



The gender puzzle in Spanish: Multiple factors conditioning the gender (in)congruence effect during sentence comprehension

Gabriela Mariel Zunino^{*}, Noelia Ayelén Stetie^{ID}

Universidad de Buenos Aires; Consejo Nacional de Investigaciones Científicas y Técnicas, Argentina

Key words

Psycholinguistics
gender morphology
gender stereotypes
gender congruence effect
gender incongruence effect
syntactic processing

Abstract

The effect of gender (in)congruence has been extensively studied in both comprehension and language production processes. Especially for the Romance languages, both theoretical and empirical discussions are based on the double status of gender as a lexical phenomenon with syntactic projection. Furthermore, for languages with grammatical gender and for nouns that refer to people, the interaction between semantic information associated with gender stereotypes and morphological markings seems unavoidable. Considering this general framework, we designed two sentence comprehension experiments in Spanish (Experiment 1: self-paced reading task; Experiment 2: L(lexical)-maze task) using a psycholinguistic approach to assess the interplay between gender stereotypes associated with the role noun semantics and morphological gender markings. Both experiments present a 2x2 factorial design: Semantic Bias (male vs female) and Gender Morphology (masculine, feminine), with two congruent and two incongruent conditions. Our results show patterns consistent with a facilitation effect for gender congruent conditions. Specifically, in Experiment 2, we found a significant asymmetrical effect of gender (in)congruence: male-biased role nouns represented the extreme conditions, showing a statistically significant (in)congruence effect. We discuss various factors that may influence gender (in)congruence effects during reading and sentence comprehension: incremental syntactic-semantic processes, task-dependent effects, incidence of the marked-unmarked axis in Spanish gender morphology paradigm. We propose, in line with previous studies, that, for Romance languages, gender processing involves the interaction of multiple levels, both semantic-conceptual and grammatical.

* Corresponding author

Cite this paper: Zunino, G. M., & Stetie, N. A. (2026). The gender puzzle in Spanish: Multiple factors conditioning the gender (in)congruence effect during sentence comprehension. *Topics in Linguistics*, 27(1), 1–30. <https://doi.org/10.17846/topling-2026-0001>

© 2026 Author(s). This is an open access article licensed under the Creative Commons Attribution-NonCommercial-NoDerivs License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

1.1. Gender morphology in the world's languages and the case of Spanish

Taxonomies describing gender marking in languages vary, but there are certain general categories with broad consensus (Dixon, 1987; Corbett, 1991; Hellinger and Bußmann, 2001; Leaper, 2014; Gygax, et al., 2019). One of the most recent classifications (Gygax, et al., 2019) distinguishes five types of languages: grammatical gender languages, natural gender languages, languages with a combination of grammatical and natural gender, genderless languages with some traces of grammatical gender, and genderless languages. Within the first group, where languages such as Spanish, German and Italian are found, gender governs grammatical agreement and nouns referring to both animate and inanimate entities have an assigned gender.

The gender paradigm in Spanish assumes a basic binary distinction (masculine/feminine) but exhibits a certain complexity, which has generated several attempts at systematization (Ambadiang, 1999; Roca, 2006; Mendívil Giró, 2020; Fábregas, 2024).

Gender in Spanish is characterized by a certain degree of complexity and this has raised broad debates, as it can be considered along two dimensions: one more closely linked to the lexicon and one strictly syntactic. On the one hand, it functions as a nominal classifier and, on the other, it entails syntactical features that guide the agreement with determiners and adjectives. Nominal classifiers, in any language, tend to be highly idiosyncratic and have no limit on the values they can acquire. In Spanish it is no different: the standard forms maintain two genders for nouns and three genders for pronouns; although in the varieties that adopt the non-binary form [-e], it is possible to consider that an additional nominal classifier is added (Fábregas, 2024). There is also an ongoing debate regarding the consideration of gender as part of derivational or inflectional morphology. But within the framework of these double lexical and syntactic dimensions, gender in Spanish seems to have both typically inflectional properties, as well as features more closely linked to derivational morphology, which represents a challenge for all analyses (Fábregas, 2024).

A traditional approach to gender inflection defines the problem in terms of marked and unmarked elements. It is from this distinction that, on many occasions, the so-called “generic masculine” has been analysed (Ambadiang, 1999; Cabeza Pereiro and Rodríguez Barcia, 2013; Márquez, 2013; Barrera Linares, 2019; Mendívil Giró, 2020). A widely accepted definition of masculine as unmarked gender in Spanish is offered by Ambadiang (1999). Fábregas (2024) makes a formalist analysis in a similar sense and argues that the only real value of the gender feature in Spanish is the feminine, while what has been linked to the masculine is, in fact, the absence of a gender marking. From another theoretical perspective, this projects the same scenario that traditional grammar poses about the masculine with a generic function.

For nouns referring to human entities, in Spanish epicene nouns have been widely studied because they present a single gender invariable form that refers indistinctly to men and

women, for example *persona* ('person'). However, most nouns that refer to people, and in general to animate entities, form gender pairs: *maestro/maestra* ('teacher').

Especially for nouns referring to people, numerous studies have pointed out that gender is also linked to extralinguistic factors (Ambadiang, 1999; Cabeza Pereiro and Rodríguez Barcia, 2013; Barrera Linares, 2019; López, 2020). Within the process of gender assignment, it seems practically indisputable that sociolinguistic, pragmatic, and grammatical factors are involved. As Ambadiang (1999) observes, from most of the elements analysed by the different studies, gender in nouns referring to people "tends to be biological". Thus, the gender assignment process should not be approached exclusively from a grammatical perspective.

1.2. The gender (in)congruence-effect in Romance languages: Still a matter of discussion

The so-called gender (in)congruency effect¹ has been extensively studied across different languages and under various experimental paradigms, though the results remain highly diverse and continue to generate debate.

(In)congruency can be established in different ways: between a role noun² and a personal pronoun, between the gender stereotype associated with a word and its grammatical gender, or between a visually defined referent and a pronoun, among other experimental configurations. In all these cases, the effect is typically defined as a differential processing cost driven by the congruency or incongruency between distinct sources of gender information (usually conceptual-semantic and grammatical). Two main points have traditionally been the subject of debate. First, the direction of the effect: is incongruency always more costly than congruency? The prevailing trend (especially among studies that analyse production processes) supports a facilitation effect for congruent pairings (Finocchiaro, et al., 2011; Bürki, et al., 2023), although some studies have found the opposite pattern (Paolieri, et al., 2010). Second, an ongoing question is whether the effect varies according to the language or language family. Romance languages have been a focal point in this discussion: some studies claim that the effect appears only in Slavic languages (Finocchiaro, et al., 2011; Bürki, et al., 2023) but not in Romance languages, whereas others have reported it in Romance languages as well (Paolieri, et al., 2010; Sá-Leite, et al., 2022; Casado, et al., 2023). Among the studies that have examined this phenomenon, many have focused on lexical-level production processes, i.e., the retrieval or activation of a lexical item for production. From this

¹ Both terms "gender (in)congruence effect" and "gender (in)congruency effect" will be used as synonyms.

² Literature on the topic has used indistinctly "role nouns" or "role names". In this paper, we use both as synonyms and we conceive them as terms referring to the function or position that someone has, or is stereotypically expected to have, in society, in the workplace, or in any context that involves interpersonal relationships in public or private domains. For example, some role names refer to family relationships, such as "daughter" or "grandfather", while others refer to professions, such as "teacher" or "engineer".

perspective, much of the discussion has been framed within the context of classical models of word production (Levelt, 1999; Levelt, Roelofs and Meyer, 1999).³

1.2.1. Gender (in)congruence effect in production processes

Alario et al. (2004) focus on the production of isolated words in three naming tasks with an experimental priming paradigm, where the prime was manipulated depending on whether it was congruent or incongruent in terms of grammatical gender relative to the target (drawing or written word). The results show that the prime activates a given gender and this gender in turn has an influence on the process of lexical selection. However, the authors note that these data do not allow evaluating whether the syntactic compatibility of prime and target is a requisite for the observation of the facilitation effect.

Paolieri et al. (2010), Finocchiaro et al. (2011) and Wu and Schiller (2023) work under the picture-word interference (PWI) paradigm and discuss in particular the extent to which empirical evidence supports the effect of (in)congruence in Romance languages and, specifically, what is the direction of the effect. PWI is usually used to study language production processes, especially semantic and grammatical, during word retrieval. Participants have to name a picture while a distractor is superimposed. If the distractor shares some feature with the word-picture target, it may cause interference and a competition is expected at the lexical selection level and, therefore, naming is slower.

