

UC San Diego

UC San Diego Previously Published Works

Title

Grammatical gender processing: More than meets the eye

Permalink

<https://escholarship.org/uc/item/26f3n60t>

Journal

Cognition, 272

ISSN

0010-0277

Authors

Beatty-Martínez, Anne L

Mulík, Stanislav

Bajo Molina, M Teresa

et al.

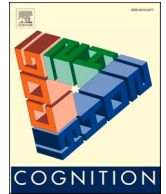
Publication Date

2026-07-01

DOI

10.1016/j.cognition.2026.106520

Peer reviewed



Brief article

Grammatical gender processing: More than meets the eye

Anne L. Beatty-Martínez^{a,*}, Stanislav Mulík^b, M. Teresa Bajo Molina^{c,d}, Paola E. Dussias^b^a Department of Cognitive Science, University of California, San Diego, La Jolla, CA 92037, United States of America^b Department of Spanish, Italian, and Portuguese, Pennsylvania State University, University Park, PA 16802, United States of America^c Mind, Brain, & Behavior Center (CIMCYC), University of Granada, Campus Cartuja s/n, 18071 Granada, Spain^d Department of Experimental Psychology, Faculty of Psychology, University of Granada, Campus Cartuja s/n, 18071 Granada, Spain

ARTICLE INFO

Keywords:

Language processing

Grammatical gender

Spanish

Pupillometry

ABSTRACT

Studies of morphosyntactic processing have shown that speakers rely on grammatical gender cues and are sensitive to gender agreement violations during language processing. However, past research has predominantly focused on words whose gender assignment (e.g., between nouns and their modifiers) is unambiguous, overlooking cases with variable gender assignment. This study takes advantage of the unique case of Spanish dual-gendered nouns, such as *agua* (water) and *área* (area). These nouns are prescriptively feminine but, due to historical processes largely unknown by most Spanish speakers, are sometimes preceded by (presumed to be) masculine determiners. A search of dual-gendered nouns on X (formerly Twitter) confirmed previous findings that Spanish speakers exceptionally alternate between feminine and masculine forms in everyday language use. Using examples extracted from the corpus search, we employed pupillometry to measure Spanish speakers' sensitivity to gender violations in both regular and dual-gendered nouns. Results showed that the pupil response was sensitive to grammatical gender violations for regular but not dual-gendered nouns. This is an important boundary condition because it suggests that there is no single processing strategy for grammatical gender. Instead, speakers may rely on context-dependent strategies that are finely attuned to distributional information. The observed differences underscore the value of studying variable phenomena using naturally-occurring data, as they provide insights into the mechanisms of language processing that might otherwise remain obscured.

1. Introduction

Research on grammatical gender contributes to the scientific understanding of analogy, our ability to perceive and use relational patterns. Analogical reasoning plays a central role in human cognition, serving as a core process in categorization and decision-making (Gentner & Smith, 2012). It has been widely recognized that speakers are sensitive to the systematic correspondences between a noun and its gender, and that grammatical gender violations result in processing costs (Beatty-Martínez & Dussias, 2019; Sá-Leite et al., 2019). Several influential psycholinguistic models treat grammatical gender as a purely symbolic, categorical feature, with each noun specified for a single discrete value (i.e., masculine, feminine). In Levell et al. (1999), for example, gender is a fixed lexical-syntactic attribute retrieved in an all-or-none manner to guide agreement (see also Schriefers, 1993). In these models, variability in agreement should arise from performance factors (e.g., retrieval error, processing noise) or from optional rule application,

rather than from the lexical representation itself. However, to the extent that gender-related variation is systematic and frequency-sensitive, it instead suggests that lexical representations may encode competing gender tendencies with different strengths rather than strictly categorical values, consistent with usage-based accounts in which frequency shapes gradient lexical representations (Bybee, 2010).

In line with this account, grammatical gender processing shows substantial inter- and intra-subject variability (Beatty-Martínez et al., 2020). In Spanish, for example, nouns are organized into masculine and feminine categories based on their phonological shape. Spanish nouns whose phonological characteristics are more strongly associated with a specific gender class (referred to in the literature as morphophonologically transparent) are typically recognized more quickly (Alarcón, 2009; Luque et al., 2023) and have been shown to elicit larger electrophysiological responses for agreement violations compared to nouns with opaque phonological shape (Caffarra & Barber, 2015). Moreover, while the two gender categories are comparable in their numerical

* Corresponding author at: 9500 Gilman Drive, Mail Code 0515, La Jolla, CA 92093, United States of America.

E-mail addresses: abeattymartinez@ucsd.edu (A.L. Beatty-Martínez), sxm6739@psu.edu (S. Mulík), mbajo@ugr.es (M.T. Bajo Molina), ped10@psu.edu (P.E. Dussias).

<https://doi.org/10.1016/j.cognition.2026.106520>

Received 24 April 2025; Received in revised form 21 February 2026; Accepted 2 March 2026

Available online 9 March 2026

0010-0277/© 2026 Elsevier B.V. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

distribution, they vary in their schematic organization (Trawick & Bero, 2021). Specifically, the feminine category is schematically constrained to a narrow range of phonological forms, whereas the masculine category has a more open schematic structure, allowing all other forms, including those of most loanwords. This difference in category schematicity tips the balance in favor of masculine gender when there is ambiguity or novelty (e.g., when referring to generic nouns, loanwords, etc.), with non-trivial consequences for language processing (Beatty-Martínez & Dussias, 2019).

The observation that Spanish speakers rarely make gender agreement errors (Montrul et al., 2014) suggests a strong relational similarity between a noun's representation and its gender category. However, this holds true only for nouns whose gender is unambiguous and whose phonological shape conforms to its gender schema. While grammatical gender processing has been extensively studied from a variety of perspectives, much of the research has focused on such nouns, excluding those whose gender can vary depending on (socio)linguistic factors (e.g., preceding context, sociogeographical influences). Thus, little is known about how analogy operates in relation to gender representations that are context-sensitive and ambiguous. Here, we capitalize on the case of dual-gendered nouns (also called hermaphroditic nouns; Eddington, 2022; Eddington and Hualde, 2008; Rini, 2016) to examine how variation in gender assignment of dual-gendered nouns affects the signatures of gender processing.

