and P vs C in the P600 time window; DELE = *Diploma de Español como Lengua Extranjera*; LUR = Language Use Ratio; SCD = Spanish – English schooling difference; LD = Language Dominance.

APPENDIX B: Animacy and Grammatical Gender Comprehension Experiment

Supplementary Materials

Gender-Inclusive –e Frequency Use Questions

1. Instructions
    a. Original: ***Uso del lenguaje inclusivo de género:*** *En este estudio usted ha leído sustantivos en español escritos con lenguaje inclusivo de género. Algunos ejemplos son: 'les latines', 'une joven llamade Max', 'les estudiantes', entre otros. El propósito de esta investigación es entender cómo les hablantes de herencia de español en California procesan y utilizan estas formas. Por esta razón, le pedimos que conteste con sinceridad las siguientes preguntas. Recuerde que no hay respuestas correctas ni incorrectas.*

    b. Translation: **Use of gender-inclusive language:** In this study, you read nouns in Spanish written with gender-inclusive language. Some examples are: 'les latines', 'une joven llamade Max', 'les estudiantes', among others. The purpose of this study is to understand how Spanish heritage speakers in California process and use these forms. For this reason, we ask you to answer honestly the following questions. Remember, there are no right or wrong answers.

2. Question 1: Frequency of –e in humans
    a. Original: *¿Qué tan frecuente usa el lenguaje inclusivo para referirse* ***a seres humanos****? Por ejemplo, el uso de 'les latines' en lugar de 'los latinos.'*

    b. Translation: How often do you use inclusive language when referring **to human beings**? For example, the use of 'les latines' [the_{incl} Latines] instead of 'los latinos' [the_{masc} Latinos].

3. Question 2: Frequency of –e in animals
    a. Original: *¿Qué tan frecuente usa el lenguaje inclusivo para referirse* ***a animales****? Por ejemplo, el uso de 'les perres' en lugar de 'los perros.'*

    b. Translation: How often do you use inclusive language when referring **to animals**? For example, the use of 'les perres' [the_{incl} $\text{dogs}_{\text{incl}}$] instead of 'los perros' [the_{masc} $\text{dogs}_{\text{masc}}$].

4. Question 3: Frequency of –e in objects
    a. Original: *¿Qué tan frecuente usa el lenguaje inclusivo para referirse* ***a objetos****? Por ejemplo, el uso de 'les micrófones' en lugar de 'los micrófonos.'*

b. Translation: How often do you use inclusive language when referring **to objects**? For example, the use of 'les micrófones' [the$_{incl}$ microphones$_{incl}$] instead of 'los micrófonos' [the$_{masc}$ microphones$_{masc}$].

5. 7-point Likert scale
   a. *No lo uso en absoluto* 'I don't use it at all'
   b. *Lo uso muy poco* 'I use it very rarely'
   c. *Lo uso poco* 'I use it rarely'
   d. *Lo uso ocasionalmente* 'I use it occasionally'
   e. *Lo uso con frecuencia* 'I use it frequently'
   f. *Lo uso muy frecuentemente* 'I use it very frequently'
   g. *Lo uso todo el tiempo* 'I use it all the time'

Table B1: Noun Lemma Frequency by Animacy Category

Animacy	Lemma	Lemma Frequency	Average Lemma Frequency
Humans			2.99 (4.29)
	argentin_	1.69	
	bolivian_	1.04	
	californian_	1.12	
	chilen_	1.05	
	colombian_	3.49	
	cuban_	7.44	
	dominican_	1.14	
	ecuatorian_	.15	
	hispan_	2.98	
	hondureñ_	.11	
	latin_	7.46	
	mexican_	16.61	
	panameñ_	.44	
	peruan_	.93	
	puertorriqueñ_	1.67	
	venezolan_	.56	
Animals			53.10 (78.60)
	burr_	15.56	
	*búh_*	3.90	
	*caball_*	140.10	
	conej_	28.22	
	gat_	79.58	
	lob_	41.13	
	mon_	60.01	
	*mosquit_*	6.44	
	os_	44.21	
	pat_	58.87	
	perr_	313.91	
	*pingüin_*	6.98	
	*pulp_*	3.04	
	sap_	6.11	
	venad_	3.55	
	zorr_	37.74	
Objects			52.30 (50.0)
	*autógraf_*	8.59	
	*cuadern_*	7.57	
	*cuart_*	189.78	
	*cuchill_*	56.33	
	*diccionari_*	4.61	
	*escritori_*	29.18	
	*espej_*	37.30	
	*helad_*	41.35	