Paolieri et al. (2010) find an unexpected direction of the effect, but verify that an effect does appear in Romance languages: their results show that items in incongruent gender condition are processed faster than those presented in congruent condition. This interference pattern in the congruent condition, in principle, is not the expected result, but the inverse. The authors explain this by postulating that there is an active inhibition of the lexical representation of the distractor noun that extends to the lexical representation of the target. Overall, they argue that this study constitutes additional evidence to support the notion that grammatical gender is an intrinsic lexical property and not a pure syntactic feature to be processed only in the production of the noun phrase, and is consistent with the hypothesis that the gender congruence effect in Italian and Spanish reflects access to appropriate nominal completion which depends on the selection of the grammatical gender. However, Finocchiaro et al. (2011) propose to revisit the findings of Cubelli et al. (2005) and investigate the possibility (or not) of generalizing these results and their implications for the understanding of the phenomena involved in lexical access. A series of PWI experiments in several languages (Spanish, Italian, French, German, and Dutch) find no evidence to support the same postulates and revisit the effect in Romance languages. They reported that participants were slower to

³ However, as one of the reviewers of this manuscript pointed out, it is worth clarifying that classical models of production are based on languages with natural gender and not with grammatical gender, so predictions may not be straightforwardly projectable to grammatical gendered languages.

produce target picture nouns when the distractor was gender-congruent in comparison with the gender-incongruent condition. The authors postulate that the gender interference effect is a rather ephemeral phenomenon and its instability cannot be attributed to specific experimental conditions. Gender interference may not be observable as reliably as, for example, the semantic interference effect. Wu and Schiller (2023), in an experimental study for Spanish, also consider transparency as a conditioning factor. With a PWI experiment, they study noun phrase production and find a significant gender congruency effect for naming latencies.

Sá-Leite et al. (2021, 2022) also work around production processes and evaluate these processes in European Portuguese, another Romance language, considering the mixed results that emerge from experimental studies. Sá-Leite et al. (2021) focus specifically on animacy as an essential factor to consider lexical processing and gender (in)congruence effects. They ran two PWI experiments in which the percentage of animate targets was manipulated (0, 25, 50, and 100%) and expected gender competitive effects only if the percentage of animate targets was low, as animacy may induce the skipping of gender selection by capturing the attention of the participants, thus diminishing activation and the competitive role of the distractors. They find gender competitive effects restricted to inanimate targets and to the 0% animacy condition, which pose an interesting interaction between animacy and gender during lexical processing in which animacy shows a semantic prioritization over gender. Sá-Leite et al. (2022), on the other hand, concentrate on the transparency of the gender-determiner mapping system and grammatical agreement restrictions. With three PWI tasks, participants were asked to name or describe each picture by producing a noun phrase (Experiment 1) or produce a bare noun (Experiments 2 and 3), while they had to ignore the distractor noun superimposed onto the picture. The results show a gender congruency effect restricted to noun phrases and that this effect is modulated by transparency, regardless of the agreement condition.

Both for this study and that of Wu and Schiller (2023), it is important to highlight the potential difference arising from the specific instruction: the case of gender recovery in Romance languages may not be the same when producing a noun phrase, which requires explicit agreement between determinant and noun, as opposed to producing a bare noun, in which the agreement requirement at least does not become explicit.

1.2.2. Gender (in)congruence effect in comprehension processes

If we consider studies that analyse the gender (in)congruence effect during language comprehension processes, papers such as Carreiras et al. (1996) and Duffy and Keir (2004) have been pioneering in many ways. Both studies analysed the influence of gender stereotypes on the construction of mental models during sentence and text comprehension in English and Spanish. Both of them analysed congruency and incongruency of explicit and implicit semantic information provided by gender stereotypes with respect to a morphologically

gendered personal pronoun. They reported the existence of a consistent and significant incongruency effect, reflected in a processing cost when the personal pronoun mismatched the stereotype associated with the role name. Specifically, Carreiras et al. (1996) found different results for Spanish and English: for Spanish, a language that morphologically marks gender through inflections in nouns and determiners, the clues about a referent's gender come not only from representations associated with stereotypes but also from explicit morphological marking.

Kreiner, Sturt and Garrod (2008) take up some of these findings to specify (using the eye-tracking technique in a language comprehension task) how different types of semantic information involved may affect the incongruency effect. On the one hand, they consider the information provided by gender stereotypes as non-specific linguistic information. On the other hand, they consider the gender information encoded in the lexicon itself as another semantic feature of some words, which they call definitional gender. Unlike Romance languages, in English, gender is usually encoded in the root of lexical forms, not in the inflectional morphology. The experimental paradigm they use is also based on analysing the effects of (in)congruence of role names (definitional, such as “king” or “mother”, or stereotypical, such as “minister” or “surgeon”) with respect to a personal pronoun, as a basic process for establishing coreference. While a significant incongruency effect was found for both types of nouns, the effect occurred earlier for definitional nouns. Additionally, incongruency was found to be more difficult to repair in order to achieve a coherent interpretation. They also presented a particularly interesting finding for cataphoric order (pronoun-noun): the incongruency effect for stereotypical nouns disappears. This suggests that the grammatical gender information provided by the pronoun consistently and unambiguously guides the construction of a mental representation of gender and the establishment of coreference with a noun that is only pragmatically associated with a particular gender but does not lexically encode that information.

Furthermore, Canal et al. (2015) and Siyanova-Chanturia et al. (2012) conducted some comprehension tasks with the electroencephalography technique (EEG-ERP),⁴ in English and Italian, to investigate the neural correlates of the incidence of gender stereotypes during language processing. They also use congruence (between nouns and pronouns) and the type of gender information encoded in the noun (definitional vs stereotypical) as central factors in the experimental design. The patterns found in each study presented particularities as well as similarities. Both found neural correlates of the gender incongruency effect found in behavioural experiments and reported consistent differences due to the type of gender information (definitional and stereotypical). Distinctly, they exhibited N400 components

⁴ The technique involves recording brain electrical activity called Event Related Potentials (ERP). Specifically, the electrical potentials generated following the presentation of a linguistic stimulus are measured (Kutas et al., 2006).

associated with gender stereotypes and P300 or P600 for words with definitional gender information, interpreted along the lines of agreement violations.

Pesciarelli et al. (2019) also worked with the EEG-ERP technique and conducted tasks within a masked and unmasked priming paradigm in Italian to explore the implicit effect of grammatical and stereotypical gender during processing of gender violations. Participants were asked to judge the grammatical gender of the target pronoun following a prime that could be either a grammatically marked or stereotypically biased role name. Gender incongruent stereotypical primes elicited larger N400 responses in both masked and unmasked conditions, which was interpreted as compelling evidence for the influence of gender stereotypes even under unconscious conditions. Casado et al. (2023), based on the study by Pesciarelli et al. (2019), also explored implicit activation of gender stereotypes during the lexical processing of nouns that refer to animate entities (animals or people) in Spanish. They analysed role names with and without stereotypicality bias, and epicene nouns referring to animals. They found a gender congruency effect for all prime conditions, even with epicene nouns. The authors conclude that grammatical gender features are activated even when bare nouns (no agreement restrictions) are processed and it can bias the gender congruency effect even when the noun is not conceptually related to gender. This might show that not just high-level conceptual information but also lexical and grammatical information are involved in the gender congruency effect.

Finally, it is worth mentioning a very recent study where Serafini and Pesciarelli (2025) proposed an adapted word-word priming task to test the (in)congruence effect not just for words but also for faces in Italian. Materials were taken from Siyanova-Chanturia et al. (2012) and Pesciarelli et al. (2019), but one half of the targets were pronouns (congruent or incongruent with respect to a previous noun) and the other half of the target were faces. Interestingly, the results found for faces are in line with previous priming studies (Banaji and Hardin, 1996; Cacciari and Padovani, 2007; Siyanova-Chanturia, et al., 2012; Pesciarelli, et al., 2019; Casado, et al., 2023). One point to remark upon is the gender asymmetry found, also in line with behavioural studies in Spanish (Stetie and Zunino, 2023; Zunino, et al., 2026): there appeared an N400 effect for violations of the female stereotype by male faces (*babysitter*—male face), but a P300 effect for the male stereotype – female faces condition (*falegname* ‘carpenter’—female face). This shows the need to consider another factor when studying the gender (in)congruence effect with role names: it might be different to represent men in female roles compared with representing women in male roles.

2. The current study

Focusing on (psycho)linguistic processing, which analyses language comprehension and production processes from a broader cognitive perspective, links with the organization of mental representations and concepts, prior world knowledge and beliefs begin to emerge. Just

as it is possible to analyse the relationship between beliefs and language by studying the effects of morphology or lexical semantics on the construction of mental representations, it is also possible to reflect on how gender stereotypes, as part of our beliefs and prior knowledge about the world, condition and influence language comprehension (Carreiras, et al., 1996; Duffy and Keir, 2004; Siyanova-Chanturia, et al., 2012; Menegatti and Rubini, 2017; Lewis and Lupyan, 2020; Stetie and Zunino, 2022; Zunino and Stetie, 2022). Within this group of studies, two phenomena stand out. The first addresses the potential gender interpretation bias for languages with mandatory gender markings and binary gender paradigms. The second focuses on a more specific point of the same problem: the construction of reference using the generic masculine in role names associated with different gender stereotypes.

Considering this general framework and the debates around the gender congruency effect (GCE) in Romance languages presented in Section 1, we designed two sentence comprehension experiments (a self-paced reading task and a L(lexical)-maze task) using a psycholinguistic approach to assess the interplay between gender stereotypes present in the role noun semantics and morphological gender markings in Spanish. Spanish seems to have a more transparent system of desinential gender marking (in comparison to Italian, Thornton, 2009; Finocchiaro, et al., 2011) and role names that form gendered pairs (rather than heteronyms, which instead mark gender on the root) are ideal examples to analyse the difference between gender information encoded on the root or through inflectional morphology. For example, *abogad[-o] / abogad[-a]* ('lawyer') does not differ in the gender information on the root at all, but only on the desinential inflectional marking. If there is any stereotypical gender information on the root it won't be related to grammatical gender but to purely conceptual-semantic information. Therefore, the gender incongruence can be generated intra-lexically: the semantic information associated with gender stereotypes and the grammatical gender features encoded on the inflexion can be congruent or incongruent.