1.1. Dual-gendered nouns

In Spanish, feminine nouns beginning with a stressed /a/ (e.g., *agua* 'water'; henceforth dual-gendered nouns) pose an intriguing conundrum (Álvarez de Miranda, 1993; Rini, 2016). Prescriptive grammar holds that dual-gendered nouns are inherently feminine¹ but at the same time take the masculine definite article *el* 'the_M' where the latter directly precedes the former (i.e., *el agua* 'the_M water' and not **la agua* 'the_F water'; RAE & ASALE, 2026a, 2026b). This apparent incongruity is commonly explained based on euphony²: sequences of word-final /a/ followed by stressed word-initial /a/ are phonetically infelicitous. In reality, there is a historical parallel rooted in the evolution of the language (see Table 1). Historically, *el* and *la* derived from Latin demonstratives ILLE and ILLA. Over time, ILLE became *el*, but the feminine form divided into two phonologically conditioned allomorphs: *la* preceded consonants whereas *el* preceded vowels.³ Thus, though identical in form, the *el* used in *el agua* is not the masculine definite article, but a

Table 1
Evolution of *el* and *la* from Latin to Modern Spanish.

Gender	LATIN			Modern Spanish	Translation
masculine	ILLE REX	>	<i>el(e) rei</i>	>	<i>el rey</i> the king
feminine	ILLA	>	<i>(e)la</i>	>	<i>la casa</i> the house
	CASA		<i>casa</i>		
dual-gendered	ILLA	>	<i>el(a)</i>	>	<i>el agua</i> the water
	AQUA		<i>agua</i>		

Note: The preference of *el* over *la* before dual-gendered nouns can be traced as far back as the 13th century (Eddington, 2022).

¹ This is generally evidenced in the use of diminutives (*agüita*, 'little water_F'), accompanying modifiers, and anaphoric reference (e.g., *Esta agua está rica. Me la voy a beber entera.* 'This_F water is delicious_F. I will drink it_F all_F.').

² Examples of present-day internet discussions are listed in Appendix A.

³ In medieval Spanish, feminine-*el* had a wider distribution and was used even before unstressed vowels (e.g., *el harina* 'the flour_F', *el espada* 'the sword_F'). Over time, its use became restricted to only occurring before feminine nouns beginning with stressed /a/ (Real Academia Española and Asociación de Academias de la Lengua Española, 2026b; Álvarez de Miranda, 2014).

coincidentally homophonous product of diachrony.

A second complicating factor is that the masculine indefinite article *un* 'a(n)_M' and, by extension, *algún* 'some_M/any_M' and *ningún* 'not any_M' also appear before dual-gendered nouns, albeit for different reasons. According to the Royal Spanish Academy (RAE & ASALE, 2026a, 2026c, 2026d), these forms are apocopated allomorphs of feminine forms (i.e., $(-)\textit{una} + \textit{área} > (-)\textit{un} + \textit{área}$) that again happen to be identical to their masculine counterparts. To summarize, the stressed /a/ in dual-gendered nouns triggers a rare use of feminine allomorphs, which are homophonous with their masculine forms. Prescriptively, these allomorphs occur only in cases where the determiner directly precedes the noun, and are thus considered to be the exception, not the rule.⁴

In principle, these criteria seem straightforward; in practice, they result in a complex network of analogical relations. The fact is that most Spanish speakers today are largely unaware of the etymology of feminine *el*, *un*, and other varieties of indefinite forms. This leads some to believe that *el* and *un* are truly masculine forms, prompting the interpretation of sequences such as *el agua* 'the_M water' or *un arpa* 'a_M harp' as showing masculine agreement (Álvarez de Miranda, 1993, 2014). The result is that the use of masculine gender determiners is extended to contexts where the prescriptive rule does not apply, most notably with demonstratives (e.g., *este agua* 'this_M water'), possessives (e.g., *nuestro agua* 'our_M water'), and in cases with intervening material (e.g., *el refrescante agua* 'the refreshing water'). In other words, the use of *el agua* and *un agua* has opened the gateway for other masculine modifiers through analogical influence.

Spanish sources are replete with instances of prescriptively incorrect gender agreement (see Table 2). This variability is observed in different varieties of Spanish and across different registers (e.g., informal speech, advertisements, media, and even scholarly works; Álvarez de Miranda, 2014; Eddington & Hualde, 2008; RAE & ASALE, 2026b; see Fig. 1).

Behavioral studies have shown that, for dual-gendered nouns, Spanish speakers are more likely to accept non-prescriptive forms, particularly those with prenominal masculine modifiers (Eddington & Hualde, 2008). Indeed, when asked to identify the gender of a bare dual-gendered noun, Spanish speakers generally assign them masculine gender (Eddington, 2022). Nonetheless, at a macro-level, prescriptive correct forms outweigh non-prescriptive forms (see Fig. 2). It is also true that not all modifiers associated with dual-gendered nouns display

Table 2
Examples of prescriptively correct and incorrect agreement with dual-gendered nouns.

Construction	Correct	Incorrect/Unexpected	Translation
Def + N	<i>el agua</i>	<i>*la agua</i>	the water
Indef + N	<i>un ala</i>	<i>**una ala</i>	a wing
Dem + N	<i>aquella área</i>	<i>*aquel área</i>	that area
Poss + N	<i>nuestra alma</i>	<i>*nuestro alma</i>	our soul
Def + Mod + N	<i>la rara águila</i>	<i>*el raro águila</i>	the rare eagle
Indef + Mod + N	<i>una vieja aula</i>	<i>*un viejo aula</i>	an old classroom

Abbreviations: Def, definites; Dem, demonstratives; F, feminine gender; Indef, indefinites; M, masculine gender; Mod, modifiers; N, noun; Poss, possessives; "**" indicates prescriptively incorrect gender agreement. **Note, while the use of *una*, *alguna*, and *ninguna* is considered prescriptively correct (RAE & ASALE, 2026a, 2026c), these forms are relatively less common than their apocopated counterparts *un*, *algún*, *ningún*. For the purposes of this study, we treat such instances as violations of a strong expectation.

