Articles as flankers: The effect of grammatical gender depends on the task

Marina Serrano Carot, Pablo Gomez, and Bernhard Angele

Faculty of Education and Language, Universidad Antonio de Nebrija

Department of Psychology, Skidmore College

Author Note

Marina Serrano Carot https://orcid.org/0009-0001-9706-2413

Pablo Gomez https://orcid.org/0000-0003-4180-1560

Bernhard Angele https://orcid.org/0000-0001-8989-8555

The authors have no conflicts of interest to disclose. This research has been supported by Grant XXXXX from California State University San Bernardino and Grant 2022-T1/SOC-23980 from the Community of Madrid. Author roles were classified using the Contributor Role Taxonomy (CRediT; credit.niso.org) as follows: Marina Serrano Carot: conceptualization, methodology, investigation, software, formal analysis, writing; Pablo Gomez: conceptualization, editing; Bernhard Angele: conceptualization, methodology, software, validation, writing, supervision, editing

Correspondence concerning this article should be addressed to Marina Serrano Carot, Email: mserranc@nebrija.es

Abstract

We report three experiments using a novel variation of the flanker paradigm (Dare & Shillcock, 2013) in order to test whether readers process articles and nouns together, taking advantage of the fact that in Spanish, as in many languages with grammatical gender, articles must be congruent with the nouns they precede. In Experiment 1, we asked Spanish native speakers to make grammatical gender judgments about centrally presented Spanish nouns (e.g. 'mesa'). The central nouns could be presented with congruent article flankers ('la mesa la'), incongruent article flankers ('el mesa el') or no flankers. We found that the presence of incongruent flankers led to slower and less accurate responses compared to the other two conditions. In Experiment 3, a semantic categorization task, we found no evidence of the article flanker effect. This suggests that articles and nouns are processed jointly, but the impact of this joint processing depends on the task.

Keywords: word recognition, flanker task, articles, nouns, word grouping hypothesis, multi-component unit hypothesis, parafoveal processing

Articles as flankers: The effect of grammatical gender depends on the task

Articles are fundamental building blocks of many languages, particularly those that are members of the Indo-European family. They belong to the syntactic category of determiners and are commonly used to indicate the grammatical definiteness of a noun phrase. Definiteness has been the subject of extensive investigation, both in philosophy (Russell, 1905) and linguistics (e.g., Lyons, 1999). Very broadly speaking, a definite noun phrase refers to a specific and known object or entity, while an indefinite noun phrase refers to a newly introduced or previously unspecified object or entity.

In languages where every noun has a grammatical gender, such as German and most Romance languages, the articles are gendered as well, and every noun needs to be accompanied by the article corresponding to its grammatical gender (as well as number, and, in languages like German, case). This is known as agreement (for an overview, see, e.g., Wechsler & Zlatić, 2003). In practice, this means that a Spanish feminine noun such as *mesa* (table) needs to be preceded by the feminine definite article *la* or the feminine indefinite article *una*, and using the masculine articles, as in *el* mesa* or *un* mesa,* is incorrect. When a noun is presented together with an incongruent article, this leads to delayed processing, known as the *gender congruency effect* (Beatty-Martínez & Dussias, 2019; for a review, see Friederici & Jacobsen, 1999).

Are articles and words processed together?

Because articles and nouns are so closely associated, one might wonder if they are processed as a single unit. This question is particularly relevant considering the ongoing debate on whether word identification during reading is serial, i.e., occurs one word at a time, or parallel, with multiple words being identified at the same time (for an overview of the debate, see Reichle et al., 2009; Schotter & Payne, 2019; Snell & Grainger, 2019; White et al., 2019). The idea that articles and nouns are processed as one unit was first put forward by Radach (1996) as the *Word Grouping Hypothesis*. In a large eye-movement data set, Radach observed that groups of short words followed by longer words, such as groups of articles and nouns (e.g., *the table*) are targeted as if they were one longer word, with fixations aimed at the center or slightly left of center (the *preferred viewing location,* Rayner, 1979) of the entire group (e.g. the *t* or the *a* in *table*), rather than the reader targeting the article and the noun separately. This often results in the short word being not fixated at all, i.e., being skipped, which is a common observation in studies of eye-movements during reading. Radach (1996) claimed that this happened for all groups of shorter words followed by longer words, not just articles and nouns. Drieghe et al. (2008) performed an experiment to test this claim for different types of short words, comparing articles (*the*) to other high-frequency three-letter words (e.g., *two*) and to longer words (e.g., *other*) in the exact same sentence context. Drieghe et al. (2008) replicated Radach (1996)'s findings for the articles, but not the other short words; however, they argued that their findings could be explained by the high skipping rate associated with articles due to their short length and high word frequency rather than due to a different saccade targeting mechanism. Still, Drieghe et al. (2008)'s results are compatible with articles and nouns being processed jointly.

Recently, the idea of certain groups of familiar words such as *teddy bear* being processed as a single unit has received renewed attention due to the Multi-Constituent Unit (MCU) hypothesis (Zang, 2019; see also Cutter et al., 2014; Yu et al., 2016). The MCU assumes that frequently co-occurring words can be processed as a single unit if they all refer to the same concept. At the unit level, processing is still serial. The MCU hypothesis is particularly relevant for reading in Chinese, where there are many idioms and common phrases whose characters co-occur so frequently that they can be considered lexicalized units (and, indeed, many Chinese dictionaries have entries for common three and four character idioms). Following this perspective, sequences of common articles and nouns (such as *la mesa*) may well occur frequently enough to be considered multi-constituent units.

Parafoveal processing of articles

A different approach to the processing of articles during reading considers how deeply they are processed, especially given that they are so frequently skipped. Angele and Rayner (2013a) used the gaze-contingent display change paradigm (Rayner, 1975) to manipulate the previews readers received of a three-letter target verb, e.g. *ace* in *She was sure she would ace all* *the tests*. The preview could either be identical (*ace*), a random letter string (*fda*) or the definite article *the*. Importantly, this last preview condition resulted in a situation where readers received a parafoveal preview of the upcoming word that looked like an article (and would therefore be likely to be skipped), but appears in a position where an article would be syntactically illegal (*she would the*). If readers take the preceding sentence context into account when making a skipping decision, they should not skip a preview looking like *the* in this case. However, Angele et al.'s (2013a) results show that readers do in fact skip the target word very frequently when its preview looked like *the*, suggesting that they are unable to take the incompatibility between the preceding sentence context and the upcoming parafoveal word into account when making the decision to either fixate or skip the next word. One may think that this shows that the preceding sentence context never plays a role when making a skipping decision, but this is not the case: Words that are predictable from the sentence context are reliably skipped more often than words that are not (Rayner et al., 2011; Slattery & Yates, 2018). Abbott et al. (2015) tested this directly, combining Angele et al.'s (2013a) preview manipulation with a predictability manipulation for the target word (e.g. *If you are shot in the heart you will surely die immediately* vs. *If you are shot in the foot you will surely die immediately*, with *die* being the target word). They found that readers were still very likely to skip a syntactically illegal *the*-preview even when the actual target word *die* was highly predictable.

This leads us to the next possible account of how articles and nouns are processed together. Perhaps the article is simply not processed at any great depth, with readers focusing their processing effort on the noun. In this account, readers merely extract the relevant information on definiteness from the article, which they can often do parafoveally, and then do not process the article any further. This would be compatible with findings involving the Missing Letter Effect (Healy, 1994; Koriat & Greenberg, 1993, 1994; Saint-Aubin & Klein, 2008), which indicate that participants are less likely to notice a letter (e.g. the *t* in *the*) when it is part of a high-frequency function word than when it is part of a lower frequency content word. Similarly, Staub et al. (2019) found that readers failed to detect a repetition error involving *the* (e.g., *Amanda jumped off* *the the swing and landed on her feet.*) in as many as 54.2% of all trials, while repeated nouns (e.g., *jumped off the swing swing*) were noticed 90% of the time. Even only considering trials where both instances of *the* were fixated, readers still failed to report 34% of the repetitions.