As one of the points of debate (Sá-Leite, et al., 2022; Bürki, et al., 2023; Wu, et al., 2023) focuses on whether the (in)congruence effect arises during lexical access or, instead, during agreement processing (mandatory in Romance languages), we aim to analyse nominal phrases in sentences using two different experimental paradigms. This offers an experimental context that allows us to study effects at the word level, while also capturing and analysing potential syntactic and semantic effects (both local and global) at the sentence level. Furthermore, using two different experimental paradigms (Experiment 1: self-paced reading and Experiment 2: L-maze) allows us to compare methodological factors and potential differential effects on the results. The maze paradigm is known to be a more demanding task but a very rigorous technique for assessing effects on the word, avoiding spillover effects (Forster, et al., 2009; Boyce, et al., 2020). In contrast, cumulative self-paced reading is considered a more ecological paradigm, but less precise for detecting fine-grain local effects and may not ensure semantic integration, especially when presented without comprehension questions (Boyce, et al., 2020).

Therefore, the use of both experimental paradigms will allow us to discern whether the possible gender (in)congruence effect occurs at the determiner, the role noun or the subsequent word, which, in turn, will enable us to add evidence to the discussion about whether gender recovery occurs during lexical access or agreement processing.

Lastly, it is important to note that we wanted to test potential differences depending on the distributional structure of the sentence, considering whether the nominal phrase with the role noun is presented at an initial or final position. While effects stemming from local grammatical processes like agreement should not be affected, the process of integration of conceptual-semantic information (such as gender stereotypical information on role nouns) might change due to incremental processing (Pickering and Traxler, 2000; Karimi and Ferreira, 2016). In the case where the noun phrase appears at the beginning of the sentence, there is no semantic information that can be used to compute gender, and the grammatical gender information and the agreement restrictions between determiner and role name would be the basic key. On the other hand, when the noun phrase appears at the end of the sentence, the semantic scaffolding constructed from the incremental integration during reading can provide elements that operate at a higher order conceptual level in the retrieval of gender information of the role name.

To summarize, there are three general research questions that guide this study: [a] What conditioning imposes the requirement of gender agreement (a morphological-lexical phenomenon) within the NP in Spanish (as in all Romance languages) for the generation/detection of the GCE?; [b] What effect could the linear organization of the sentence (distributional phenomenon) have on GCE generation/detection?; [c] What effect could the type of task and the experimental paradigm used (underlying process and task related phenomenon) have on GCE generation/detection?

Based on the previous literature and the theoretical considerations discussed above, we propose the following experimental hypotheses:

1. There will be a positive congruence effect between stereotypical bias and gender morphology: when stereotypical bias and gender morphology are congruent, times will be shorter, thus congruence will facilitate the process.
2. There will be a negative incongruence effect between semantic bias and gender morphology: when stereotypical bias and gender morphology are incongruent, times will be larger, thus incongruence will hinder the process.
3. There will be differences between noun phrases presented in initial versus final position: when presented in final position, role nouns are expected to show a stronger stereotype bias effect due to ongoing global semantic integration mechanisms.

2.1. Experiment 1

2.1.1. Method

2.1.1.1. Participants

The task involved 301 Spanish speakers from Argentina. Fifteen participants had to be removed: 5 persons participated twice in the experiment and 10 participants systematically did not reach the minimum time of 100 ms, considered the lower time limit for word reading (Hartley, et al., 1994). Of the remaining 286 participants, 181 were women, 66 were men, 9 were non-binary, and 30 did not report their gender identities. Participants were between 18 and 76 years old ($M = 32.93$; $SD = 12.03$) and all had completed high school. Participants were recruited through social media, email and other free digital media. Participation was voluntary and unpaid.

2.1.1.2. Materials

We designed 40 sentences with role nouns, 10 for each condition according to Semantic Bias (male vs female) and Gender Morphology (masculine, feminine). All sentences began with a noun phrase (determiner + role noun) and then a predicate with no subordinate clauses. The only reference to gender appeared in the subject, both in the determiner and the role noun. Sentence length was controlled: between 11 and 16 words ($M = 13.50$; $SD = 1.26$) and between 85 and 90 characters ($M = 87.47$; $SD = 1.29$), with the same mean for all conditions (87.4 characters). Female or male semantic bias is dependent on cultural knowledge, as was already noted in the Introduction. Therefore, we conducted a prior normative study with a preselection of 12 role names (see Stetie and Zunino, 2023; Tzinavos Muñoz, et al., 2025). From the data obtained in the normative task, we selected 10 final stimuli for each level of Semantic Bias. Examples (1) and (2) present items with semantic female and male bias, respectively, all with both, masculine and feminine morphology.

(1) *Los enfermeros / Las enfermeras sostienen gran parte de la demanda en*
The.M nurses.M / The.F nurses.F support much of the demand in
los hospitales y centros de salud.
hospitals and healthcare centres.

(2) *Los carniceros / Las carniceras solo desgrasan bien*
The.M butchers.M / The.F butchers.F only defat well
los cortes de carne y pollo para sus mejores clientes.
beef and chicken cuts for their best customers.

Regarding the fillers, we used 80 stimuli that corresponded to another experiment, as shown in example (3). In addition, to control that participants were paying attention to the task and constructing a global representation of the sentences they read, we included comprehension questions with four response options (see example 4) for half of the stimuli, both experimental and filler. We divided the sentences into three counterbalanced lists (i.e., each participant was exposed to only one of the two versions of each item) with 40 items each, of which 20 presented comprehension questions. To ensure a proper balance, two lists contained 13 experimental items and 27 fillers, while the third consisted of 14 experimental items and 26 fillers. No participant saw the same role name twice.

(3) *La joven defendió al hijo del profesor que se copió*
The young woman defended the son of the teacher.M who cheated
durante la evaluación de matemática.
during the math test.

(4) *¿En qué evaluación se copiaron?*
In which test did they copy?
a. matemática; b. física; c. química; d. geografía.
a. mathematics; b. physics; c. chemistry; d. geography.

2.1.1.3. Procedure

The task was designed using IBEX software (Internet Based Experiments: Drummond, 2013) and one of the three lists was randomly assigned to each participant. First, we presented an informed consent that had to be accepted to access the demographic data questions and the experiment. The experiment consisted of a self-paced reading task with a cumulative moving window paradigm: participants read a sentence word by word at their own pace. In half of the cases, after each sentence, they moved to another screen, where they were presented with a comprehension question with four options. Once they selected the answer, they moved to the next screen with a fixation point. At this point they could either rest or continue with the next trial. We made it clear to them that they would have four practice sentences, which were followed by three more training items, which, for them, were already part of the experiment. The task could only be performed on a computer with an internet connection.

2.1.2. Results and discussion

We considered two different sentence regions as dependent variables; as we wanted to focus on the noun phrase (NP) processing, here we analysed the NP reading time and the spillover region reading time, that corresponds to the verb in all sentences.

Statistical analysis was performed using R software version 4.4.2 in the R Studio interface (R Core Team, 2024). We used the following packages: tidyverse (Wickham, et al., 2019), lme4 (Bates, et al., 2015), lmerTest (Kuznetsova, et al., 2017), sjPlot (Lüdtke, 2023) and MASS (Venables and Ripley, 2002). Data and analysis code are available at Open Science Framework (OSF): <https://osf.io/rk7uz/>.

For the final analysis, we discarded items that were answered erroneously (1.66% of the data). In addition, because the task was performed remotely without being able to control the situational context during data collection, we set time limits for its completion (lower limit: 1000 ms; upper limit: 10000 ms). This involved the removal of 11 data points (0.30% of the total sample). Additionally, we conducted an identification of outliers for a possible posterior imputation, however, there was no measure that exceeded the 2.5 standard deviations for each participant in each condition (Baayen and Milin, 2010; Cousineau and Chartier, 2010).

We analysed the time participants took to read the role noun and the verb, which we treated as the spillover region. As shown in Figure 1, as expected, no effect is generated on the determiner. Instead, male-biased role nouns with masculine morphology had the shortest processing times, both in the role noun and in the subsequent word, the verb. In the place of the noun the other three conditions are assimilated, while in the spillover zone, the verb, the two incongruent conditions differ from the two congruent conditions. For the statistical analysis, we tested the assumptions of normality and homoscedasticity based on Linear Mixed Models. This led us to perform a logarithmic transformation (Winter, 2019; Vasishth, et al., 2021). Additionally, we coded fixed factor levels as scaled sum contrasts (Schad, et al., 2020). The models used for the analysis included Semantic Bias and Morphology as fixed effects and Participants and Items as random effects.⁵ The resulting model was: $\text{lmer}(\log(\text{reading_time}) \sim \text{Semantic Bias} / \text{Morphology} + (1 | \text{Participants}) + (1 | \text{Items}))$.⁶ The results for each variable are presented in Table 1.⁷ Although there is a clear pattern that favours congruent conditions

⁵ Although it was not the primary focus of the present study, we included a supplementary analysis, available at OSF (<https://osf.io/rk7uz/>), to explore potential differences in reading times across the gender identities of the participants. No statistically significant differences were found based on the participants' gender identity; however, this evidence should be interpreted with caution, as the experiment was not specifically designed to test this variable and, hence, the sample was not balanced across gender groups.