⁴ The stressed /a/ in dual-gendered nouns does not affect plural articles (i.e., not **los aguas* 'the_{M-PL} waters' but *las aguas* 'the_{F-PL} waters') or any accompanying modifiers (e.g., *el agua fría* 'the_M cold_F water'; see also Eddington & Hualde, 2008; Real Academia Española and Asociación de Academias de la Lengua Española, 2026d).

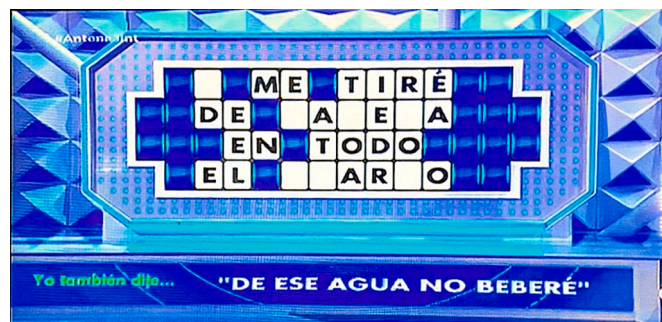


Fig. 1. A photo of a Spanish game show where the key phrase *de *ese agua no beberé* ‘I will not drink from *that_M water’ features a prescriptively incorrect use of masculine agreement in an idiomatic expression.

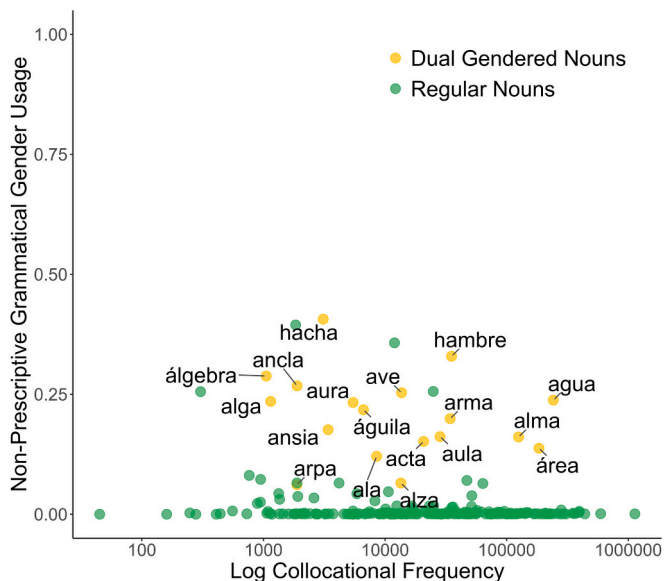


Fig. 2. Proportion of non-prescriptive grammatical gender usage in dual-gendered nouns (yellow) and regular nouns (green). Proportions were calculated from the determiner-noun collocational frequency values extracted from the Corpus del Español (Davies, 2016). Despite broadly overlapping frequency distributions, the two noun groups differ in their rates non-prescriptive usage. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

similar rates of gender marking variability. As a case in point, dual-gendered nouns exhibit what has been referred to as number sensitive lateral hermaphroditism, with greater gender vacillations⁵ on prenominal modifiers, but a stronger proclivity toward feminine gender assignment in postnominal modifiers. These results suggest that the processing of grammatical gender is a far more fluid and context-sensitive mechanism than previously thought. To our knowledge, no study to date has investigated the implications of the differential behavior of regular and dual-gendered nouns for the processing of gender violations.

1.2. The present study

We use pupillometry to investigate grammatical gender processing in

⁵ We use vacillations to denote greater variety in gender assignment overall (see, Poplack et al., 1982). Of relevance, non-prescriptive forms with the article *la* (e.g., **la agua*,) also occur at a non-trivial frequency (Álvarez de Miranda, 1993).

native Spanish speakers. Task-evoked changes in pupil size have been shown to reliably reflect attentional effort and processing difficulty in a variety of linguistic tasks (e.g., Beatty-Martínez et al., 2021; López-Beltrán & Dussias, 2023). An increasing body of research suggests that larger pupil responses index a discrepancy between expected and actual outcomes. We extend existing evidence by examining pupil responses to determiner-noun gender violations, focusing on dual-gendered nouns, which remain largely underrepresented in models of grammatical gender processing (e.g., Levelt et al., 1999; Schriefers, 1993). Importantly, our approach distinguishes between two distributional dimensions illustrated in Fig. 2: regular and dual-gendered nouns are comparable in overall determiner-noun collocational frequency, yet they differ markedly in the reliability with which they map onto a particular gender form. This distinction allows us to isolate the role of variability in gender assignment as a driver of processing outcomes, over and above lexical frequency.

Because not all violations stem from the same source, we further differentiate between two determiner categories: (1) definites (*el/la*), where an error constitutes failure to apply a phonological rule (e.g., **la agua*) and indefinites (*un/una*, *algún/alguna*, *ningún/ninguna*), where both forms are prescriptively admissible, but the feminine form is considerably less frequent and less expected; and (2) demonstratives (*ese/esa*, *este/esta*, *aquel/aquella*) and possessives (*nuestro/nuestra*), where an error constitutes extension of the phonological rule to contexts where it does not apply (e.g., **ese agua*). If processing strategies are modulated by analogy to distributional cues based on prior usage experience (Bybee, 2010), we hypothesized that gender violations would elicit larger pupil responses for regular but not for dual-gendered nouns. For dual-gendered nouns, participants may not interpret gender mismatches as true errors but as expected variation. Such a pattern would indicate that gender agreement processing is sensitive to the reliability of distributional mappings and that variability in gender assignment constitutes a meaningful input to the processing system.

2. Method

2.1. Participants

We tested 68 Spanish speakers from the University of Granada, Spain, all of whom gave informed consent. Procedures were approved by the Pennsylvania State University Institutional Review Board. Participants were paid 10€ per hour. Data from 11 participants were excluded from further analysis per pupillometry exclusion criteria, which resulted in a total of 57 participants (40 female). All participants reported being functionally monolingual in Spanish (see Table 3 for summary of participant characteristics).

2.2. Materials and procedure

2.2.1. Stimuli

Experimental stimuli followed a 2 (Noun type: Regular or Dual-Gendered) x 2 (Grammaticality: Correct or Incorrect) design. Dual-

Table 3
Participant self-reported characteristics.