Using the Flanker paradigm to study processing of articles

Have readers learned to systematically ignore articles, especially when they are presented together with nouns? In order to investigate this question, we used the *Flanking Letters paradigm* (also referred to simply as the *Flanker paradigm*) which was introduced by Dare and Shillcock (2013). This paradigm, which recently has become very popular in studying how parafoveal information affects foveal word identification, is an adaptation of the classic Flanker task (Eriksen & Eriksen, 1974) and involves presenting a central target word alongside a number of flanker stimuli, usually to the left and right, but also above and below the target word (e.g., Snell, Mathôt, et al., 2018). If the presence and the type of the flankers influence the response time and response accuracy on a task involving the target word, most commonly the lexical decision task (LDT), this is taken as evidence in favor of parallel processing of the target and the flankers (e.g., Snell & Grainger, 2018; cf. Schotter & Payne, 2019).

So far, the Flanker paradigm has been used frequently to investigate orthographic processing by comparing the effect flankers that were orthographically related or unrelated to the target word (Dare & Shillcock, 2013; Snell, Mathôt, et al., 2018; Snell, Bertrand, et al., 2018; Snell et al., 2021; Snell & Grainger, 2018; Snell & Simon, 2025), but there are also studies that investigate the effect of syntactic congruency by manipulating whether flankers and target are the same part of speech on a syntactic categorization task (Snell, 2024) or by manipulating whether the flankers form a grammatical context for the target word Vandendaele et al. (2025). The Flanker paradigm has also been used to study semantic processing: Meade et al. (2021) presented flankers that were either of the same semantic category as the target word (*coyote wolf coyote*) or a different semantic category (*coyote sofa coyote*). In a semantic categorization task (animal vs. not an animal), they found both slower response times and higher amplitude negative ERPs in the N400 window.

Present study

The fact that, in languages such as Spanish, there must be gender agreement between articles and the subsequent nouns makes it possible to use the Flanker paradigm to test whether Spanish readers obligatorily process articles and nouns together. According to the *gender congruency effect*, if we show participants a noun flanked by an incongruent article (e.g. *el mesa el*), this should lead to disruption in a task that relies on the joint processing of article and noun compared to either a flanker condition with congruent articles (*la mesa la*) or a control condition without flankers. We tested this in three experiments, each with a different task. In Experiment 1, we asked participants to respond based on the grammatical gender of a target noun (grammatical gender categorization task). When performing this task, it is likely that speakers attempt to recall the article that precedes the noun. This is supported, for example, by the fact that Spanish native speakers are often confused about the gender of words such as *agua,* which are feminine, but, for phonological reasons, are preceded by *el* instead of *la* (Eddington & Hualde, 2008). The flankers, if they are processed along with the target words, will likely interfere with this process. In Experiment 2, participants performed a lexical decision task (LDT) on the central target stimulus, with the same flanker conditions as in Experiment 1. In an LDT, incongruent flankers may either interfere with with lexical access or cause a general error signal that may make it more difficult to correctly identify word stimuli. Finally, in Experiment 3, participants performed a semantic categorization task, indicating whether the target word referred to an animal or not. One might assume that the grammatical gender of the target word is irrelevant for this task, but many models of word identification assume, at least implicitly, the lexical representation of a word needs to be activated in order for it to be processed semantically.

Experiment 1: Gender categorization task

In Experiment 1, participants were shown target words flanked by articles that were either congruent or incongruent with the grammatical gender of the target word and were asked to indicate the grammatical gender of the target word. Our predictions for the outcome of this experiment were as follows: (1) We expected that this task would maximize the effect of flanker-noun gender congruency as grammatical gender was directly relevant to the task, with incongruent flankers leading to longer response times and more errors than congruent flankers. (2) Additionally, there was a no-flanker control condition. We expected both response time and accuracy to be in between the congruent and incongruent conditions, however, there is some evidence that the mere presence of any flanker can inhibit processing of the central word (Snell & Grainger, 2018). (3) There is evidence that, in Spanish, the masculine gender has a special status as the default grammatical gender (Beatty-Martínez & Dussias, 2019). For example, most loanwords are assigned the masculine gender (e.g., *el bit vs. la web;* de la Cruz Cabanillas et al., 2007) and children tend to assign the masculine gender to new, ambiguous nouns (Pérez-Pereira, 1991). If there is a systematic difference between processing grammatically male and grammatically female words, we expected to observe this in reaction time or accuracy or both. (4) If the grammatical gender of the target word influences how participants process the flankers (for example because there is a default grammatical gender that can be modified by the information from the flankers, such that the flankers only have an effect if they do not correspond to the default gender), we expected that we would observe an interaction between word gender and flanker effect.

Method

We report here how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the two experiments. The experiment was preregistered; the preregistration form can be found at https://osf.io/5j9q2/?view_only=017755699aa1467b8de9f4a0584729bc. Data, experiment scripts, and stimuli can be found at https://osf.io/avbh5/?view_only=f5c00d9af2e144cc969c96fbd04452d2. The experiment (as well as Experiments 2 and 3) was conducted in accordance with the ethical guidelines laid out in the Helsinki Declaration and was approved by the ethics boards of the California State University San Bernardino (IRB:) and Nebrija University (UNNE-2024-0012).

Rationale for sample size

Brysbaert (2019) established a simple rule of thumb for designs involving both random effects for participants and for language (e.g. words or sentences). They recommended a minimum of 1,600 observations per condition to reliably detect small effects. Our design has a total of six conditions (3 flanker conditions X 2 grammatical genders) and 450 items, meaning that each participant contributes 75 observations per condition. This means that 22 participants would theoretically be sufficient to reach the 1,600 observations. However, as there is no previous research on articles as flankers and we had no clear expectations regarding the size of the flanker effect (and, particularly, the interaction between the flanker manipulation and the grammatical gender of the target word), we decided to increase the target number of observations to 6,400, which would translate to a total of 86 participants. We estimated that, with a budget of $400, we would be able to collect sufficient observation even with a moderate amount of data loss due to low accuracy and participant drop-out. We therefore set a stopping rule to finish data collection when the budget was spent. In the end, participants were more accurate and more likely to finish the experiment than expected, leaving us with more than 100 participants even after exclusions.

Participants

A total of 116 participants were recruited through Prolific. On the Prolific platform, the experiment was only shown to participants who had indicated that they were over 18 years of age, that Spanish was their first language and that they had no diagnosis of a language-related disorder. Participants were paid $2.50 (or the equivalent in their currency) for their participation, corresponding to $10/hour given that the experiment lasted about 15 minutes on average. Out of all the participants, seven were excluded from the analysis due to low accuracy (less than 80% correct responses). The final sample consisted of 109 participants (67 male, 41, 1 non-binary). The median age of the participants was 30 years, with a minimum age of 18 and a maximum age of 54. The majority of the participants stated that their first language was only Spanish (97 participants), with 7 reporting both Spanish and English as their first languages and 5 reporting Spanish and another language as their first languages. All participants were naive to the purpose of the experiment and gave informed consent before the experiment.

Materials

We used a total of 450 high-frequency Spanish target words (e.g., *mesa*) with three, four, and five letters. Half of the target words were female and the other half were male. Table 1 shows the number of target words in each category and their word frequency.

Apparatus

Participants accessed the experiment via a link on the Prolific platform. Participants were instructed to complete the experiment on a laptop or desktop computer, and to use a physical keyboard. The experiment was presented using PsychoJS (the online Javascript-based version of PsychoPy, Peirce et al., 2019) and hosted on Pavlovia. The experiment was presented in the full screen mode of the browser, and participants were informed that the experiment would terminate if the exited full screen mode before completing the experiment.