⁶ The model used is similar to the more usual one: $\text{lmer}(\log(\text{reading_time}) \sim \text{Semantic bias} * \text{Morphology} + (1 | \text{Participants}) + (1 | \text{Items}))$, since it also considers interactions. Our decision to use nested models was guided by clearly defined hypotheses with specific planned comparisons. Addressing these analyses separately allowed us to directly address the different research questions avoiding additional comparisons that are not theoretically motivated. We avoided omnibus analyses with full models that usually have overparameterization and to increase the risk of Type I errors (Schad et al., 2020; Vasishth & Nicenboim, 2016). Random slopes were not considered due to convergence issues.

⁷ As we did not expect effects on the determiner, we did not report here statistical analysis for the determiner, but we presented it in Supplementary Materials. One of the key points for the Romance languages is precisely the computation of grammatical gender information through the agreement between determiner and name in NPs. The determiner conveys information that generates expectations about the grammatical gender of the name

(in line with the gender congruence effect, Bürki, et al., 2023; Casado, et al., 2023), we found no statistically significant effects due to semantic bias in interaction with gender morphology on the role noun. For the verb reading time (spillover effect), our attention focused on morphology nested within the male bias. In other words, in sentences with male-biased role nouns, those with masculine morphology (congruent) were read faster than those with feminine morphology (incongruent), but no statistically significant effect was found. Although not significant, there is a clear pattern. In Figure 1, it is observed that sentences with male-biased role nouns and congruent morphology were the ones that were read the fastest among the four conditions, while those with male bias and feminine morphology (incongruent) had the longest reading times.

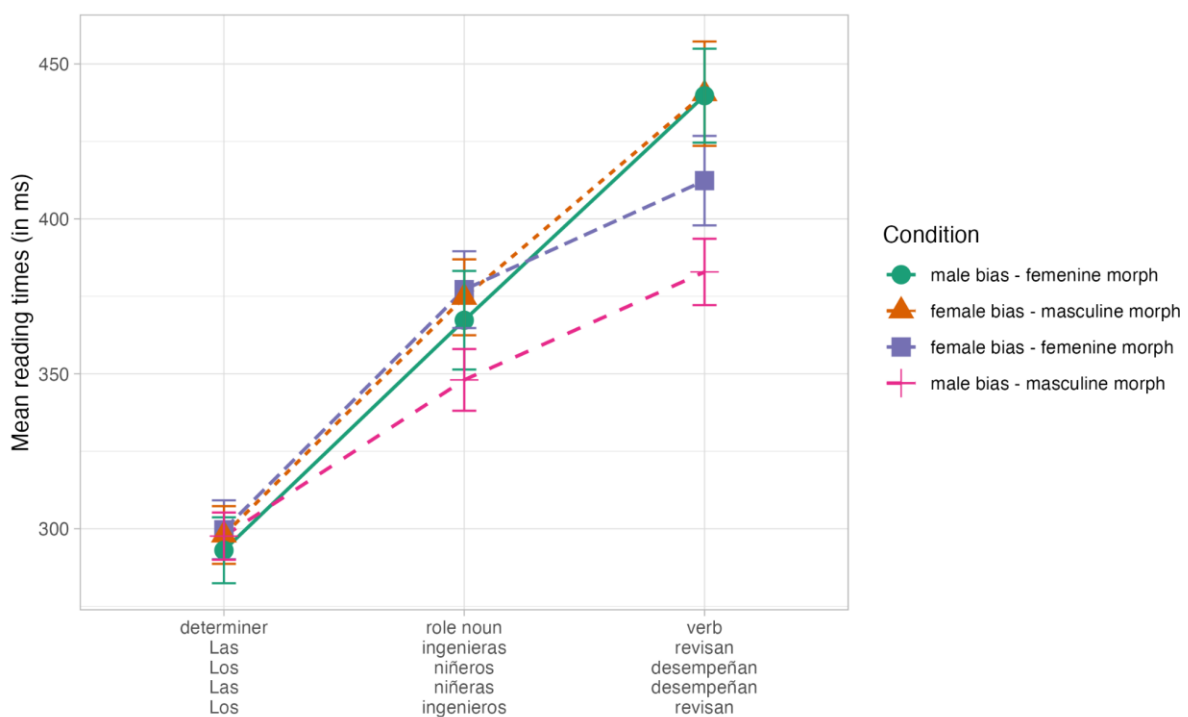


Figure 1. Mean reading times for determiner, role noun and the verb by Semantic Bias and Morphology with error bars (in ms).

to be found in an incremental processing scheme and, therefore, operates when processing the gender information in the role name (Alario, et al., 2004, Wicha, et al., 2004).

Table 1. Summary of LMM analysis for role noun and verb reading times

Predictors	log(RT_noun)				log(RT_verb)			
	Estimates	SE	T	P	Estimates	SE	T	P
(Intercept)	5.67	0.03	193.80	<0.001	5.78	0.03	191.21	<0.001
Semantic Bias	-0.03	0.02	-1.77	0.084	-0.02	0.03	-0.88	0.384
SBias _{male} / morphology	-0.01	0.03	-0.21	0.836	-0.06	0.04	-1.71	0.096
SBias _{female} / morphology	0.00	0.03	0.02	0.984	0.03	0.04	0.79	0.435
Random Effects								
σ^2	0.14				0.19			
τ_{00}	0.22 _{part}				0.21 _{part}			
	0.00 _{item}				0.00 _{item}			
ICC	0.62				0.53			
N	286 _{part}				286 _{part}			
	40 _{item}				40 _{item}			
Observations	3669				3669			
Marginal R ² / Conditional R ²	0.001 / 0.619				0.002 / 0.535			

2.2. Experiment 2

2.2.1. Method

2.2.1.1. Participants

The task involved 145 Spanish speakers from Argentina, from which three had to be removed because they were not Argentinian or were not living in the country. Of the remaining 140 participants, 106 were women, 31 were men, 3 were non-binary. Participants were between 21 and 75 years old ($M = 37.24$; $SD = 13.01$) and all had completed high school. Participants were recruited through social media, email and other free digital media. Participation was voluntary and unpaid.

2.2.1.2. Materials

We designed 48 sentences with role nouns, 12 for each condition according to Semantic Bias (male vs. female) and Gender Morphology (masculine, feminine). All sentences were presented in passive voice in order to end with a noun phrase (determiner + role noun + adjective). The only reference to gender appeared in this noun phrase. All sentences were of similar length: between 11 and 15 words ($M = 12.58$; $SD = 1.23$) and between 82 and 97 characters ($M = 90.77$; $SD = 3.88$), with the same mean for all conditions (12.6 words). To select role names, we used the same prior normative studies from which we chose the items

for Experiment 1 (see Stetie and Zunino, 2023; Tzinavos Muñoz, et al., 2025). Examples (5) and (6) present items with semantic female and male bias, respectively.

(5) *Gran parte de la demanda en el área pediátrica es sostenida por los enfermeros universitarios / las enfermeras universitarias.*
Large part of the demand in the paediatric area is sustained by the.M nurses.M university-trained.M / the.F nurses.F university-trained.F

(6) *La suba del precio de la nafta fue resistida enérgicamente por los camioneros afectados / las camioneras afectadas.*
The increase in gas prices was strongly resisted by the.M truck drivers.M affected.M / the.F truck drivers.F affected.F

Since the L-maze task requires participants to choose each word that continues a sentence, they must choose between the word that actually belongs in the sentence and a pseudoword. In this case, the pseudowords were constructed from all the words in the experimental items and matched in length and number of syllables. In (7), we present an example of the competitor pseudowords that was associated with item (5).

(7) x-x-x minte ta sa ñostica an ke miba merrarpina us miudinido rin yan lerrircute vonfaderamalte.

Additionally, we used 48 distractors that corresponded to the stimuli from another experiment, as illustrated in (8). We organized all the items into four lists with 24 items each: 12 experimental and 12 fillers.

(8) *Lucía defendió al pasante de la funcionaria que estaba muy convencido de su opinión.*
Lucía defended the intern of the official who was very convinced of his opinion.

2.2.1.3. Procedure

In this case we designed an L-maze task and used the PCibex platform (Zehr and Schwarz, 2018). First, we presented an informed consent that had to be accepted to access the demographic data questions and the experiment.

The L-maze task consists in selecting one of two items (one word and one distractor, in this case a pseudoword) presented on a screen in order to form a complete sentence, as shown in Figure 2. Participants had, at their own pace, to select which word was the adequate one on each screen to construct the sentence. The experiment began with four practice sentences, and to help participants understand the dynamics of the task, they were given feedback, i.e., they

were informed if they had selected an incorrect option and were allowed to correct their choice. After the practice, two training sentences followed, which participants perceived as already part of the task, though they were not included in the data analysis. In all cases, participants saw two words at the centre of the screen, as shown in Figure 2. At the end of each item, a fixation point appeared in the middle of the screen for 500 ms before moving on to the next sentence. If they selected the pseudoword, a message appeared indicating they had made an error. In addition, since the task was very demanding, we included three mandatory breaks during which participants could take as much time as they needed. The task could only be performed on a computer with an internet connection.