Measure	M	SD	95% CI
Age, years	23.19	5.67	[21.69, 24.69]
Spanish speaking (/10)	9.82	0.56	[9.67, 9.97]
Spanish understanding (/10)	9.74	0.62	[9.57, 9.90]
Spanish reading (/10)	9.83	0.46	[9.71, 9.95]
Spanish writing (/10)	9.74	0.61	[9.58, 9.90]

Note: Means, standard deviations, and 95% confidence intervals for age and proficiency self-ratings. Proficiency self-ratings were on a 10-point scale ranging from 1 (not proficient) to 10 (highly proficient). All values represent raw, non-standardized scores. One participant did not complete the proficiency self-ratings.

gendered nouns were compiled from Eddington (2022) and Spanish grammar books, and experimental stimuli were curated using real-world examples from social media posts from X (formerly Twitter) in prescriptively incorrect contexts.⁶ Nouns with biological gender, loanwords, regionalisms, restricted polysemy, idiomatic uses, or homonyms were excluded, leaving 18 dual-gendered nouns (Appendix B). Tweets were minimally edited for clarity and length. For each edited stimulus, we selected a regular noun alternative that was contextually plausible (see Table 4).

We created 224 trial quartets distributed across four experimental lists in Latin square design. Half of the quartets contained determiners for which masculine use is prescriptively accepted (definites: *el/la* and indefinites: *un/una, algún/alguna, ningún/ninguna*). The other half contained determiners for which masculine use is prescriptively incorrect (demonstratives: *ese/esa, este/esta, aquel/aquella* and possessives: *nuestro/nuestra*). Determiner types were equally distributed across conditions. Trials within each list were pseudo-randomized so that no more than two consecutive items from the same condition. To ensure that participants paid attention, 25% of trials were followed by a yes-no comprehension question. Stimuli were recorded by a female speaker from Andalucía using a Shure SM35 condenser microphone and normalized for intensity. The onset and offsets for target determiners and nouns were identified and hand-coded using both auditory cues and visual inspection of sound spectrographs in Praat (Boersma & Weenink, 2023).

2.2.2. Procedure

The experiment lasted 2 h and took place in a quiet room. The session began with informed consent followed by the pupillometry experiment. Participants were seated with their head stabilized on a chinrest approximately 65 cm from the monitor. Pupil data were recorded from the right eye at 1000 Hz using an Eyelink Portable Duo eye-tracker (SR Research Data Ltd., 2004–2015). A standard nine-point calibration and

validation procedure was performed at the start of the experiment, requiring fixation on a black dot appearing randomly on a 3 × 3 grid. On each trial, an ear icon appeared, followed by an auditory sentence. Afterward, a mouth icon cued participants to repeat the sentence aloud (see Fig. 3 for trial procedure). The experiment consisted of two counterbalanced blocks separated by a short break. Every block began with two practice trials followed by 112 experimental trials, with calibration repeated at the start of each block and as needed throughout the experiment.

2.3. Pupillometry analysis

2.3.1. Data processing

Pupil data were extracted using EyeLink Data Viewer v3.1.246 (SR Research). The interest period began 500 ms prior to the target noun and lasted 5000 ms. Data were downsampled to 50 Hz and binned in 20 ms intervals. Baseline correction was applied by subtracting the mean pupil size in the 500 ms pre-target period from each time bin during the 4500 ms target period (van Rij et al., 2019). Trials with >35% samples lost to blinks, saccades, or off-area fixations in either the baseline or target interest periods were excluded. Participants with <50% valid trials were excluded, leaving 9464 trials from 57 participants.

2.4. Modeling

Normalized pupil size was modeled as a time-dependent variable using generalized additive mixed-effects models (GAMMs) in R (v.4.4.2; R Core Team, 2024). Unlike linear mixed-effects models, GAMMs can capture non-linear patterns in the data by incorporating smooth functions built from sets of basis functions. This feature is advantageous for pupillometry, where responses often unfold nonlinearly over time. GAMMs enable the examination of non-linear interactions between the pupil response and experimental variables, while also capturing non-linear random effects and addressing autocorrelation inherent in time-series measurements. Models were fitted using the *bam* function (*mgcv* v.1.9–1; Wood, 2017) with *fREML* estimation and discretized covariates. To improve model accuracy, fixed smooth terms used thin plate regression splines, while random smooth terms used cubic regression splines (Sóskuthy, 2021). Model diagnostics were performed with *mgcv::gam.check* and the *itsadug* package (v2.4, van Rij et al., 2020), and autocorrelation was addressed using a first-order AR(1) error model ($\rho = 0.95$; Wieling, 2018). Sudden shifts in pupil size due to gaze position were accounted for via an isometric non-linear interaction between X- and Y-gaze coordinates (van Rij et al., 2019). Random effects included by-participant and by-determiner factor smooths, allowing the shape of the average time course to vary across participants and determiners. Model specifications and output summaries are available in the Open Science Framework repository (see Data availability statement).

3. Results

Model 1 examined pupil response trajectories to gender violations in regular and dual-gendered nouns. To distinguish the grammaticality effects for the two noun groups, we constructed two ordered factors: *regular_incorrect_ord*, which was set to “TRUE” for violations involving regular nouns and “FALSE” otherwise, and *dual_incorrect_ord*, which was set to “TRUE” for violations involving dual-gendered nouns and “FALSE” otherwise. This factorization allowed the model to assess differences in both intercepts (i.e., curve height) and smooth functions (i.e., curve shape) over time for the two noun groups. As shown in Table 5, parametric coefficients and difference smooths for Model 1 both indicate a significant grammaticality effect for regular but not dual-gendered nouns. Fig. 4 shows this interaction: violations had larger pupil responses compared to non-violations for regular nouns (left panel; significant time window indicated by the vertical red lines) but not for

Table 4
Example stimuli.