Procedure

The experiment was conducted online. It was presented in Spanish, and all instructions were given in Spanish. After confirming their participation on Prolific, participants were directed to the experiment. Before the start of the experiment, participants gave informed consent and confirmed that they were in a quiet environment without distractions and would be able to focus on the experiment for 15-20 minutes. They responded to a short survey with demographic data, and confirmed that they fulfilled the selection criteria (Spanish native speakers, no diagnosis of language-related disorders). The experimental procedure was very similar to that used by Snell and Grainger (2018): During each experimental trial, participants first saw a fixation target (two vertical bars located above and below the center point of the screen, with the center of the screen marked by the gap between the bars) for 500 ms. The target word then appeared for in the center of the screen, between the two vertical bars, either on its own (in the control condition) or along with the flankers to the left and right. The flankers were either congruent articles (e.g., *la mesa la*) or incongruent articles (e.g., *el mesa el*). The assignment of target words to the three conditions was counterbalanced such that every participant saw 150 target words each in the control, congruent, and incongruent conditions and that each target word was presented the same number of times in each condition across participants. After 150 ms, both the target word and the flankers disappeared. This presentation time was short enough that participants were not be able to plan and execute a saccade before the stimulus offset. Participants were instructed to make their decision solely based on the target word (ignoring the flankers) and as soon as the target word disappeared. They were asked to respond as quickly and as accurately as possible. Participants responded by using using either the “A” and “L” keys on the physical keyboard of their devices. The assignment of these keys was counterbalanced such that half of the participants received one assignment (A = masculine and L = feminine) and the other half of participants received the other (A = feminine and L = masculine). After making the response, participants received feedback: a green tick mark emoji for correct responses or a red “x” emoji for incorrect responses were presented for 500 ms in the center of the screen. If no response was made within 2000 ms, the trial ended and participants received the feedback “Too slow!” (“¡Muy lento!”) and the response was counted as an incorrect response. After the trial end, the experiment proceeded to the next trial until all trials had been completed. The trials were presented in random order. Participants completed 15 practice trials to familiarize themselves with the procedure and then performed the 450 experimental trials. Participants were offered the opportunity to take two breaks during the experiment, after 150 trials and after 300 trials. After the experiment, participants were thanked for their participation and returned to the Prolific platform. The experiment took about 15 minutes on average.

Data analysis and dependent variables

We analysed the data by fitting Bayesian linear mixed models using the *brms* package (Bürkner, 2017, 2018, 2021) in R[1]. These models included the flanker condition (control

[1] The versions of R and all packages used are as follows: R (Version 4.5.1; R Core Team, 2021) and the R-packages *brms* (Version 2.22.0; Bürkner, 2017, 2018, 2021), *conflicted* (Version 1.2.0; Wickham, 2023a), *dplyr* (Version 1.1.4; Wickham, François, et al., 2023), *flextable* (Version 0.9.9; Gohel & Skintzos, 2024), *forcats* (Version 1.0.0; Wickham, 2023b), *ftExtra* (Version 0.6.4; Yasumoto, 2024), *ggplot2* (Version 3.5.2; Wickham, 2016), *gt* (Version

vs. congruent vs. incongruent) as well as the grammatical gender of the target word (feminine vs. masculine) and their interaction as categorical fixed effects. We used Helmert contrasts for the flanker condition, with the first contrast comparing the incongruent flanker condition (represented as -1) to the mean of the congruent and the control conditions (represented as .5 and .5) and the second contrast comparing the no-flanker control condition (represented as 1) to the congruent flanker condition (represented as -1). For the grammatical gender, we used a sum contrast comparing the feminine (-1) to the masculine grammatical gender (1). We used the ex-Gaussian distribution to model correct response times (in s) after excluding excessively short (response time < .25 s; 705 of 45450 total correct responses) and long responses (response time > 1.8 s; 6 responses), with both the mean of the Gaussian component μ_{RT} and the scale parameter of the exponential component β_{RT} (equaling the inverse of the rate parameter λ) being allowed to vary between conditions. To model response accuracy, we used the Bernoulli distribution with a logit link after excluding no-response trials (116 trials) and short and long responses as above (resulting in 48052 out of 48934 no-timeout responses being included). We used the default priors suggested by *brms* except for the coefficients for the fixed effects, for which we applied weakly informative priors of $\beta \sim N(0, 1)$ in order to rule out improbably large effect sizes. Each model was fitted using four chains with 5000 iterations each with 1000 warmup iterations. The models converged successfully (all $\hat{R}$s = 1.00). We report the mean, the estimates (b) and the 95% Bayesian Credible Intervals (95% CrIs) based on the posterior distribution of each model parameter. In order to simplify the interpretation of the posterior distribution, we highlight whether 0 is not a credible value for its coefficient (i.e., if it is not part of the 95% CrI). Additionally, we report the proportion of the posterior distribution on the same side of 0 (i.e., the proportion of the posterior distribution

1.0.0; Iannone et al., 2024), *here* (Version 1.0.1; Müller, 2020), *kableExtra* (Version 1.4.0; Zhu, 2021), *knitr* (Version 1.50; Xie, 2015), *lubridate* (Version 1.9.4; Grolemund & Wickham, 2011), *papaja* (Version 0.1.3; Aust & Barth, 2023), *purrr* (Version 1.0.4; Wickham & Henry, 2023), *Rcpp* (Eddelbuettel & Balamuta, 2018; Version 1.0.14; Eddelbuettel & François, 2011), *readr* (Version 2.1.5; Wickham, Hester, et al., 2023), *stringr* (Version 1.5.1; Wickham, 2023c), *tibble* (Version 3.3.0; Müller & Wickham, 2023), *tidyr* (Version 1.3.1; Wickham, Vaughan, et al., 2023), *tidyverse* (Version 2.0.0; Wickham et al., 2019), and *tinylabels* (Version 0.2.5; Barth, 2023).

that is greater than 0 or less than 0). We always report the greater proportion (e.g., if 49% of the posterior distribution is < 0 and 51 % is > 0 we report .51). We also report the ratio of the two proportions (with the greater proportion always in the numerator), which effectively corresponds to a Bayes Factor for the directional hypothesis b > 0 or b < 0, respectively. Finally, we report the Bayes Factor (BF) calculated using the Savage-Dickey density ratio which tests the evidence for the point-null hypothesis b = 0. These measures all provide complementary information about the effect in question. We interpret BF values of 3-10 as moderate evidence, 10-30 as strong evidence, and >30 as very strong evidence for the effect (Jeffreys, 1961).

Results

In general, and after excluding the seven low-accuracy participants (see above), accuracy was very high, with 45450 correct responses out of 48934 total responses. The mean response time for correct answers was 0.51 seconds. The mean response time for incorrect answers was 0.54.

Table 2 summarizes the means and standard deviations of the correct response times by flanker condition and grammatical gender of the target word, as well as the number of observations in each condition.

Table 3 shows the results of the Bayesian linear mixed model for response time, while Table 4 shows the results of the Bayesian generalized linear mixed model for accuracy.

Effects of the flanker condition

In response time, we observed a substantial effect of the incongruent flanker condition compared to the two other conditions in both the mean of the Gaussian component, μ_{RT} ($b = -0.05$, 95% CrI $[-0.06,\ -0.05]$, $P(b < 0) = 1.00$, $ER_{b<0} > 1000$, $BF_{10} > 1000$), indicating longer response times in the incongruent condition, and the scale component, β_{RT} ($b = -0.25$, 95% CrI $[-0.28,\ -0.23]$, $P(b < 0) = 1.00$, $ER_{b<0} > 1000$, $BF_{10} > 1000$), indicating a longer right tail of the response time distribution. Participants also made more errors in the incongruent compared to the congruent condition ($b = 0.70$, 95% CrI $[0.59,\ 0.82]$, $P(b > 0) = 1.00$, $ER_{b>0} > 1000$, $BF_{10} > 1000$). The effect of the congruent flanker condition compared to the control condition was much smaller in μ_{RT} ($b = -0.003$, 95% CrI $[-0.0059, -0.0002]$, $P(b < 0) = 0.98$, $ER_{b<0} = 52.69$, $BF_{10} = 0.01$) and virtually absent in β_{RT} ($b = -0.0037$, 95% CrI $[-0.03, 0.02]$, $P(b < 0) = 0.60$, $ER_{b<0} = 1.52$, $BF_{10} = 0.02$), indicating that the congruent condition was not much different from the control condition in terms of response time. However, participants tended to be slightly more accurate in the congruent condition compared to the control condition ($b = -0.22$, 95% CrI $[-0.35, -0.09]$, $P(b < 0) = 1.00$, $ER_{b<0} > 1000$, $BF_{10} = 16.86$).

Effects of grammatical gender

We found small marginal effects of grammatical gender across flanker conditions: Response times were slightly lower for feminine target words compared to masculine target words (μ_{RT}: $b = 0.0064$, 95% CrI $[0.0034, 0.0094]$, $P(b > 0) = 1.00$, $ER_{b>0} > 1000$, $BF_{10} = 125.67$; β_{RT}: $b = 0.03$, 95% CrI $[0.01, 0.04]$, $P(b > 0) = 1.00$, $ER_{b>0} > 1000$, $BF_{10} = 45.60$), but accuracy was slightly higher for masculine compared to feminine target words ($b = 0.14$, 95% CrI $[0.05, 0.23]$, $P(b > 0) = 1.00$, $ER_{b>0} > 1000$, $BF_{10} = 4.23$).