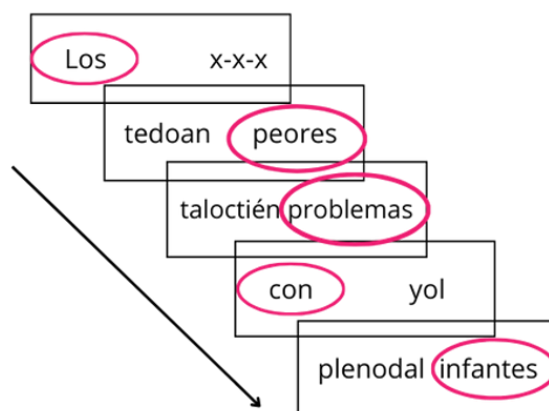


Figure 2. Representation of the lexical maze task

2.2.2. Results and discussion

We considered two different sentence regions reading time as dependent variables: the role noun and the posterior adjective (see Figure 3). Statistical analysis was performed using the same software and procedures as in Experiment 1. Data and analysis code are also available at OSF: <https://osf.io/rk7uz/>.

For the final analysis, we discarded items that were answered erroneously (13.56% of the data).⁸ We applied the same procedures for controlling extreme values and identifying outliers than in Experiment 1. Only one data point (0.07% of the total sample) was removed due to an extreme value and no outliers required imputation.

As shown in Figure 3, male-biased role nouns with masculine morphology had the shortest processing times, both in the role noun and in the adjective. In this case it was not considered as a mere spillover zone (in fact, the maze technique significantly reduces this type of effect: Forster, et al., 2009; Boyce, et al., 2020). The adjective in Spanish requires gender agreement with determiner and name within a NP, so from analysing the processing time there, it is

⁸ This result is not surprising, as maze tasks typically yield a high number of errors (Forster, et al., 2009; Boyce, et al., 2020).

possible to obtain data on the potential facilitation in some of the conditions for the processing of more complex NPs, with a triple agreement restriction. For example, if the particular condition of (in)congruence in the role name may affect gender computation in the adjective.

The reading times for the role noun showed the same pattern as in Experiment 1: the next fastest condition was that of female-biased role nouns with feminine morphology, followed by the two incongruent conditions, with the male-biased role noun with feminine morphology showing the longest reading times. However, these patterns did not replicate in the adjective.

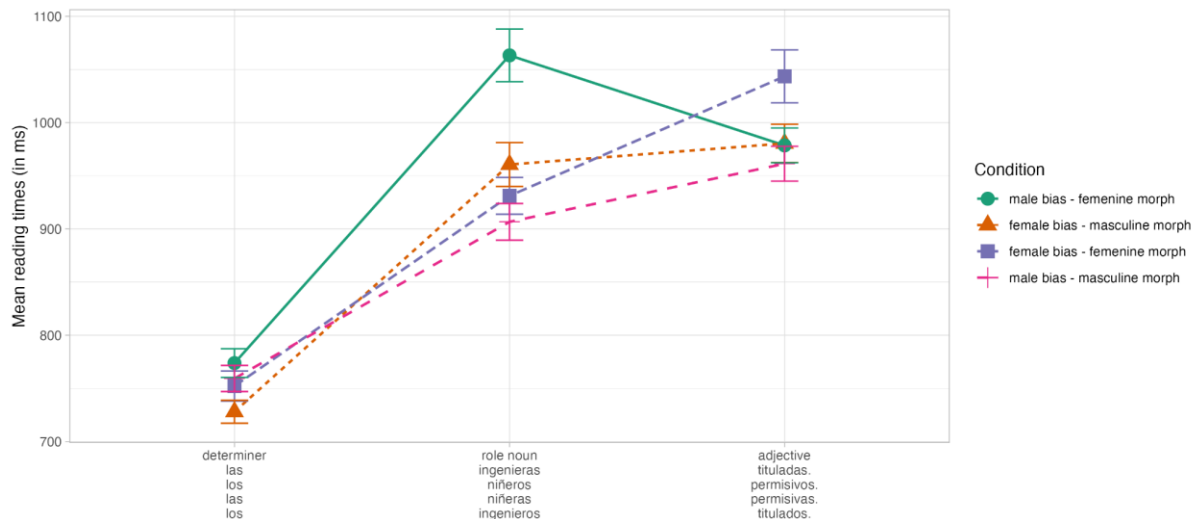


Figure 3. Mean reading times for determiner, role noun and adjective by Semantic Bias and Morphology with error bars (in ms).

For the statistical analysis, we followed the same procedure as in Experiment 1 and used Linear Mixed Models (Winter, 2019; Vasishth, et al., 2021). We also coded fixed factor levels as scaled sum contrasts (Schad, et al., 2020). The models used for the analysis also included Semantic Bias and Morphology as fixed effects and Participants and Items as random effects: $\text{lmer}(\log(\text{reading_time}) \sim \text{Semantic Bias} / \text{Morphology} + (1 | \text{Participants}) + (1 | \text{Items}))$.⁹ The results for each variable are presented in Table 2.¹⁰ We didn't find statistically significant differences in the adjective reading times and, for the role noun, the only statistically significant difference was found between the congruent and incongruent conditions involving male bias, specifically, between male-biased role nouns with masculine morphology and those with feminine morphology (i.e., *ingenieros* vs. *ingenieras*, 'male engineers vs. female

⁹ In Experiment 2, we also used a complementary analysis to assess whether there were statistically significant differences due to the participants' gender identity variable. The analysis is available at OSF (<https://osf.io/rk7uz/>), and here too we found no statistically significant differences.

¹⁰ See footnote 1.

engineers’). In other words, we find a significant effect of gender (in)congruence only for male-biased role names.

Table 2. Summary of LMM analysis for role noun and adjective reading times

Predictors	log(rt_noun)				log(rt_verb)			
	Estimates	SE	T	P	Estimates	SE	T	P
(Intercept)	6.81	0.02	303.66	<0.001	6.85	0.02	351.03	<0.001
Semantic Bias	-0.03	0.03	-1.04	0.304	0.03	0.02	1.20	0.236
SBias _{male} / morphology	0.14	0.04	3.76	<0.001	0.02	0.03	0.50	0.620
SBias _{female} / morphology	-0.02	0.04	-0.47	0.642	0.04	0.03	1.18	0.245
Random Effects								
σ^2	0.05				0.06			
τ_{00}	0.05 _{part}				0.03 _{part}			
	0.01 _{item}				0.00 _{item}			
ICC	0.50				0.40			
N	140 _{part}				140 _{part}			
	48 _{item}				48 _{item}			
Observations	1492				1492			
Marginal R ² / Conditional R ²	0.024 / 0.509				0.004 / 0.339			

3. General discussion

As we developed in the Introduction, the phenomena linked to gender processing are manifold and research on the gender (in)congruence effect (GCE) has been approached from various paradigms and models (Siyanova-Chanturia, et al., 2015; Bürki, et al, 2023; Casado, et al., 2023), both in production processes and in language comprehension. One of the central debates is related to defining where and when gender information is computed and retrieved (Alario, et al., 2004; Finocchiaro, et al., 2011; Sá-Leite, et al., 2021). The search for gender (in)congruences has been a good clue to analyse this phenomenon.

In the different studies, the locus at which (in)congruence is experimentally generated also varies. In some of them, gender (in)congruence between noun and pronoun is studied. Others explore the effect of context and several conditions are considered: lexical, phrasal, sentential (Carreiras, et al., 1996; Paolieri, et al., 2010; Pesciarelli, et al., 2019). Some studies also focus on animacy as an independent variable (congruence effect in inanimate nouns: Casado, et al., 2021) or explore semantic information projected from nouns and faces (Serafini and Pesciarelli, 2025). Several factors have also been analysed as conditioning for the generation of the GCE during production (Finocchiaro, et al. 2011). The language under study is perhaps the most coarse-grained one: there is an intense debate about whether Romance languages do

indeed exhibit GCEs. This grouping is not random; these are languages with grammatical gender, with particular gender morphology paradigms. In those cases, the discussion about whether gender is a purely lexical phenomenon or also syntactical is profuse (Thornton, 2009; Finocchiaro, et al., 2011; Fábregas, 2024).

Among the factors considered to be influential, studies have addressed animacy (Sá-Leite, et al., 2021), transparency of the morphological gender system (Sá-Leite, et al., 2022; Wu, et al., 2023), type of noun in terms of gender (Kreiner, Sturt and Garrod, 2008; Siyanova-Chanturia, et al., 2015; Casado, et al., 2023), agreement restrictions with the determiner (Alario, et al., 2004; Wu, et al. 2023), and even the interplay between linguistic and non-linguistic stimuli (Sá-Leite, et al., 2021; Serafini and Pesciarelli, 2025), which make it possible to consider gender processing, especially linked to stereotypes, also as a semantic-conceptual phenomenon of high order, which in Romance languages, may interact with morphological marking (Zunino, et al., 2025; Zunino, et al., 2026).