Exception determiners (definites and indefinites)	
Original tweet	<i>pero si parece señora esperando que se caliente la agua para el cafe.</i>
Edited stimulus	<i>La señora tuvo que calentar *la/el agua para el café.</i>
Regular noun	<i>La señora tuvo que calentar la/*el leche para el café.</i>
English translation	The woman had to heat the _{F/M} water _{DG} /milk _R for the coffee.
Standard determiners (demonstratives and possessives)	
Original tweet	<i>—Lo primero que hizo al entrar en aquel aula... ¿se podría considerar aula?, fue mirar a su alrededor para ver si estaba en sintonía con el resto de la clase. Si había que llevar algún libro especial, cuaderno para tomar notas o si había + .</i>
Edited stimulus	<i>Lo primero que hizo al entrar en *aquel/aquella aula fue saludar a sus alumnos.</i>
Regular noun	<i>Lo primero que hizo al entrar en aquel/*aquella salón fue saludar a sus alumnos.</i>
English translation	The first thing he did when he entered that _{M/F} classroom _{DG} /room _R was greet his students.

Abbreviations: F, feminine gender; M, masculine gender; DG, dual-gendered; R, regular; “*” indicates prescriptively incorrect gender agreement. The complete list of experimental stimuli is available via the Open Science Framework platform (see Data availability statement).

⁶ Keywords searches for each dual-gendered noun included feminine marked definite article (*la*) and indefinites (*una, ninguna, alguna*) and masculine marked demonstratives (*este, ese, aquel*) and possessives (*nuestro*).

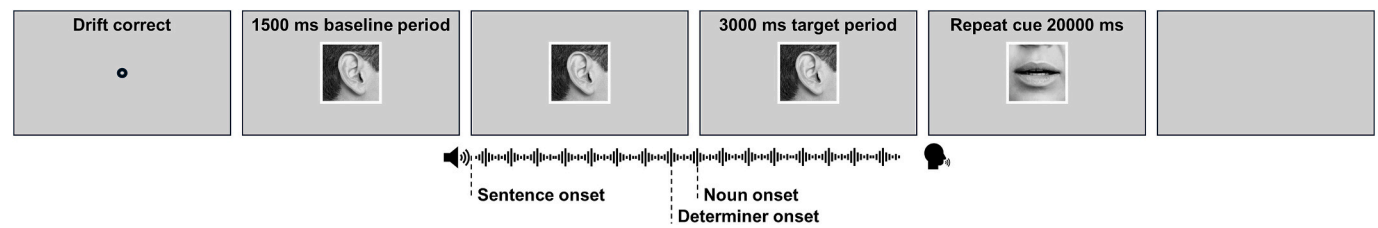


Fig. 3. Trial Procedure. Each trial began with a fixation point, followed by a 1500 ms baseline period. During the baseline, no audio was played, and an image of an ear was displayed in a central 300 × 300 pixel interest area. After the baseline period, the stimulus sentence was presented. The target period began at the noun onset and lasted for 3000 ms. The ear icon remained on screen throughout both the baseline and target periods, maintaining constant luminance. Following the target period, the display showed a mouth icon, cueing participants to repeat the sentence aloud. The mouth remained on screen until participants pressed a button or until 20,000 ms elapsed.

Table 5
Ordered-Factor Difference Smooth Model 1 summary: Parametric coefficients and approximate significance of smooth terms: estimated degrees of freedom (edf), reference degrees of freedom (Ref.df), *F*- and *p*-values for smooth terms.

Parametric coefficients	Estimate	SE	<i>t</i> -value	<i>p</i> -value
(Intercept)	−12.15	2.15	−5.65	<0.001***
noun_groupdual	−0.08	0.90	−0.09	0.926
regular_incorrect_ordTRUE	1.84	0.90	2.04	0.041*
dual_incorrect_ordTRUE	−0.90	0.89	−1.01	0.311
Smooth terms	Edf	Ref.df	<i>F</i> -value	<i>p</i> -value
s(bin):noun_groupregular	14.19	18.65	2.69	<0.001***
s(bin):noun_groupdual	13.70	18.19	2.40	<0.001***
s(bin):	16.89	22.12	4.70	<0.001***
regular_incorrect_ordTRUE				
s(bin):dual_incorrect_ordTRUE	1.05	1.08	0.74	0.382
s(bin,participant)	634.90	2280.00	8.24	<0.001***
s(bin,det)	41.30	159.00	1.41	<0.001***
s(gaze_x,gaze_y)	28.95	29.00	3404.62	<0.001***

Note: The rows **regular_incorrect_ordTRUE** (parametric coefficient) and **s(bin):regular_incorrect_ordTRUE** (smooth term) correspond to the grammaticality effect for regular nouns, reflecting differences in overall pupil response size (curve height) and response over time (curve shape), respectively.

dual-gendered nouns (right panel). To further confirm these results, we conducted a follow-up analysis combining the intercept and non-linear differences using a binary difference smooth approach (Wieling, 2018). This confirmatory test showed that pupillary responses were significantly elevated for gender violations involving regular nouns (*Edf* = 17.90, *F* = 4.71, *p* < .001), whereas violations involving dual-gendered nouns had no effect on pupil size (*Edf* = 2.05, *F* = 0.64, *p* =

.535). The complete binary difference approach model summary is available in the OSF repository.

One possible explanation for the lack of sensitivity to gender violations involving dual-gendered nouns is that not all determiners follow the same grammatical rules. As noted in the introduction, although indefinite determiners typically take masculine forms when paired with dual-gendered nouns, both genders are prescriptively admissible (RAE & ASALE, 2026c). To address this possibility, we reran the analysis excluding indefinite determiners. The results remained unchanged (see Table S1 in the Open Science Framework repository for model summary).

To further rule out the possible effects of collapsing across different kinds of determiners, we conducted a second analysis focusing exclusively on dual-gendered nouns. Recall that the Royal Spanish Academy does not endorse the usage of masculine demonstrative and possessive determiners. Thus, our second analysis set out to evaluate grammaticality effects for standard determiners (i.e., demonstratives and possessives) versus exception determiners (i.e., definites and indefinites). For Model 2, we constructed two ordered factors: *standard_incorrect_ord*, which is set to “TRUE” for violations involving demonstratives and possessives and “FALSE” otherwise, and *exception_incorrect_ord*, which is set to “TRUE” for violations with definites and indefinites and “FALSE” otherwise. All other parameters were kept the same as in Model 1. The estimated *p*-values for the parametric and smooth difference terms are displayed in Table 6. As shown in Fig. 5, we found that the violations involving dual-gendered nouns had no effect on pupil size for any of the determiner groups. A follow-up binary difference model confirmed this result: (standard determiners: *Edf* = 2.02, *F* = 0.21, *p* = .825; exception determiners: *Edf* = 2.07, *F* = 0.19, *p* = .799; see OSF repository for model summary).