Interaction between flanker condition and grammatical gender

We did not observe meaningful interactions between grammatical gender of the target word and the flanker condition in response time, but, in terms of accuracy, there was an interaction between grammatical gender of the target word and the contrast comparing the congruent and control flanker conditions, with there being a larger increase in accuracy in the congruent condition compared to the control condition for feminine target words compared to masculine target words ($b = 0.12$, 95% CrI $[0.01, 0.24]$, $P(b > 0) = 0.98$, $ER_{b>0} = 61.99$, $BF_{10} = 0.57$).

Discussion

In the grammatical gender categorization task, we found clear evidence for an effect of the article flankers. While the congruent flankers did not affect reaction times much compared to the control condition, and only led to a small increase in accuracy compared to the control condition, the incongruent flankers had a major disruptive effect on processing the target word's grammatical gender. This suggests that the article flankers are obligatorily processed and that the conflict between the incongruent flanker and the correct article has a major effect on performance on the task. We found some evidence in favor of a bias in gender assignment as suggested by Beatty-Martínez and Dussias (2019), with feminine target words being processed slightly faster but slightly less accurately than masculine target words, especially in the no-flanker control condition. Overall, however, these biases were very small compared to the effect of the incongruent flankers. In summary, Experiment 1 shows clear parafoveal processing of article flankers during a grammatical gender categorization task, with participants being unable to ignore the parafoveal article information even when it disrupts the task. This is compatible with the assumption that articles and nouns are processed jointly.

Experiment 2: Lexical decision task

In Experiment 2, participants were shown target stimuli that could either be words or pseudowords and were flanked by articles that were either congruent or incongruent with the gender of the central word or the apparent grammatical gender of the matched pseudoword. Our predictions for the outcome of this experiment were as follows: (1) Compared to Experiment 1, the lexical decision task does not require participants to process the grammatical gender of the target explicitly, as it is irrelevant for the decision of whether the target stimulus is a word or not. However, if the flanker articles are processed obligatorily and at the same time as the central stimulus, the error signal from the mismatch between incongruent article and noun may interfere with the decision of whether the target stimulus is a word, delaying the decision or even resulting in an incorrect rejection of a word stimulus. As the pseudowords were matched with the words to have the same apparent gender as the matched target word (e.g. *cosa* was matched with *poga* and *pelo* was matched with *meno*), every pseudoword had an apparent gender. A pseudoword presented with a "congruent" flanker may be harder to reject than a pseudoword presented with a flanker incongruent with its apparent gender. (2) As in Experiment 1, we may observe a systematic difference between processing grammatically male and grammatically female words. If this is the case, we expected to observe this in reaction time or accuracy or both. (3) Finally, and again as in Experiment 1, if the grammatical gender of the target word influences how participants process the flankers, we expected that we would observe an interaction between subject gender and flanker effect.

Method

We report here how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the two experiments. The experiment was preregistered; the preregistration form can be found at https://osf.io/wt2qg/?view_only=d39d63d23f674e2ab47e7d06f1c22b6d. Data, experiment scripts, and stimuli can be found at https://osf.io/nsu4p/?view_only=c5f86ce479cf4866b6c30523f375219c.

Rationale for sample size and stopping rule

In Experiment 2, our design had a total of 480 items (240 words and 240 pseudowords) and eight conditions (congruent vs. incongruent flankers X masculine vs. feminine (apparent) gender X word vs. pseudoword). This means that every participant contributed 60 observations per condition. Using the rule of thumb established by Brysbaert (2019), this means that 27 participants would theoretically be sufficient to reach the 1,600 observations. In Experiment 1, we observed a clear effect of the flanker manipulation, and, in an exploratory analysis on the Experiment 1 data, found that we could detect the main effect of flanker condition conclusively even at 20 participants (corresponding to 1,500 observations per condition). Therefore, if the effect in the lexical decision task has the same size as in the gender categorization task, 1,600 observations should be sufficient. Out of an abundance of caution, we instead opted for 3,000 observations (50 participants). Additionally, we performed a (frequentist) power simulation using SimR with the data from the first experiment and assuming a much smaller effect size (10 ms instead of 52 ms), which showed that, with 50 participants, we could expect a power of .97. Even through the frequentist simulation is not an ideal predictor for the outcome of our Bayesian analyses, it provides a guideline for the number of participants that may be needed to detect an effect. We planned to collect data from slightly more than 50 participants to account for potential data loss and exclusions, e.g. due to low response accuracy and to ensure a balanced design.

Participants

A total of 59 participants were recruited through Prolific. On the Prolific platform, the experiment was only shown to participants who had indicated that they were over 18 years of age, that Spanish was their first language and that they had no diagnosis of a language-related disorder. Participants were paid $2.50 (or the equivalent in their currency) for their participation, corresponding to $10/hour given that the experiment lasted about 15 minutes on average. Out of all the participants, five were excluded from the analysis due to low accuracy (less than 80% correct responses). The final sample consisted of 54 participants (33 male, 20 female, 1 non-binary). The median age of the participants was 27 years, with a minimum age of 18 and a maximum age of 41. The majority of the participants stated that their first language was only Spanish (49 participants), with 2 reporting both Spanish and English as their first languages and 3 reporting Spanish and another language as their first languages. All participants were naive to the purpose of the experiment and gave informed consent before the experiment.

Materials

We used a total of 240 high-frequency Spanish target words (e.g., *mesa*) with four and five letters. Half of the target words were female and the other half were male. Table 5 shows the number of target words in each category and their word frequency as well as the OLD20 (Yarkoni et al., 2008). For each target word, we created a matched pseudoword using the Wuggy pseudoword generator (Keuleers & Brysbaert, 2010) with the Spanish word data base and default settings. We ensured that each pseudoword matched the corresponding word in terms of its word ending (which, for most words, is indicative of grammatical gender in Spanish). We included this "apparent gender" as a predictor in the analyses for the pseudoword trials (see below).

Apparatus

The apparatus was identical to Experiment 1.

Procedure

The procedure was identical to Experiment 1, except that participants performed a lexical decision instead of a grammatical gender categorization. Thet responded by using using either the “A” and “L” keys on the physical keyboard of their devices. As in Experiment 1, the assignment of these keys was counterbalanced such that half of the participants received one assignment (A = word and L = nonword) and the other half of participants received the other (A = nonword and L = word). In Experiment 2, we only included the congruent and incongruent flanker conditions, as the results in the no-flanker control condition were very similar to those in the congruent flanker condition in Experiment 1.

Data analysis and dependent variables

As in Experiment 1, we analyzed the data by fitting Bayesian linear mixed models including the flanker condition (congruent vs. incongruent) as well as the grammatical gender of the target word or the apparent gender of the pseudoword (feminine vs. masculine) and their interaction as categorical fixed effects. We fitted separate models for words and pseudoword stimuli. We used a sum contrast for the flanker condition, with the incongruent flanker condition represented as -1 and the congruent flanker condition represented as 1. For the grammatical gender, we again used a sum contrast comparing the feminine (-1) to the masculine grammatical gender (1). The rest of the model specifications were identical to those used in Experiment 1. All models converged successfully (all $\hat{R}$s = 1.00). We report the results as in Experiment 1.

Results

In general, and after excluding the low-accuracy participants (see above), accuracy was very high, with 24268 correct responses out of 26345 total responses (for an average of 3,033.50 included observations per condition). The mean response time for correct answers was 0.47 seconds. The mean response time for incorrect answers was 0.49.

Table 6 summarizes the means and standard deviations of the correct response times by flanker condition and grammatical gender of target words, as well as the number of observations in each condition, while Table 7 shows the same for the pseudoword trials.

Table 8 shows the results of the Bayesian linear mixed model for response time, while Table 10 shows the results of the Bayesian generalized linear mixed model for accuracy. Table 9 and Table 11 show the same for response time and accuracy in pseudoword trials, respectively.