In this work, we focus on exploring the occurrence of GCE on role nouns during sentence reading and comprehension. As discussed in Section 1, role nouns that are formed using generic pairs in Spanish (i.e., *abogado-o*; *abogada-a*, ‘lawyer’) are particularly interesting for analysing the GCE and discussing the locus of where (and what kind of) gender information would be encoded. In these cases, the lexical root seems not to carry any grammatical gender information; this information is encoded in the desinence. However, the nominal base can carry gender information at a semantic-conceptual level through stereotyping bias (Cacciari and Padovani, 2007; Brauer, 2008; Misersky, Majid and Snijders, 2019; Pinheiro and Freitag, 2020; Stetie and Zunino, 2022; Zunino and Stetie, 2022). This characteristic of role names in a language with a high degree of morphological transparency such as Spanish (Sá-Léite, et al., 2022; Wu, et al., 2024) leads to the fact that the (in)congruence can be generated intralexically between the semantic information at the root and the morphological gender marking.

We designed two experiments assessing the effects of gender stereotypicality on lexical semantics and gender morphology. First, we performed a self-paced reading task with a cumulative moving window paradigm. Second, we ran an L-maze task. In both we manipulated two factors: semantic bias and gender morphology. Furthermore, in Experiment 1 we tested sentences with NPs containing the target role noun in initial position and in Experiment 2 we analysed sentences with NPs in final position and triple gender agreement (Det + N + Adj) restrictions. These allowed us to evaluate the possible GCE on role nouns considering different conditioning variables: type of process due to the task, distributional effect within the sentence, grammatical gender, and gender stereotype of the noun.

In this work we focus on the detection of the effect on the role noun, however, from its insertion into an NP and with a sentence processing task, the syntactic projection of the gender is necessarily involved. Although we did not expect to find any effect at the determiner, it is decisive that in this place grammatical gender information due to agreement restrictions

begins to be computed (Alario, et al. 2004; Wicha, et al., 2004; Sikos, et al., 2016; Fábregas, 2024). So, at the time of reading the role noun, there should be pre-activated information (which might generate a predictive expectation) about the noun that should follow. From this information, used as a predictive track, two sources of information in the role noun could be coordinated to generate the GCE: on the one hand, predictive expectation generated by the determiner (under both syntactic and semantic constraints) and, on the other hand, the intra-lexical (in)congruence between the noun stereotype bias (encoded in the root) and the grammatical gender inflexion. In addition, in Experiment 2, we were also able to verify what may happen with gender agreement computation within the NP but with an adjective presented after the noun.

In general terms, our first two hypotheses predicted the existence of a GCE in which the congruent condition is facilitating, while the incongruent condition is hindering, in line with numerous previous findings (Bürki, et al., 2023; Casado, et al., 2023; Wu, et al., 2023). In addition, we considered the possibility that the position of the NP within the sentence would make a difference in processing, in terms of what information is available at the time the role noun is processed. Taken as a whole, we can say that our results support our hypotheses just partially. We did find a general pattern of reaction times on the role noun consistent with the GCE predicted in our hypotheses, but the differences exhibited by this pattern were only statistically significant for male-biased role nouns in Experiment 2. Furthermore, this fact yields the prediction made in our third exploratory hypothesis: there were differences between Experiment 1 and Experiment 2 that might be due to the NP distributional factor. Based on our results, and in the context of the current discussions on this phenomenon, we propose some more fine-grained reflections organized around three elements that will also open up questions for future research.

1. Could it be a distributional factor conditioning gender (in)congruence effect for role names in noun phrases?

In the first experiment, where the NP was in initial position and computation might be considered more lexically than syntactically driven, the differences are not statistically significant. This could be interpreted in contrast to other experimental studies that did report an early effect (Siyanova-Chanturia, et al., 2012; Canal, et al., 2015). However, in Experiment 2, where NPs were in final position and appeared after incremental integration, we found a similar pattern but with a statistically significant gender incongruence effect for male-biased role names.

It is insightful to compare our analysis with findings such as those reported by Kreiner, Sturt and Garrod (2008). The authors show different effects depending on whether stereotypical or definitional role names are involved. Casado et al. (2023) also analysed the potential difference between the source of information involved in gender retrieval:

grammatical gender feature or more high order semantic-conceptual information. The authors discuss that GCE may indeed exhibit a double incidence of lexical-grammatical and conceptual-semantic levels to compute gender. When analysing the animacy effect, Sá-Leite et al. (2021) also considered the complexity of multiple sources to compute gender information and proposed a competition effect between purely grammatical gender features and conceptual encoding of gender information associated with animacy.

Apart from the interplay between conceptual (gender stereotype) and grammatical (morphological gender) information, our results might also be exhibiting distinctions associated with the psycholinguistic process conditioned by the distributional organization of the sentence. For Experiment 1, with a process more anchored in lexical processing of role names and more locally restricted by NP agreement requirements, no significant effect was found. In Experiment 2, where processing of role names is conditioned by the syntactic-semantic scaffolding of the incremental sentence processing where a global conceptual representation of a whole event is generated, we did find a significant gender incongruence effect. Predictive processing proposals (Huettig, 2015) have analysed how a more profuse and fine-grained context generates more precise (and also more biased) semantic expectations (Kamide, et al., 2003; Matusalem, et al., 2012). Both semantic and syntactic features are involved in these incremental processes, and both seem to interact to generate anticipatory predictions, so the amount of previous information computed before reaching the NP can clearly affect the GCE on the role name. Taken together, these results can be interpreted as evidence that gender is a complex construct that co-involves multiple sources of information during language processing.

2. Might gender (in)congruence effect be task-dependent?

Many differences have been reported in terms of the psycholinguistic process involved (production and comprehension). But there is also a variety of experimental techniques and tasks that can significantly affect the results found (Casado, et al., 2023; Bürki, et al., 2024). Our results show that the task can indeed condition the effects found.

In this study, we proposed to analyse the phenomenon in the context of processing NPs within sentences in which, besides varying the distributional position of the NP, we tested two different techniques. The decrease in spillover effects and the requirement for immediate integration of syntactic-semantic information offered by the maze tasks (Forster, et al., 2009) may have helped to exhibit effects on the role noun.

Although the processing pattern conditioned by the two main factors in the two tasks is the same, only with the maze technique was it possible to register statistically significant differences. This makes it possible to assume that many of the previous findings could be strongly influenced by the experimental technique used, by the paradigm of presentation of the items or by the task itself. The self-paced reading paradigm may not be accurate enough

to record effects that do exist or record them without too much spillover. The maze technique has been highlighted for its precision and for its ability to ensure syntactic-semantic integration in the incremental course of sentence construction (Forster, et al., 2009; Boyce, et al., 2020). The results of this study seem to support this.

In addition, it allowed us to analyse potential effects of agreement within the NP between the noun and the posterior adjective, without this data being confused with spillover effects. Analysing NP with triple agreement is of interest in the framework of studying at what point gender information is recovered and to what extent in the temporal course of the process. In this case, we could observe that the gender incongruence effect found for nouns with male bias is not projected to the adjective, the point at which the semantic information of the stereotype seems to be dismissed, and only the grammatical gender information is relevant for the agreement checking. The technique facilitated especially the interpretation of these results. The change in the pattern and the absence of significant differences in the adjective, where, in principle, there is no semantic bias, can also be interpreted as reinforcing the idea that in the role noun the factor which effectively tracts the differences is the semantic bias linked to gender stereotypes and not the recovery of the grammatical gender *per se*. We will return to this point in the next sub-section in relation to the proposal of an unmarked stereotype, in addition to an unmarked grammatical gender in Spanish.

3. Exploring the gender asymmetry of the gender (in)congruence effect

Consistently, in both tasks, the male-masculine congruent condition shows the shortest reading times. As shorter reading times would indicate an easier process, we can propose that this particular congruent condition (male-masculine) is taken as the default one. Interestingly, we did not find a main GCE but an asymmetrical one: only male-biased nouns with feminine grammatical gender (i.e., *camioneras*, ‘female truck-drivers’) showed a significant incongruence effect. Therefore, congruence is not *per se* easier or by default, instead this particular congruent condition that presents a male stereotypical noun role with masculine morphology seems to be.

Seen from the grammatical gender perspective, feminine morphology causes a difficulty that is strongly linked to the role name’s semantic bias, and that is not transversal for all nouns. In other words, we may interpret that processing role nouns that construct a reference to women performing stereotypically masculine roles showed significantly higher processing times, thus a greater cost. Siyanova-Chanturia et al. (2015) and Serafini and Pesciarelli (2025) also reported asymmetrical patterns in their results. The first study found that female-biased professions carried out by men were more bearable than the opposite, as per our results. They explained these findings as an effect of men being regarded as the “standard” social category, that is, the unmarked normative group. The second study discusses a differential pattern in ERP components conditioned by the direction of the incongruence: N400 for violations of the

female stereotype by male faces (e.g., *babysitter*—male face), but P300 for violations of the male stereotype by female faces (e.g., *falegname* ‘carpenter’—female face). These findings enable us to discuss the notion of unmarkedness, not only for the morphological gender marking (Ambadiang, 1999; Fábregas, 2024), but also for the semantic male bias that role nouns may carry.

Furthermore, the differential behaviour between the two binary morphologies might be reflecting the imbalance of functions between the feminine and masculine morphologies in Spanish (also discussed by Casado et al., 2023). Masculine, as unmarked in Spanish, may have the double function of “generic” and “exclusively men” that the feminine morphology does not. “*Las camioneras*” (male-biased and feminine morphology condition, ‘female truck-drivers’) imposes the restriction of representing exclusive groups of women in stereotypically male roles, and this seems especially costly. As the masculine morphology is an ambiguous form in Spanish, its generic function is tied to information of diverse nature, such as world knowledge, and not just grammar. In this sense, the generic function of the masculine morphology is not always identifiable, and, in many situations, this form causes a male bias conditioned by the role names’ level of stereotypicality (Stetie and Zunino, 2022; Zunino and Stetie, 2022).