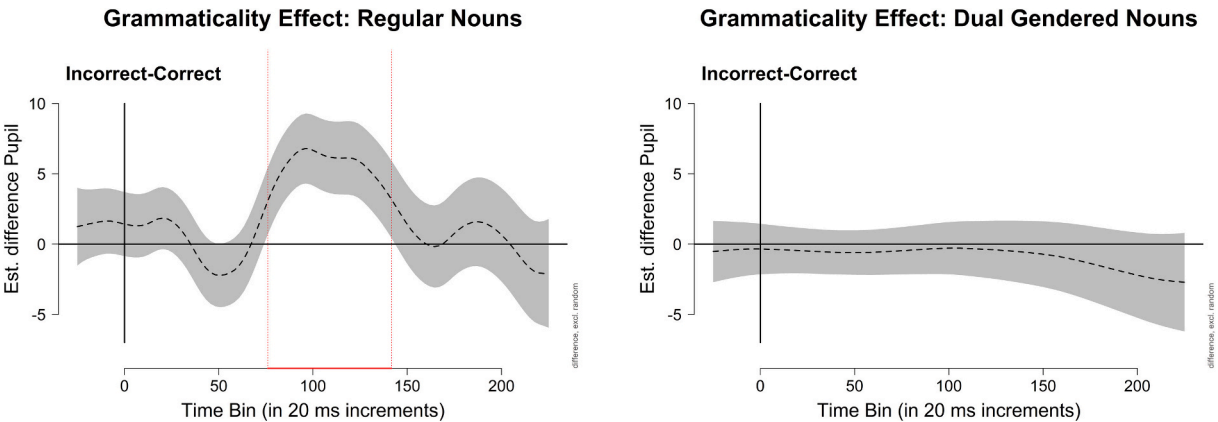


Fig. 4. Model 1 difference smooths representing the estimated grammaticality effect (incorrect-correct) for regular nouns (left) and dual-gendered nouns (right). Shading represents pointwise 95% confidence intervals. The vertical red dotted lines indicate the time window (bins 76–141; 1520–2820 ms after target noun onset) in which the difference is significantly different from 0. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 6

Ordered-Factor Difference Smooth Model 2 summary: Parametric coefficients and approximate significance of smooth terms: estimated degrees of freedom (edf), reference degrees of freedom (Ref.df), *F*- and *p*-values for smooth terms.

Parametric coefficients	Estimate	SE	<i>t</i> -value	<i>p</i> -value
(Intercept)	−14.87	2.35	−6.32	<0.001***
det_groupeexception	−0.30	1.43	−0.21	0.840
standard_incorrect_ordTRUE	−0.77	1.43	−0.54	0.590
exception_incorrect_ordTRUE	−0.57	1.423	−0.40	0.690
Smooth terms	Edf	Ref.df	<i>F</i> -value	<i>p</i> -value
s(bin):det_groupstandard	13.53	18.09	2.07	0.004**
s(bin):det_groupeexception	9.03	12.19	2.23	0.008**
s(bin):standard_incorrect_ordTRUE	1.02	1.04	0.44	0.560
s(bin):exception_incorrect_ordTRUE	1.05	1.08	0.23	0.538
s(bin,participant)	462.58	2280.00	4.36	<0.001***
s(bin,det)	39.50	158.00	0.99	<0.001***
s(gaze_x,gaze_y)	28.88	29.00	1782.06	<0.001***

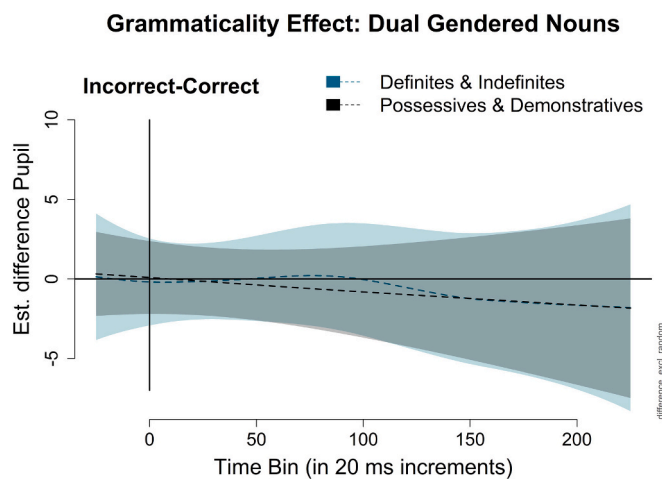


Fig. 5. Model 2 difference smooths representing the estimated grammaticality effect (incorrect-correct) for standard determiners (i.e., possessives and demonstratives; black) and exception determiners (i.e., definites & indefinites; teal) Shading represents pointwise 95% confidence intervals. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4. Discussion

Grammatical gender processing has been studied using a variety of experimental methods, but prior work largely focused on nouns with clear and unambiguous gender, relegating exceptions to the periphery. Here, we examined the case of Spanish dual-gendered nouns, using pupillometry as an index of processing violations. While determiners reliably mark gender of most nouns (>99%; Franck et al., 2008), gender-matching between determiners and dual-gendered nouns is considerably less consistent in real-world language. If processing relies on analogy to distributional cues from prior experience, we expected speakers to be differentially sensitive to violations for regular and dual-gendered nouns. Consistent with this prediction, pupil responses were sensitive to violations for regular but not for dual-gendered nouns. For dual-gendered nouns, both error types (i.e., failure to apply the phonological rule and its overextension, differentiated by determiner types) were similarly insensitive to grammaticality manipulations. Importantly, this difference emerged despite similar collocational frequencies (Fig. 2), indicating that pupil responses reflect variability in gender usage rather than lexical frequency per se.