Effects of the flanker condition

On word trials, we observed an effect of the incongruent flanker condition compared to the congruent flanker in the mean of the Gaussian component of response time, μ_{RT} ($b = -0.0049$, 95% CrI $[-0.007,\ -0.003]$, $P(b < 0) = 1.00$, $ER_{b<0} > 1000$, $BF_{10} > 1000$), indicating longer response times in the incongruent condition, and a slightly larger effect in the scale component, β_{RT} ($b = -0.04$, 95% CrI $[-0.07,\ -0.02]$, $P(b < 0) = 1.00$, $ER_{b<0} > 1000$, $BF_{10} = 3.50$), indicating some evidence for a longer right tail of the response time distribution. There was little evidence for effects of any of the manipulations on accuracy for word trials in Experiment 2 ($b = 0.07$, 95% CrI $[-0.03,\ 0.16]$, $P(b > 0) = 0.91$, $ER_{b>0} = 10.48$, $BF_{10} = 0.12$).

On pseudoword trials, we observed the same pattern of effects as for words, but in the opposite direction, with shorter response times in the incongruent flanker condition compared to the congruent flanker condition (μ_{RT}: $b = 0.003$, 95% CrI $[0.001,\ 0.005]$, $P(b > 0) = 1.00$, $ER_{b>0} = 515.13$, $BF_{10} = 0.07$; β_{RT}: $b = 0.04$, 95% CrI $[0.02,\ 0.07]$, $P(b > 0) = 1.00$, $ER_{b>0} > 1000$, $BF_{10} = 2.24$), indicating that it was slightly easier to correctly reject a pseudoword when it was presented together with the incorrect articles for its apparent gender. As for word trials, little evidence that the flanker condition affected accuracy for pseudoword trials in Experiment 2 ($b = -0.06$, 95% CrI $[-0.15,\ 0.03]$, $P(b < 0) = 0.92$, $ER_{b<0} = 10.92$, $BF_{10} = 0.12$).

Effects of grammatical gender

For word trials in Experiment 2, there was little evidence for a marginal effect of grammatical gender across flanker conditions in either response times (μ_{RT}: $b = 5.7e-05$, 95% CrI $[-0.0033,\ 0.0034]$, $P(b > 0) = 0.51$, $ER_{b>0} = 1.06$, $BF_{10} = 0.0017$; β_{RT}: $b = -0.00093$, 95% CrI $[-0.03,\ 0.03]$, $P(b < 0) = 0.53$, $ER_{b<0} = 1.11$, $BF_{10} = 0.02$) or accuracy ($b = 0.05$, 95% CrI $[-0.08,\ 0.17]$, $P(b > 0) = 0.76$, $ER_{b>0} = 3.12$, $BF_{10} = 0.08$).

The same was true for response times in pseudoword trials (μ_{RT}: $b = 0.0037$, 95% CrI $[-0.00036,\ 0.0078]$, $P(b > 0) = 0.96$, $ER_{b>0} = 26.21$, $BF_{10} = 0.0098$; β_{RT}: $b = 0.01$, 95% CrI $[-0.02,\ 0.05]$, $P(b > 0) = 0.79$, $ER_{b>0} = 3.85$, $BF_{10} = 0.02$). However, there was an effect of the apparent gender of the pseudoword on accuracy, with apparent feminine pseudowords being slightly harder to correctly reject than apparent masculine pseudowords over all ($b = -0.16$, 95% CrI $[-0.31,\ -0.01]$, $P(b < 0) = 0.98$, $ER_{b<0} = 56.76$, $BF_{10} = 0.72$).

Interaction between flanker condition and grammatical gender

We did not observe meaningful interactions between grammatical gender of the target word and the flanker condition in response time or accuracy for either words (μ_{RT}: $b = 0.0011$, 95% CrI $[-0.00071,\ 0.003]$, $P(b > 0) = 0.88$, $ER_{b>0} = 7.36$, $BF_{10} = 0.0019$; β_{RT}: $b = 0.02$, 95% CrI $[-0.0083,\ 0.04]$, $P(b > 0) = 0.90$, $ER_{b>0} = 9.24$, $BF_{10} = 0.03$; $\theta_{accuracy}$: $b = -0.03$, 95% CrI $[-0.11,\ 0.04]$, $P(b < 0) = 0.80$, $ER_{b<0} = 3.95$, $BF_{10} = 0.06$) or pseudowords (μ_{RT}: $b = -0.00039$, 95% CrI $[-0.0024,\ 0.0016]$, $P(b < 0) = 0.65$, $ER_{b<0} = 1.82$, $BF_{10} = 0.0011$; β_{RT}: $b = -0.01$, 95% CrI $[-0.04,\ 0.02]$, $P(b < 0) = 0.79$, $ER_{b<0} = 3.74$, $BF_{10} = 0.02$; $\theta_{accuracy}$: $b = 0.01$, 95% CrI $[-0.07,\ 0.09]$, $P(b > 0) = 0.61$, $ER_{b>0} = 1.54$, $BF_{10} = 0.04$).

Discussion

In Experiment 2, we found that the article flankers affected processing of both words and pseudowords in a lexical decision task. For words, having an incongruent flanker increased response times, while the opposite was true for pseudowords. This suggests that participants use, at least to some degree, a type of “error signal” to perform the lexical decision task, and that the incongruent articles increase this error signal, interfering with the correct acceptance of words and facilitating the correct rejection of pseudowords. It is important to note that, compared to the large flanker effect we observed in Experiment 1, where grammatical gender needed to be identified explicitly as part of the task, the effect in Experiment 2 was much smaller. Unlike in Experiment 1, this effect were also only observed in response time and not in accuracy. There was again some evidence in favor of a bias in gender assignment, with apparent female pseudowords being harder to reject than apparent male pseudowords, regardless of the flanker condition. In summary, Experiment 2 shows clear evidence of parafoveal processing of article flankers during the lexical decision task, with participants being unable to ignore the parafoveal article information even though it was irrelevant to the task. Just as the results from Experiment 1, our findings in Experiment 2 are compatible with the assumption that articles and nouns are processed jointly, but it is clear that the task plays a major role in determining the extent of this obligatory joint processing.

Experiment 3

In Experiment 3, participants were shown target words flanked by articles that were either congruent or incongruent with the grammatical gender of the target word and were asked to indicate whether the target words referred to animals or not. After the grammatical categorization task in Experiment 1, which required participants to make an explicit response about the grammatical gender of the target word, and Experiment 2, which required participants to decide whether the letter string flanked by articles was a word or not, we hypothesized that this final task would require the least explicit or implicit processing of grammatical gender out of our three tasks. However, as described above, many models of word recognition (such as the Interactive Activation Model) assume that a word needs to be identified (i.e., lexical access needs to occur) before semantic information about it can be retrieved. Experiment 2 suggests that the incongruent flankers can interfere with lexical access. If they also interfere with the retrieval of semantic properties of a word, we would (1) expect incongruent flankers leading to longer response times and more errors than congruent flankers on the semantic categorization task. (2) There may also be a systematic difference between processing grammatically male and grammatically female words syntactically, which we may observe in response time or accuracy or both. (3) If the grammatical gender of the target word influences how participants process the flankers, and if this affects semantic processing of the target word, we expected that we would observe an interaction between word gender and flanker effect.

Method

Rationale for sample size

In Experiment 3, we showed each participant a total of 416 non-animal target nouns, 208 of them male and 208 female (we did not plan on analyzing the 140 animal noun trials). As in Experiment 2, there were two flanker conditions (congruent vs. incongruent), resulting in a total of 104 observations per participant and condition. Using the rule of thumb established by Brysbaert (2019), this means that 16 participants would theoretically be sufficient to reach the 1,600 observations. Given the smaller effect observed in Experiment 2 compared to Experiment 1, we suspected that the effect size in Experiment 3 could be even smaller. Because of this, we set our target number of observations per condition (flanker condition * word gender) to 6,400. This corresponds to 60 participants. Since there was likely to be some data loss/drop out, we collected data from slightly more than 60 participants to account for potential data loss and exclusions, e.g. due to low response accuracy and to ensure a balanced design.