Animacy	Lemma	Lemma Frequency	Average Lemma Frequency
	*huev_*	69.35	
	*jalapeñ_*	.51	
	*micrófon_*	16.81	
	*pian_*	26.16	
	*pis_*	108.78	
	*sombrer_*	54.59	
	*vestid_*	99.47	
	*zapat_*	85.73	

Note. Italiziced words exist only in masculine singular and masculine plural so their lemma frequency reflects the total frequency of these two versions. All the other lemma frequencies are made up of grammatical gender (i.e., masculine, feminine) and number (i.e., singular, plural) quadruplets. An exception to this were the feminine singular of 'argentin_' which referred to the country's name, and the feminine plural of 'os_' which referred to the second person conjugation of the verb *osar* ('to dare'). The feminine plural of the lemma 'hondur_' did not return any counts in the Latin American Spanish EsPal subtitles corpus.

Table B2: Model Fit Indices for the Gender-Inclusive Spanish Animacy Study

Model	Parameters	AIC	BIC	Log likelihood	Deviance	Chi-square	df	*p* value
1a. log10RT ~ Animacy*Gender + (1 + Gender | Subject) + (1 | Item)	11	-1177.7	-1116.5	599.8	-1199.7			
1b. log10RT ~ Animacy*Gender + Lemma Frequency + (1 + Gender | Subject) + (1 | Item)	12	-1175.7	-1109.0	599.9	-1199.7	.03	1	.86
1c. log10RT ~ Animacy*Gender*EIT Spanish + (1 + Gender | Subject) + (1 | Item)	17	-1170.6	-1076.2	602.3	-1204.6	5.0	6	.55
1d. log10RT ~ Animacy*Gender*Language Dominance + (1 + Gender | Subject) + (1 | Item)	17	-1171.5	-1077.0	602.7	-1205.5	5.8	6	.44
1e. log10RT ~ Animacy*Gender*AY RT + (1 + Gender | Subject) + (1 | Item)	17	-1171.8	-1077.3	602.9	-1205.8	6.2	6	.41
1f. log10RT ~ Animacy*Gender*Surgery + (1 + Gender | Subject) + (1 | Item)	17	-1169.5	-1075.1	601.8	-1203.5	3.9	6	.69
1g. log10RT ~ Animacy*Gender*Upbringing + (1 + Gender | Subject) + (1 | Item)	17	-1175.8	-1081.3	604.9	-1209.8	10.1	6	.12

Note. *p* value shows the results of the likelihood-ratio test through the ANOVA function from the lmerTest R-package (Kuznetsova et al., 2017).

APPENDIX C: Animacy and Grammatical Gender Production Experiment Supplementary Materials

Seeded Narrative Task

Prompts entered into Zot GPT:

1) *Puedes inventarte una historia corta que utilice lenguaje inclusivo?*

2) *Puedes incorporar animales dentro de la historia tambien?*

Short story created:

El festival del pueblo

En un pequeño pueblo vivía **une**[+1] joven **llamade** [+1] Max, quien amaba organizar eventos para toda la comunidad. Un día, Max tuvo una idea brillante, mientras su perro Canelo movía la cola emocionado.

—¡Vamos a hacer un festival! —anunció Max en la plaza—. Un evento para todas las personas del pueblo, sin importar quiénes sean. ¡Y también para nuestros animales!

Les [+1] **vecines** [+1] se emocionaron. La ciudadanía completa empezó a colaborar: **les** [+1] profesores de la escuela ofrecieron dar talleres de arte, **les** [+1] estudiantes preparon presentaciones de música, **les** [+1] comerciantes donaron comida y bebidas, las familias decoraron las calles con banderas de colores y la veterinaria organizó un desfile de mascotas.

El día del festival llegó. Niños, niñas y **niñes** [+1] corrían felices con sus perros, gatos y conejos. **Les** [+1] artistas pintaban murales, **les** [+1] **músiques** [+1] tocaban en vivo, y la gente bailaba sin parar. Los pájaros del parque cantaban desde los árboles, y hasta un caballo adornado con flores paseaba a **les** [+1] **niñes** [+1].

Canelo y otros perros jugaban juntos. Una gata llamada Luna descansaba tranquilamente mientras las mariposas volaban alrededor de las flores.

Max miró a su alrededor con orgullo, acariciando a Canelo. —Este festival es de **todes** [+1] **nosotres** [+1] —dijo con una sonrisa—.

Esa noche, el pueblo entero celebró bajo las estrellas, recordando que cuando la comunidad se une, la magia sucede.

Figure C1: Instagram Post from the group Mecha UCI