Overall, these findings reveal an important boundary condition,

indicating that grammatical gender processing cannot be captured by a single, uniform approach. Instead, the evidence suggests that speakers and listeners draw on multiple, sometimes competing, sources of information—including lexical entries, phonological forms, and distributional patterns based on usage experience. In production, this competition surfaces as variability, especially in cases where lexical and phonological cues diverge, as with dual-gender nouns. In comprehension, the system appears more flexible and adaptive, integrating surface-level variability while maintaining stable underlying representations. This asymmetry highlights the extent to which grammatical gender processing is shaped by domain-specific pressures: whereas production highlights the competition between lexical and phonological information, comprehension integrates these cues into a coherent representation.

More broadly, the results support accounts of language processing that emphasize flexibility, adaptation, and the interaction of multiple representational levels. While the field is increasingly attentive to variability, uniform-processing assumptions continue to shape how evidence is selected, modeled, and interpreted (see Beatty-Martínez & Dussias, 2026). As a case in point, our findings bear directly on established accounts emphasizing markedness and noun transparency (Alemán Bañón & Rothman, 2016; Caffarra & Barber, 2015). Stronger violation effects are often reported for feminine nouns (the marked category) and for morphologically transparent nouns, where the form-to-gender mapping (e.g., *—a* for feminine) is highly reliable. Dual-gendered nouns meet both criteria: they are prescriptively feminine and morphophonologically transparent. By these predictions, they should therefore pattern with regular feminine nouns and elicit robust violation effects. Yet they do not. This boundary condition indicates that surface gender category and morphophonological transparency alone are insufficient to fully predict processing outcomes when distributional variability in usage is high. Examining dual-gendered nouns directly shows that patterns appearing to “deviate” from benchmark predictions can instead be understood as principled adaptations to distributional experience, highlighting the limits of conclusions drawn from more uniform noun sets. Together, the differences observed in this study underscore the value of examining variable phenomena, which can help illuminate otherwise hidden aspects of the architecture and mechanisms of gender processing. These findings further suggest that gender processing is not a monolithic phenomenon but a system that flexibly adapts to distributional experience. Specifically, in the absence of reliable distributional cues as with dual-gendered nouns, speakers appear to treat variability in gender assignment not as true violations but as expected variation, guided by analogy to prior usage experience and adjusting their processing strategies accordingly.

To our knowledge, this is the first study to show task-evoked pupillary effects based on grammatical gender violations. Methodologically, by applying pupillometry in Spanish, our work broadens the empirical base of cognitive science research, which has historically focused on English, and demonstrates the utility of continuous physiological measures for morphosyntactic processing. Pupillometry is relatively portable, cost-effective, and well-suited for spoken language stimuli, affording greater flexibility in experimental design and making it especially suitable for data collection outside the laboratory setting (Beatty-Martínez & Dussias, 2026). A second aspect of this research was to increase the ecological validity of the experimental data by curating materials from naturally-occurring data. To this end, our dataset of X (Twitter) posts corroborated previous observations that prescriptively incorrect uses of dual-gendered nouns are common in real-world usage. Taken together, the study highlights that language variation is not just simply noise, but a window into the fundamental mechanisms of language processing strategies. By studying variable phenomena such as dual-gendered nouns, we gain a deeper understanding of how speakers interpret and adapt to distributional information in natural language contexts.

CRediT authorship contribution statement

Anne L. Beatty-Martínez: Writing – review & editing, Writing – original draft, Visualization, Supervision, Project administration, Methodology, Formal analysis, Data curation, Conceptualization. **Stanislav Mulík:** Writing – review & editing, Visualization, Supervision, Project administration, Data curation. **M. Teresa Bajo Molina:** Supervision, Resources, Funding acquisition. **Paola E. Dussias:** Writing – review & editing, Supervision, Software, Resources, Project administration, Funding acquisition, Conceptualization.

Funding sources

This work was supported by the National Science Foundation grants

NRT DGE-2125865, OISE-1545900, to P.E.D., and Spanish Ministry of Economy and Competitiveness grant PID2024-159774NB-I00 to M.T.B.

Declaration of competing interest

None.

Acknowledgments

We are grateful to Priscila López Beltrán, Cody Laudenslager, Blanca Lizanne, Kelsey Shekore, and Lucy Taylor for assistance with material creation and data collection. We are also grateful to María del Carmen Parafita Couto for helpful discussions during the preparation of the manuscript.

Appendix A. Examples of present-day discussions on dual-gendered nouns

- 1. MosaLingua. (2023, September 15). *Spanish articles: “el agua” or “la agua” #shorts* [Video]. YouTube. <https://www.youtube.com/watch?v=Bj4kADHITaw>
- 2. theonlinespanishteacher. (2025, January 22).



Why Do We Say ‘El Agua’ Instead of ‘La Agua’? It’s all about avoiding cacofonía (cacophony)!



- [Post]. Instagram. <https://www.instagram.com/theonlinespanishteacher/reel/DFJiuXSFxn/>
- 3. Di Leva, M. (2024, September 30). *Spanish Mystery Solved: Why Is it EL Agua and Not LA Agua?*. Easy Argentine Spanish. <https://easyargentinestpanish.com/definitearticles/>

Appendix B. Dual-gendered noun stimuli

Spanish	English Translation	Spanish	English Translation
acta	certificate/agreement	ansia	longing/anxiety
agua	water	área	area
águila	eagle	arma	weapon
ala	wing	arpa	harp
alga	seaweed	aula	classroom
álgebra	algebra	aura	aura
alma	soul	ave	bird
alza	rise	hacha	axe
ancla	anchor	hambre	hunger

Appendix C. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.cognition.2026.106520>.

Data availability

All experimental materials and data are available via OSF at https://osf.io/pumrj/?view_only=bef3e9ca03e647ffb397de273cac66e4.

References

Alarcón, I. (2009). The processing of gender agreement in L1 and L2 Spanish: Evidence from reaction time data. *Hispania*, 92, 814–828.

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, 1273–1298. <https://doi.org/10.1080/23273798.2016.1218032>

Álvarez de Miranda, P. (2014). *Quedarse solo (1). Rinconete*. Centro Virtual Cervantes. https://cvc.cervantes.es/el_rinconete/antiores/enero_14/10012014_01.htm.

Álvarez de Miranda, P. (1993). El alomorfo de ‘la’ y sus consecuencias. *Lingüística Española Actual*, 15, 5–44.