Participants

A total of 63 participants were recruited through Prolific. On the Prolific platform, the experiment was only shown to participants who had indicated that they were over 18 years of age, that Spanish was their first language and that they had no diagnosis of a language-related disorder. Participants were paid $3 (or the equivalent in their currency) for their participation, corresponding to $9/hour given that the experiment lasted about 20 minutes on average. Out of all the participants, three were excluded from the analysis due to low accuracy (less than 80% correct responses). The final sample consisted of 60 participants (36 male, 22 female, 2 non-binary). The median age of the participants was 28 years, with a minimum age of 18 and a maximum age of 40. The majority of the participants stated that their first language was only Spanish (53 participants), with 7 reporting both Spanish and English as their first languages and 0 reporting Spanish and another language as their first languages. As in Experiments 1 and 2, all participants were naive to the purpose of the experiment and gave informed consent before the experiment.

Materials

We used the same stimuli as in Experiment 1, with some ambiguous words that could refer refer to both animals and non-animals (e.g., "manta") removed and more animal words added, for a total of 416 non-animal (208 feminine and 208 masculine) and 140 animal trials (70 feminine and 70 masculine). No non-animal word was repeated, but the 70 unique animal words were each presented twice to ensure there were sufficient trials (25.18%) that required an "animal" response. Table 12 shows the number of target words in each category and their word frequency as well as the OLD20 (Yarkoni et al., 2008).

Apparatus

The apparatus was identical to that used in Experiments 1 and 2.

Procedure

The procedure was identical to that used in Experiment 1 and 2, except that participants performed a semantic categorization instead of a grammatical gender categorization or lexical decision. They responded by using using either the "S" and "L" keys on the physical keyboard of their devices (we avoided the "A" key in Experiment 3 due to the potential association between "A" and "animal"). As in Experiment 1, the assignment of these keys was counterbalanced such that half of the participants received one assignment (S = animal and L = non-animal) and the other half of participants received the other (S = non-animal and L = animal). As in Experiment 2, we only included the congruent and incongruent flanker conditions.

Data analysis and dependent variables

As in Experiments 1 and 2, we analyzed the data by fitting Bayesian linear mixed models including the flanker condition (congruent vs. incongruent) as well as the grammatical gender of the target word (feminine vs. masculine) and their interaction as categorical fixed effects. We only included non-animal trials in the analysis, as "yes" responses may carry an extra cost. As in Experiment 2, we used a sum contrast for the flanker condition, with the incongruent flanker condition represented as -1 and the congruent flanker condition represented as 1. For the grammatical gender, we again used a sum contrast comparing the feminine (-1) to the masculine grammatical gender (1). The rest of the model specifications were also identical to those used in Experiment 2. All models converged successfully (all $\hat{R}$s = 1.00). We report the results as in Experiments 1 and 2.

Results

In general, and after excluding the low-accuracy participants (see above), accuracy was very high, with 31077 correct responses out of 33227 total responses (for an average of 3,884.62 included observations per condition). The mean response time for correct answers was 0.47 seconds. The mean response time for incorrect answers was 0.49.

Table 13 summarizes the means and standard deviations of the correct response times by flanker condition and grammatical gender of target words, as well as the number of observations in each condition.

Table 14 shows the results of the Bayesian linear mixed model for response time for non-animal target words, while Table 15 shows the results of the Bayesian generalized linear mixed model for accuracy (i.e. the proportion of correct rejections of target words as non-animals).

Effects of the flanker condition

Unlike in the first two experiments, we did not observe any evidence for an effect of the flanker condition in the mean of the Gaussian component of response time, μ_{RT} ($b = 7.1e - 05$, 95% CrI $[-0.0013,\ 0.0015]$, $P(b > 0) = 0.54$, $ER_{b>0} = 1.17$, $BF_{10} < 0.001$) or the scale component, β_{RT} ($b = 0.0069$, 95% CrI $[-0.01,\ 0.02]$, $P(b > 0) = 0.78$, $ER_{b>0} = 3.55$, $BF_{10} = 0.01$). There was also little evidence for effects of any of the manipulations on accuracy for word trials in Experiment 2 ($b = -0.04$, 95% CrI $[-0.18,\ 0.10]$, $P(b < 0) = 0.72$, $ER_{b<0} = 2.52$, $BF_{10} = 0.08$).

Effects of grammatical gender

Just like in Experiment 1 (but unlike in Experiment 2), there was some evidence for a marginal effect of grammatical gender across flanker conditions in response times, with both the mean of the gaussian component shifted to the right for female compared to male target words

(μ_{RT}: $b = 0.0043$, 95% CrI $[0.0022,\ 0.0065]$, $P(b > 0) = 1.00$, $ER_{b>0} > 1000$, $BF_{10} = 911.96$) and the scale component for the exponential component being larger for female compared to male target words (β_{RT}: $b = 0.04$, 95% CrI $[0.02,\ 0.06]$, $P(b > 0) = 1.00$, $ER_{b>0} > 1000$, $BF_{10} = 9.50$), indicating a longer right tail of the distribution. However, there was no evidence for gender effect on accuracy ($b = -0.16$, 95% CrI $[-0.33,\ 0.0015]$, $P(b < 0) = 0.97$, $ER_{b<0} = 37.46$, $BF_{10} = 0.53$).

Interaction between flanker condition and grammatical gender

We did not observe meaningful interactions between grammatical gender of the target word and the flanker condition in response time or accuracy (μ_{RT}: $b = 0.00065$, 95% CrI $[-0.00074,\ 0.002]$, $P(b > 0) = 0.82$, $ER_{b>0} = 4.68$, $BF_{10} = 0.0011$; β_{RT}: $b = 0.01$, 95% CrI $[-0.0067,\ 0.03]$, $P(b > 0) = 0.89$, $ER_{b>0} = 7.91$, $BF_{10} = 0.02$; $\theta_{accuracy}$: $b = -0.04$, 95% CrI $[-0.15,\ 0.06]$, $P(b < 0) = 0.78$, $ER_{b<0} = 3.53$, $BF_{10} = 0.07$).

Discussion

Unlike in the previous experiments, we did not observe an effect of the article flanker condition in Experiment 3. This suggests that, for certain tasks, joint processing of article and noun is either not mandatory, or the joint processing does not influence the response. It seems that a semantic judgment task is high-level enough that the grammatical gender of the target word is either not processed or does not interfere with the semantic decision. We will discuss the implications of this finding together with those from Experiments 1 and 2 in the following.

General discussion

In the present study, we reported three experiments using a novel flanker manipulation involving articles in order to investigate whether Spanish readers process articles and nouns as a single unit. The flankers were always articles that could either be congruent or incongruent with the target stimulus (in Experiment 1, we also included a no-flanker control condition). All information required for the tasks was provided by the central target stimulus, such that the flanker articles were irrelevant and should have been ignored by an ideal observer. In fact, the participants were explicitly asked to ignore the flankers in the instructions. Despite this, there was clear evidence of the participants processing the flankers in Experiment 1 (a grammatical gender categorization task) and Experiment 2 (a lexical decision task), as indicated by the effect of the flankers on response time in both experiments and on accuracy in Experiment 1. This suggests that, at least under some circumstances, readers cannot help but process articles and nouns as a unit. However, in Experiment 3 (a semantic categorization task), we found no evidence for a flanker effect, suggesting that, in this case, the participants were able to ignore the flankers. There are two possible explanations for this difference in results between the experiments. First, it is possible that readers were able to process only the nouns in Experiment 3, ignoring the articles. The question is why they were not able to do the same in Experiments 1 and 2. It is not clear how a semantic categorization task would cause readers to engage in serial processing rather than parallel processing. Second, it is possible that the participants did process the articles in Experiment 3, but, since the article information was not at all relevant for the task, it did not interfere with their responses.

Implications for models of word recognition: Does lexical processing precede semantic processing?

There are clear implications of these results for models of word recognition. The most important finding, at the level of individual word recognition, is that the article flanker manipulation appears to interfere with a syntactic/lexical task (grammatical gender categorization) and a pure lexical task (lexical decision), but not with a semantic task. This seems to contradict models that assume that lexical processing is a necessary pre-requisite for semantic processing. For example, the original Logogen model (Morton, 1969) made the assumption that, once the logogen for a particular word is activated, all information about the word is active. This includes orthographic, syntactic, and semantic information. Similarly, all these sources of information can activate the logogen. It is not clear why the incompatible syntactic information from the article flanker should inhibit this activation process during a lexical and syntactic task, but not during a semantic task. Similarly, in Forster's Autonomous Search Model (Forster & Bednall, 1976), all information about a lexical entry becomes available once the word is found in one of the access files. While the the interactive activation model [IAM; Rumelhart and McClelland (1982); McClelland and Rumelhart (1981)] does not have an explicit semantic layer, it predicts that, during word recognition, activation spreads from visual feature nodes to letter nodes to word nodes (bottom-up processing), but the higher-level nodes can feed back activation to lower-level nodes (top-down processing). This is how the IAM explains the word superiority effect (Reicher, 1968; Wheeler, 1970), but it also means that it is unlikely that a hypothetical semantic layer would be unaffected by disruption on the lexical layer caused by the incongruent article flanker condition.