This poses a scenario in which, for Romance languages, it is not only morphology but also stereotypical mental representations of gender that may operate in terms of the traditional marked/unmarked axis (Eagly and Kite, 1987; Siyanova-Chanturia, et al., 2012), and might be conditioning the generation of GCE. Future studies will allow us to further refine these hypotheses.

These three lines of discussion have allowed us to analyse and compare the results of our two experiments, both with each other and in relation to previous studies. They have also helped us to outline some emerging certainties regarding gender processing in Romance languages, to identify nuances and influencing factors in the gender (in)congruence effect and to propose several promising directions for future research.

4. Limitations and future directions

While research on gender processing and the gender (in)congruence effect has flourished in recent years, much remains to be unravelled regarding the intricate interplay of variables that modulate this complex phenomenon.

In the future, it would be interesting to investigate distributional effects and task-dependent differences in an isolated and systematic way that was not possible in this study. Also, it would be necessary to be able to isolate distributional effects from effects due to the internal structure of the NP, for example the potential difference between a simple agreement between Det + N

versus a complex agreement such as Det + N + Adj, which requires gender activation in the three retrieved words.

Furthermore, it would be interesting to investigate possible differences linked not only to the experimental technique but to the task itself, for example through hybrid tasks of sentence reading and comprehension with the completion of a final NP. Or even sentence comprehension tasks conducted completely in auditive modality.

Finally, it seems to be key to highlight the need to deepen our understanding of potential biases linked to the speaker/listener's gender identity. Not only is greater evidence required within the binary paradigm (cis men and cis women), but it is also essential to know in what way the scene can be modified with participants who define themselves as non-binary, queer, transgender, gender fluid, among other possible gender identities.

Acknowledgement

This work has been co-funded by the European Union through the *MultiLingualGender* project (MSCA Staff Exchanges, Horizon Europe, Grant Agreement No. 101182959).

Ethics and consent

This study was conducted with voluntary, anonymous, and unpaid participation of adults. Each participant could stop the task at any time they wished. The study was under the support of the Habeas Data Law of the country. The University of Buenos Aires does not request authorization from the Ethics Committee when the tasks are not invasive, nor require access to medical records or personal identification data and the participants are of legal age. However, the Institute of Linguistics of the University of Buenos Aires has approved by the Ethics Committee of the institution the documents that must be used as informed consents and this format was the one used in this study, which ensures data protection under our country's Habeas Data Law.

References

- Alario, F-X., Matos, R., & Segui, J. (2004). Gender congruency effects in picture naming. *Acta Psychologica*, 117, 185–204. <https://doi.org/10.1016/j.actpsy.2004.06.003>
- Ambadiang, T. (1999). La flexión nominal: género y número. In I. Bosque & V. Demonte (dirs.), *Gramática Descriptiva de la Lengua Española* (pp. 4843-4914). Espasa Calpe.
- Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, 3(2), 12–28. <https://doi.org/10.21500/20112084.807>
- Banaji, M. R., & Hardin, C. D. (1996). Automatic stereotyping. *Psychological Science*, 7, 136–41. <https://doi.org/10.1111/j.1467-9280.1996.tb00346.x>

- Barrera Linares, L. (2019). Relación género/sexo y masculino inclusivo plural en español. *Literatura y Lingüística*, 40, 327–354. <http://dx.doi.org/10.29344/0717621x.40.2070>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.3929/ethz-b-000105397>
- Boyce, V., Futrell, R., & Levy, R. P. (2020). Maze Made Easy: Better and easier measurement of incremental processing difficulty. *Journal of Memory and Language*, 111, 104082. <https://doi.org/10.1016/j.jml.2019.104082>
- Brauer, M. (2008). Un ministre peut-il tomber enceinte? L’impact du générique masculin sur les représentations mentales. *L’Année Psychologique*, 108(2), 243–272. <http://doi.org/10.4074/S0003503308002030>
- Bürki, A., van den Hoven, E., Schiller, N., & Dimitrov, N. (2023). Cross-linguistic differences in gender congruency effects: Evidence from meta-analyses. *Journal of Memory and Language*, 131, 104428. <https://doi.org/10.1016/j.jml.2023.104428>
- Cabeza Pereiro, M. C., & Rodríguez Barcia, S. (2013). Aspectos ideológicos, gramaticales y léxicos del sexismo lingüístico. *Estudios Filológicos*, 52, 7–27. <http://dx.doi.org/10.4067/S0071-17132013000200001>
- Cacciari, C., & Padovani, R. (2007). Further evidence of gender stereotype priming in language: Semantic facilitation and inhibition in Italian role nouns. *Applied Psycholinguistics*, 28(2), 277–293. <http://doi.org/10.1017/S0142716407070142>
- Canal, P., Garnham, A., & Oakhill, J. (2015). Beyond gender stereotypes in language comprehension: Self sex-role descriptions affect the brain’s potentials associated with agreement processing. *Frontiers in Psychology*, 6, 1953. <https://doi.org/10.3389/fpsyg.2015.01953>
- Carreiras, M., Garnham, A., Oakhill, J., & Cain, K. (1996). The use of stereotypical gender information in constructing a mental model: Evidence from English and Spanish. *The Quarterly Journal of Experimental Psychology Section A*, 49(3), 639–663. <http://dx.doi.org/10.1080/713755647>
- Casado, A., Palma, A., & Paolieri, D. (2021). The scope of grammatical gender in Spanish: Transference to the conceptual level. *Acta Psychologica*, 218, 103361. <https://doi.org/10.1016/j.actpsy.2021.103361>
- Casado A, Sá-Leite, A. R., Pesciarelli, F., & Paolieri, D. (2023). Exploring the nature of the gender-congruency effect: implicit gender activation and social bias. *Frontiers in Psychology*, 14, 1160836. <https://doi.org/10.3389/fpsyg.2023.1160836>
- Corbett, G. G. (1991). *Gender*. Cambridge University Press.

- Cousineau, D., & Chartier, S. (2010). Outliers detection and treatment: a review. *International Journal of Psychological Research*, 3(1), 58–67. <https://doi.org/10.21500/20112084.844>
- Cubelli, R., Lotto, L., Paolieri, D., Girelli, M., & Job, R. (2005). Grammatical gender is selected in bare noun production: Evidence from the picture–word interference paradigm. *Journal of Memory and Language*, 53(1), 42–59. <https://doi.org/10.1016/j.jml.2005.02.007>
- Dixon, J. (1987). The question of genres. In I. Reid (ed.), *The place of genre in learning: Current debates* (pp. 9.21). Typereader Publications.
- Drummond, A. (2013). *Ibex farm*. Online server: <http://spellout.net/ibexfarm>
- Duffy, S. A., & Keir, J. A. (2004). Violating stereotypes: Eye movements and comprehension processes when text conflicts with world knowledge. *Memory & Cognition*, 32(4), 551–559. <https://doi.org/10.3758/BF03195846>
- Eagly, A. H., & Kite, M. E. (1987). Are stereotypes of nationalities applied to both women and men? *Journal of Personality and Social Psychology*, 53(3), 451–462. <https://doi.org/10.1037/0022-3514.53.3.451>
- Fábregas, A. (2024). *The fine-grained structure of the lexical area*. John Benjamins.
- Finocchiaro, C. Alario, F.-X., Schiller, N., Costa, A., Miozzo, M., & Caramazza, A. (2011). Gender congruency goes Europe: A cross-linguistic study of the gender congruency effect in Romance and Germanic languages. *Italian Journal of Linguistics / Rivista di linguistica*, 23(2), 161–198.
- Forster, K. I., Guerrera, C., & Elliot, L. (2009). The maze task: Measuring forced incremental sentence processing time. *Behavior Research Methods*, 41, 163–171. <https://doi.org/10.3758/BRM.41.1.163>
- Gygax, P. M., Elmiger, D., Zufferey, S., Garnham, A., Sczesny, S., Von Stockhausen, L., Braun, F., & Oakhill, J. (2019). A language index of grammatical gender dimensions to study the impact of grammatical gender on the way we perceive women and men. *Frontiers in Psychology*, 10, 1604. <https://doi.org/10.3389/fpsyg.2019.01604>
- Hartley, J. T., Stojack, C. C., Mushaney, T. J., Annon, T. A., & Lee, D. W. (1994). Reading speed and prose memory in older and younger adults. *Psychology and Aging*, 9(2), 216. <https://doi.org/10.1037/0882-7974.9.2.216>
- Hellinger, M., & Bußmann, H. (2001). *Gender across languages: The linguistic representation of women and men* Vol. 1. John Benjamins.
- Huettig, F. (2015). Four central questions about prediction in language processing. *Brain research*, 1626, 118–135. <https://doi.org/10.1016/j.brainres.2015.02.014>