Beatty-Martínez, A. L., Bruni, M. R., Bajo, M. T., & Dussias, P. E. (2020). Brain potentials reveal differential processing of masculine and feminine grammatical gender in native Spanish speakers. *Psychophysiology*, 58(3), Article e13737. <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. <https://doi.org/10.3389/fpsyg.2019.00751>

Beatty-Martínez, A. L., & Dussias, P. E. (2026). Bilingualism in context: Principles for designing ecological and transdisciplinary research in language and cognitive science. In G. Klassen, & J. W. Schwieter (Eds.), *Quantitative research methods in multilingual acquisition and processing* (pp. 7–28). Benjamins.

Beatty-Martínez, A. L., Guzzardo Tamargo, R. E., & Dussias, P. E. (2021). Phasic pupillary responses reveal differential engagement of attentional control in bilingual spoken language processing. *Scientific Reports*, 11, 23474. <https://doi.org/10.1038/s41598-021-03008-1>

Boersma, P., & Weenink, D. (2023). Praat: Doing phonetics by computer (v.6.3.10) [software]. <http://www.praat.org/>.

Bybee, J. (2010). *Language, usage and cognition*. Cambridge University Press.

- Caffarra, S., & Barber, H. A. (2015). Does the ending matter? The role of gender-to-ending consistency in sentence reading. *Brain Res.*, 1605, 83–92. <https://doi.org/10.1016/j.brainres.2015.02.018>
- Davies, M. (2016). *Corpus del Español*. Web/Dialects Corpus. <https://www.corpusdelespanol.org/web-dial/>.
- Eddington, D. (2022). Processing Spanish gender in a usage-based model with special reference to dual-gendered nouns. *Lex Mente*, 17, 34–75. <https://doi.org/10.1075/ml.21011.ell>
- Eddington, D., & Hualde, J. I. (2008). El abundante agua fría: Hermaphroditic Spanish nouns. *Studies in Hispanic and Lusophone Linguistics*, 1, 5–32. <https://doi.org/10.1515/shll-2008-1004>
- Franck, J., Vigliocco, G., Antón-Méndez, I., Collina, S., & Frauenfelder, U. H. (2008). The interplay of syntax and form in sentence production: A cross-linguistic study of form effects on agreement. *Language & Cognitive Processes*, 23, 329–374.
- Gentner, D., & Smith, L. (2012). Analogical reasoning. In V. S. Ramachandran (Ed.), *Encyclopedia of human behavior* (2nd ed., pp. 130–136). Elsevier.
- Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *The Behavioral and Brain Sciences*, 22(1), 1–38. <https://doi.org/10.1017/S0140525X99001776>
- López-Beltrán, P., & Dussias, P. E. (2023). Heritage speakers' processing of the Spanish subjunctive: A pupillometric study. *Linguistic Approaches to Bilingualism*, 14, 809–855. <https://doi.org/10.1075/lab.21030.lap>
- Luque, A., Rossi, E., Kubota, M., Nakamura, M., Rosales, C., López-Rojas, C., & Rothman, J. (2023). Morphological transparency and markedness matter in heritage speaker gender processing: An EEG study. *Frontiers in Psychology*, 14, 1114464.
- Montrul, S., Davidson, J., De La Fuente, I., & Foote, R. (2014). Early language experience facilitates the processing of gender agreement in Spanish heritage speakers. *Bilingualism: Language and Cognition*, 17, 118–138. <https://doi.org/10.1017/S1366728913000114>
- Poplack, S., Pousada, A., & Sankoff, D. (1982). Competing influences on gender assignment: Variable process, stable outcome. *Lingua*, 57, 1–28.
- R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Real Academia Española & Asociación de Academias de la Lengua Española. (2026a). El artículo ante nombres femeninos comenzados por /a/ tónica. In *El buen uso del español*. <https://www.rae.es/buen-uso-español/el-artículo-ante-nombres-femeninos-comenzados-por-a-tónica>.
- Real Academia Española & Asociación de Academias de la Lengua Española. (2026b). El. In *Diccionario panhispánico de dudas*. <https://www.rae.es/dpd/el>.
- Real Academia Española & Asociación de Academias de la Lengua Española. (2026c). Uno. In *Diccionario panhispánico de dudas*. <https://www.rae.es/dpd/uno>.
- Real Academia Española & Asociación de Academias de la Lengua Española. (2026d). 9.2.1. las formas del artículo femenino. In *Nueva gramática básica de la lengua española*. <https://www.rae.es/gramática-básica/el-artículo/particularidades/las-formas-del-artículo-femenino>.
- van Rij, J., Hendriks, P., van Rijn, H., Baayen, R. H., & Wood, S. N. (2019). Analyzing the time course of pupillometric data. *Trends in Hearing*, 23, 1–22.
- Rini, J. (2016). Are some Spanish nouns truly grammatical hermaphrodites? *Zeitschrift für Romanische Philologie*, 132, 731–754. <https://doi.org/10.1515/zrp-2016-0049>
- Sá-Leite, A. R., Fraga, I., & Comesaña, M. (2019). Grammatical gender processing in bilinguals: An analytic review. *Psychonomic Bulletin & Review*, 26, 1148–1173. <https://doi.org/10.3758/s13423-019-01596-8>
- Schriefers, H. (1993). Syntactic processes in the production of noun phrases. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 841–850. <https://doi.org/10.1037/0278-7393.19.4.841>
- Sóskuthy, M. (2021). Evaluating generalised additive mixed modelling strategies for dynamic speech analysis. *Journal of Phonetics*, 84, Article 101017.
- SR Research Data Ltd.. (2004–2015). *SR research data viewer (v.4.1.63) [software]*.
- Trawick, S., & Bero, T. (2021). Gender assignment: Monolingual constraints contribute to a bilingual outcome. *International Journal of Bilingualism*, 26, 181–197. <https://doi.org/10.1177/13670069211053466>
- van Rij, J., Wieling, M., Baayen, R. H., & van Rijn, H. (2020). *itsadug: Interpreting time series and autocorrelated data using GAMMs (R package, v2.4)*.
- Wieling, M. (2018). Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between L1 and L2 speakers of English. *Journal of Phonetics*, 70, 86–116.
- Wood, S. N. (2017). *Generalized additive models: An introduction with R*. Taylor & Francis Group.