- Kamide, Y., Scheepers, C., & Altamann, G. (2003). Integration of Syntactic and Semantic Information in Predictive Processing: Cross-Linguistic Evidence from German and English. *Journal of Psycholinguistic Research*, 32(1), 37–55. <https://doi.org/10.1023/A:1021933015362>
- Karimi, H., & Ferreira, F. (2016). Good-enough linguistic representations and online cognitive equilibrium in language processing. *Quarterly Journal of Experimental Psychology*, 69(5), 1013–1040. <https://doi.org/10.1080/17470218.2015.1053951>
- Kreiner, H., Sturt, P., & Garrod, S. (2008). Processing definitional and stereotypical gender in reference resolution: Evidence from eye-movements. *Journal of Memory and Language*, 58(2), 239–261. <https://doi.org/10.1016/j.jml.2007.09.003>
- Kutas, M., Van Petten, C. K., & Kluender, R. (2006). Psycholinguistics electrified II (1994–2005). In *Handbook of Psycholinguistics* (pp. 659-724). Academic Press.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, 82(13), 1-26. <https://doi.org/10.18637/jss.v082.i13>
- Leaper, C. (2014). Gender similarities and differences in language. In T. M. Holtgraves (ed.), *The Oxford Handbook of Language and Social Psychology* (pp. 62-81). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199838639.013.002>
- Levelt, W. J. (1999). Models of word production. *Trends in Cognitive Sciences*, 3(6), 223–232. [https://doi.org/10.1016/S1364-6613\(99\)01319-4](https://doi.org/10.1016/S1364-6613(99)01319-4)
- Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22(1), 1–38. <https://doi.org/10.1017/S0140525X99001776>
- Lewis, M., & Lupyan, G. (2020). Gender stereotypes are reflected in the distributional structure of 25 languages. *Nature Human Behaviour*, 4(10), 1021–1028. <https://doi.org/10.1038/s41562-020-0918-6>
- López, Á. (2020). Cuando el lenguaje excluye: consideraciones sobre el lenguaje no binario indirecto. *Cuarenta Naipes. Revista de Cultura y Literatura*, 3, 295–312.
- Lüdtke, D. (2021). *sjPlot: Data Visualization for Statistics in Social Science*. R package version 2.8.16. [Accessed 30 July 2025]. <https://CRAN.R-project.org/package=sjPlot>
- Márquez, M. (2013). *Género gramatical y discurso sexista*. Síntesis.
- MendivilGiró, J. L. (2020). El masculino inclusivo en español. *Revista Española de Lingüística*, 50(1), 35–64. <http://dx.doi.org/10.31810/RSEL.50.1.2>

- Menegatti, M., & Rubini, M. (2017). Gender bias and sexism in language. In *Oxford Research Encyclopedia of Communication*. <https://doi.org/10.1093/acrefore/9780190228613.013.470>
- Metusalem, R., Kutas, M., Urbach, T. P., Hare, M., McRae, K., & Elman, J. L. (2012). Generalized event knowledge activation during online sentence comprehension. *Journal of Memory and Language*, 66(4), 545-567. <https://doi.org/10.1016/j.jml.2012.01.001>
- Misersky, J., Majid, A., & Snijders, T. M. (2019). Grammatical gender in German influences how role-nouns are interpreted: Evidence from ERPs. *Discourse Processes*, 56(8), 643-654. <http://doi.org/10.1080/0163853X.2018.1541382>
- Paolieri, D., Lotto, L., Morales, L., Bajo, T., Cubelli, R., & Job, R. (2010). Grammatical gender processing in romance languages: Evidence from bare noun production in Italian and Spanish. *European Journal of Cognitive Psychology*, 22(3), 335-347. <https://doi.org/10.1080/09541440902916803>
- Pesciarelli, F., Scorolli, C., & Cacciari, C. (2019). Neural correlates of the implicit processing of grammatical and stereotypical gender violations: A masked and unmasked priming study. *Biological Psychology*, 146, 107714. <https://doi.org/10.1016/j.biopsycho.2019.06.002>
- Pickering, M., & Traxler, M. (2000). Parsing and incremental understanding during reading. In M. Crocker, M. Pickering & C. Clifton (eds.), *Architectures and Mechanisms for Language Processing* (pp. 238-258). Cambridge University Press.
- Pinheiro, B. F. M., & Freitag, R. M. K. (2020). Estereótipos na concordância de gênero em profissões: efeitos de frequência e saliência. *Linguística*, 16(1), 85-107. <http://doi.org/10.31513/linguistica.2020.v16n1a31637>
- R Core Team (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.
- Roca, I. M. (2006). La gramática y la biología en el género del español (2da parte). *Revista Española de Lingüística*, 35(2), 397-492.
- Sá-Leite, A. R., Tomaz, Â., Hernández-Cabrera, J. A., Fraga, I., Acuña-Fariña, C., & Comesaña, M. (2022). What a transparent Romance language with a Germanic gender-determiner mapping tells us about gender retrieval: Insights from European Portuguese. *Psicologica*, 43(2), e14777. <http://doi.org/10.20350/DIGITALCSIC/14777>
- 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, 661175. <https://doi.org/10.3389/fpsyg.2021.661175>

- Schad, D., 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*, 104038. <http://doi.org/10.1016/j.jml.2019.104038>
- Serafini, L., & Pesciarelli, F. (2025). The influence of gender stereotypical primes on the neural processing of words and faces. *Social Cognitive and Affective Neuroscience*, *20*(1), nsaf031. <https://doi.org/10.1093/scan/nsaf031>
- Sikos, L., Duffield, C. J., & Kim, A. E. (2016). Grammatical predictions reveal influences of semantic attraction in online sentence comprehension: evidence from speeded forced-choice sentence continuations. *Language, Cognition and Neuroscience*, *31*(8), 1055-1073. <https://doi.org/10.1080/23273798.2016.1186808>
- Siyanova-Chanturia, A., Pesciarelli, F., & Cacciari, C. (2012). The electrophysiological underpinnings of processing gender stereotypes in language. *PloS one*, *7*(12), e48712. <https://doi.org/10.1371/journal.pone.0048712>
- Siyanova-Chanturia, A., Warren, P., Pesciarelli, F., & Cacciari, C. (2015). Gender stereotypes across the ages: On-line processing in school-age children, young and older adults. *Frontiers in Psychology*, *6*, 1388. <https://doi.org/10.3389/fpsyg.2015.01388>
- 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). <https://doi.org/10.16995/glossa.6144>
- Stetie, N. A., & Zunino, G. M. (2023). Estereotipos y morfología de género en nombres de rol: un estudio psicolingüístico. *Lexis*, *47*(2), 678–716. <https://doi.org/10.18800/lexis.202302.006>
- Thornton, A. (2009). Constraining gender assignment rules. *Language Sciences*, *31*(1), 14–32. <https://doi.org/10.1016/j.langsci.2007.09.003>
- Tzinavos Muñoz, S. M., Stetie, N. A., & Tomé Cornejo, C. (2025). Diferencias en la carga de estereotipicidad de género en nombres de rol: un estudio comparativo entre hablantes de español de Argentina y España. *Clepsydra: Revista Internacional de Estudios de Género y Teoría Feminista*, *29*, 179–200. <https://doi.org/10.25145/j.clepsydra.2025.29.08>
- Vasishth, S., & Nicenboim, B. (2016). Statistical methods for linguistic research: Foundational ideas—Part I. *Language and Linguistics Compass*, *10*(8), 349-369. <https://doi.org/10.1111/lnc3.12201>
- Vasishth, S., Schad, D., Bürki, A., & Kliegl, R. (2021). *Linear Mixed Models in Linguistics and Psychology: A Comprehensive Introduction* [Accessed 30 July 2025]. https://vasishth.github.io/Freq_CogSci/

- Venables, B. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S*, Fourth Edition. Springer. <http://doi.org/10.1007/978-0-387-21706-2>
- Wicha, N. Y. Y., Moreno, E. M., & Kutas, M. (2004). Anticipating words and their gender: an event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. *Journal of Cognitive Neuroscience*, 16(7), 1272–1288. <https://doi.org/10.1162/0898929041920487>
- Wickham, H., Averick, M., Bryan, J., Chang, W., D'Agostino McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., & Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686. <http://doi.org/10.21105/joss.01686>
- Winter, B. (2019). *Statistics for linguists: An introduction using R*. Routledge. <http://doi.org/10.4324/9781315165547>
- Wu, R., & Schiller, N. O. (2023). Gender congruency effects in Spanish: Behavioral evidence from noun phrase production. *Brain Sciences*, 13(4), 696. <https://doi.org/10.3390/brainsci13040696>
- Zehr, J., & Schwarz, F. (2018). *PennController for Internet Based Experiments (IBEX)*. <https://doi.org/10.17605/OSF.IO/MD832>
- Zunino, G. M., & Stetie, N. A. (2022). Binary or non-binary? Gender morphology in Spanish: Differences dependent on the task. *ALFA: Revista de Lingüística*, 66, e14546. <https://doi.org/10.1590/1981-5794-e14546>
- Zunino, G. M., Aguilar, M., Stetie, N. A., Martínez Rebolledo, C., & Hinojosa Podeva, J. A. (2025). Dresses and ties: the effect of grammatical gender and stereotypical semantic bias in three Spanish-speaking communities. *Language and Cognition*, 17, e35. <https://doi.org/10.1017/langcog.2025.4>
- Zunino, G. M., Gagliardi, G., Stetie, N. A., & Miola, E. (2026). Female truck drivers and male babysitters? Interactions between gender stereotypes and grammatical gender in Spanish and Italian: a psycholinguistic proposal. *Linguistik Online*, 144(3), 89-103. <https://doi.org/10.13092/q8we5311>