Compared to models that propose a particular, all-or-nothing moment of "lexical access" at which all information about a word becomes available simultaneously (and before of which no top-down information is available), our results are more compatible with distributed models such as the triangle model (Seidenberg & McClelland, 1989), a model of word recognition with sets of orthographic, phonological, and semantic units, which are all connected by hidden units. There is no central "logogen"-type representation; rather, the information about a word that is available is represented by the simultaneous activation of the different nodes, and a reader's knowledge about words is represented by the connections between the nodes and the hidden layers. While the triangle model does not have an explicit syntactic layer, one could easily imagine an additional set of units representing the grammatical gender of a word. Alternatively, the knowledge about grammatical gender could be implicit in the knowledge about the context in which words occur ("mesa" occurs frequently in the context of "la" and "una", but never in the context of "el" or "un"). If participants perform the semantic task by exclusively evaluating the state of the semantic units (which are activated by the orthographic information about the noun), their performance might not be affected by the incongruent flankers. On the other hand, the grammatical gender categorization task directly relies on information about the word context, and the lexical decision task may, more indirectly, rely on a general "error signal" which is be triggered by the incongruent article flankers.

Implications for models of reading: Is word recognition serial or parallel?

Our findings suggest a differentiated picture regarding the processing of articles during reading: The presence of an incorrect article most likely does not interfere much with processing the meaning of the subsequent noun, but the grammatical gender information may become relevant as a reader tries to process the syntactic structure of a sentence as a whole. This is precisely what Serrano-Carot and Angele (cite talk?) observed in a gaze-contingent display change experiment similar to that done by Angele and Rayner (2013b), where the preview for an article in Spanish sentences was manipulated such that it could be either congruent or incongruent: There was no immediate effect of the incongruent article previews on ongoing processing, as reflected by skipping probability for the article, but later fixation time measures showed a clear inhibitory effect of the incongruent previews further downstream. In principle, and always with the caveat that a flanker task is of course not the same as sentence reading, our results (especially those of Experiments 1 and 2) are compatible with a parallel account of reading, where articles and nouns are processed at the same time. However, Experiment 3 shows that such parallel processing does not always have to be reflected in behavior. Alternatively, our results are also compatible with a multi-constituent unit approach such as those proposed specifically for articles and nouns by Radach (1996) and, more generally, for frequent combinations of lexical units by Zang (2019), predicting that articles and nouns are processed as one single unit. Multi-constituent units themselves may well be processed either serially or in parallel.

Conclusion

In summary, using our novel article flanker manipulation, we found evidence for obligatory joint processing of articles and nouns. However, this only seems to be the case during certain syntactic and lexical tasks and not during a purely semantic task. Our results are compatible with distributed models of word recognition such as the Triangle model, but less so with models that assume a specific moment of "lexical access". During sentence reading, our results suggest that articles could be processed together with the subsequent nouns, but that the consequences of this processing may not become apparent in eye-movement behavior until the sentence is syntactically integrated. In other words, it may not be more difficult to determine that "el mesa" is a piece of furniture, but it may be more difficult to know what role "el mesa" plays in a sentence compared to "la mesa".

References

Abbott, M. J., Angele, B., Ahn, Y. D., & Rayner, K. (2015). Skipping syntactically illegal the previews: The role of predictability. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(6), 1703–1714. https://doi.org/10.1037/xlm0000142

Angele, B., & Rayner, K. (2013a). Processing the in the parafovea: Are articles skipped automatically? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(2), 649–662. https://doi.org/10.1037/a0029294

Angele, B., & Rayner, K. (2013b). Processing the in the parafovea: Are articles skipped automatically? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(2), 649–662. https://doi.org/10.1037/a0029294

Aust, F., & Barth, M. (2023). *papaja: Prepare reproducible APA journal articles with R Markdown*. https://github.com/crsh/papaja

Barth, M. (2023). *tinylabels: Lightweight variable labels*. https://cran.r-project.org/package=tinylabels

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

Brysbaert, M. (2019). How many participants do we have to include in properly powered experiments? A tutorial of power analysis with reference tables. *Journal of Cognition*, *2*(1), 16. https://doi.org/10.5334/joc.72

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package brms. *The R Journal*, *10*(1), 395–411. https://doi.org/10.32614/RJ-2018-017

Bürkner, P.-C. (2021). Bayesian item response modeling in R with brms and Stan. *Journal of Statistical Software*, *100*(5), 1–54. https://doi.org/10.18637/jss.v100.i05

Cutter, M. G., Drieghe, D., & Liversedge, S. P. (2014). Preview benefit in English spaced compounds. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *40*(6), 1778–1786. https://doi.org/10.1037/xlm0000013

Dare, N., & Shillcock, R. (2013). Serial and parallel processing in reading: Investigating the effects of parafoveal orthographic information on nonisolated word recognition. *Quarterly Journal of Experimental Psychology*, *66*(3), 487–504. https://doi.org/10.1080/17470218.2012.703212

de la Cruz Cabanillas, I., Tejedor Martínez, C., Díez Prados, M., & Cerdá Redondo, E. (2007). English loanwords in spanish computer language. *English for Specific Purposes*, *26*(1), 52–78. https://doi.org/10.1016/j.esp.2005.06.002

Drieghe, D., Pollatsek, A., Staub, A., & Rayner, K. (2008). The word grouping hypothesis and eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(6), 1552–1560. https://doi.org/10.1037/a0013017

Eddelbuettel, D., & Balamuta, J. J. (2018). Extending R with C++: A Brief Introduction to Rcpp. *The American Statistician*, *72*(1), 28–36. https://doi.org/10.1080/00031305.2017.1375990

Eddelbuettel, D., & François, R. (2011). Rcpp: Seamless R and C++ integration. *Journal of Statistical Software*, *40*(8), 1–18. https://doi.org/10.18637/jss.v040.i08

Eddington, D., & Hualde, J. I. (2008). El abundante agua fría: Hermaphroditic Spanish Nouns. *Studies in Hispanic and Lusophone Linguistics*, *1*(1), 5–32. https://doi.org/10.1515/shll-2008-1004

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, *16*(1), 143–149. https://doi.org/10.3758/BF03203267

Forster, K. I., & Bednall, E. S. (1976). Terminating and exhaustive search in lexical access. *Memory & Cognition*, *4*(1), 53–61. https://doi.org/10.3758/BF03213255

Friederici, A. D., & Jacobsen, T. (1999). Processing Grammatical Gender During Language Comprehension. *Journal of Psycholinguistic Research*, *28*(5), 467–484. https://doi.org/10.1023/A:1023264209610

Gohel, D., & Skintzos, P. (2024). *Flextable: Functions for tabular reporting*. https://CRAN.R-project.org/package=flextable

Grolemund, G., & Wickham, H. (2011). Dates and times made easy with lubridate. *Journal of Statistical Software*, *40*(3), 1–25. https://www.jstatsoft.org/v40/i03/

Healy, A. F. (1994). Letter detection: A window to unitization and other cognitive processes in reading text. *Psychonomic Bulletin & Review*, *1*, 333–344. https://doi.org/10.3758/BF03213975

Iannone, R., Cheng, J., Schloerke, B., Hughes, E., Lauer, A., Seo, J., Brevoort, K., & Roy, O. (2024). *Gt: Easily create presentation-ready display tables*. https://CRAN.R-project.org/package=gt

Jeffreys, H. (1961). *The Theory of Probability* (3rd ed.). Oxford University Press.

Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods*, *42*(3), 627–633. https://doi.org/10.3758/brm.42.3.627

Koriat, A., & Greenberg, S. N. (1993). Prominence of leading functors in function morpheme sequences as evidenced by letter detection. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(1), 34–50. https://doi.org/10.1037/0278-7393.19.1.34

Koriat, A., & Greenberg, S. N. (1994). The extraction of phrase structure during reading: Evidence from letter detection errors. *Psychonomic Bulletin & Review*, *1*, 345–356. https://doi.org/10.3758/BF03213976

Lyons, C. (1999). *Definiteness*. Cambridge University Press.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of the effect of context in perception: Part 1. *Psychological Review*, *88*, 375407.

Meade, G., Declerck, M., Holcomb, P. J., & Grainger, J. (2021). Parallel semantic processing in the flankers task: Evidence from the N400. *Brain and Language*, *219*, 104965.

https://doi.org/10.1016/j.bandl.2021.104965

Morton, J. (1969). Interaction of information in word recognition. *Psychological Review*, *76*(2), 165178.

Müller, K. (2020). *Here: A simpler way to find your files*. https://CRAN.R-project.org/package=here

Müller, K., & Wickham, H. (2023). *Tibble: Simple data frames*. https://CRAN.R-project.org/package=tibble

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Pérez-Pereira, M. (1991). The acquisition of gender: what Spanish children tell us. *Journal of Child Language*, *18*(3), 571–590. https://doi.org/10.1017/S0305000900011259

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Radach, R. (1996). Blickbewegungen beim lesen: Psychologische aspekte der determination von fixationspositionen [eye movements in reading: Psychological factors that determine fixation locations]. *Münster/New York: Waxmann*.

Rayner, K. (1975). Parafoveal identification during a fixation in reading. *Acta Psychologica*, *39*(4), 271–281. https://doi.org/10.1016/0001-6918(75)90011-6

Rayner, K. (1979). Eye guidance in reading: Fixation locations within words. *Perception*, *8*(1), 2130.

Rayner, K., Slattery, T. J., Drieghe, D., & Liversedge, S. P. (2011). Eye movements and word skipping during reading: Effects of word length and predictability. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(2), 514–528.

Reicher, G. M. (1968). *Perceptual Recognition as a Function of Meaningfulness of Stimulus Material* [PhD thesis]. https://www.proquest.com/docview/302331273/citation/87A54446E91A453EPQ/1

Reichle, E. D., Liversedge, S. P., Pollatsek, A., & Rayner, K. (2009). Encoding multiple words simultaneously in reading is implausible. *Trends in Cognitive Sciences*, *13*(3), 115119. https://doi.org/10.1016/j.tics.2008.12.002

Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: Part 2. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review*, *89*(1), 6094.

Russell, B. (1905). II.—on denoting. *Mind*, *XIV*(4), 479–493. https://doi.org/10.1093/mind/XIV.4.479

Saint-Aubin, J., & Klein, R. M. (2008). The influence of reading skills on the missing-letter effect among elementary school students. *Reading Research Quarterly*, *43*, 132–146. https://doi.org/10.1598/RRQ.43.2.2

Schotter, E. R., & Payne, B. R. (2019). Eye movements and comprehension are important to reading. *Trends in Cognitive Sciences*, *23*(10), 811–812. https://doi.org/10.1016/j.tics.2019.06.005

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, *96*(4), 523–568. https://doi.org/10.1037/0033-295X.96.4.523

Slattery, T. J., & Yates, M. (2018). Word skipping: Effects of word length, predictability, spelling and reading skill. *Quarterly Journal of Experimental Psychology*, *71*(1), 250–259. https://doi.org/10.1080/17470218.2017.1310264

Snell, J. (2024). The reading brain extracts syntactic information from multiple words within 50 milliseconds. *Cognition*, *242*, 105664. https://doi.org/10.1016/j.cognition.2023.105664

Snell, J., Bertrand, D., Meeter, M., & Grainger, J. (2018). Integrating orthographic information across time and space. *Experimental Psychology*, *65*(1), 32–39. https://doi.org/10.1027/1618-3169/a000386

Snell, J., Cauchi, C., Grainger, J., & Lété, B. (2021). Attention extends beyond single words in beginning readers. *Attention, Perception, & Psychophysics*, *83*(1), 238–246.

https://doi.org/10.3758/s13414-020-02184-y

Snell, J., & Grainger, J. (2018). Parallel word processing in the flanker paradigm has a rightward bias. *Attention, Perception, & Psychophysics*, *80*(6), 1512–1519. https://doi.org/10.3758/s13414-018-1547-2

Snell, J., & Grainger, J. (2019). Readers are parallel processors. *Trends in Cognitive Sciences*, *23*(7), 537–546. https://doi.org/10.1016/j.tics.2019.04.006

Snell, J., Mathôt, S., Mirault, J., & Grainger, J. (2018). Parallel graded attention in reading: A pupillometric study. *Scientific Reports*, *8*(1), 3743. https://doi.org/10.1038/s41598-018-22138-7

Snell, J., & Simon, J. (2025). Readers encode absolute letter positions. *Cognition*, *254*, 105960. https://doi.org/10.1016/j.cognition.2024.105960

Staub, A., Dodge, S., & Cohen, A. L. (2019). Failure to detect function word repetitions and omissions in reading: Are eye movements to blame? *Psychonomic Bulletin & Review*, *26*(1), 340–346. https://doi.org/10.3758/s13423-018-1492-z

Vandendaele, A., Declerck, M., Grainger, J., & Snell, J. (2020). How readers process syntactic input depends on their goals. *Acta Psychologica*, *203*, 103006. https://doi.org/10.1016/j.actpsy.2020.103006

Vandendaele, A., Ortega, S. E., Midgley, K. J., Grainger, J., & Holcomb, P. J. (2025). Parallel Syntactic Processing in the Flankers Task: Insights From ERP Decoding. *Psychophysiology*, *62*(3), e70019. https://doi.org/10.1111/psyp.70019

Wechsler, S., & Zlatić, L. (2003). *The many faces of agreement*. Center for the Study of Language; Information. http://catdir.loc.gov/catdir/toc/ecip046/2003016201.html

Wheeler, D. D. (1970). Processes in word recognition. *Cognitive Psychology*, *1*(1), 59–85. https://doi.org/10.1016/0010-0285(70)90005-8

White, A. L., Boynton, G. M., & Yeatman, J. D. (2019). You can't recognize two words simultaneously. *Trends in Cognitive Sciences*, *23*(10), 812–814. https://doi.org/10.1016/j.tics.2019.07.001

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org

Wickham, H. (2023a). *Conflicted: An alternative conflict resolution strategy*. https://CRAN.R-project.org/package=conflicted

Wickham, H. (2023b). *Forcats: Tools for working with categorical variables (factors)*. https://CRAN.R-project.org/package=forcats

Wickham, H. (2023c). *Stringr: Simple, consistent wrappers for common string operations*. https://CRAN.R-project.org/package=stringr

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., 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., … Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2023). *Dplyr: A grammar of data manipulation*. https://CRAN.R-project.org/package=dplyr

Wickham, H., & Henry, L. (2023). *Purrr: Functional programming tools*. https://CRAN.R-project.org/package=purrr

Wickham, H., Hester, J., & Bryan, J. (2023). *Readr: Read rectangular text data*. https://CRAN.R-project.org/package=readr

Wickham, H., Vaughan, D., & Girlich, M. (2023). *Tidyr: Tidy messy data*. https://CRAN.R-project.org/package=tidyr

Xie, Y. (2015). *Dynamic documents with R and knitr* (2nd ed.). Chapman; Hall/CRC. https://yihui.org/knitr/

Yarkoni, T., Balota, D., & Yap, M. (2008). Moving beyond Coltheart's N: A new measure of orthographic similarity. *Psychonomic Bulletin & Review*, *15*(5), 971–979. https://doi.org/10.3758/pbr.15.5.971

Yasumoto, A. (2024). *ftExtra: Extensions for 'flextable'*. https://CRAN.R-project.org/package=ftExtra

Yu, L., Cutter, Yan, Bai, Fu, Drieghe, & Liversedge, S. P. and. (2016). Word n + 2 preview effects in three-character chinese idioms and phrases. *Language, Cognition and Neuroscience*, *31*(9), 1130–1149. https://doi.org/10.1080/23273798.2016.1197954

Zang, C. (2019). New Perspectives on Serialism and Parallelism in Oculomotor Control During Reading: The Multi-Constituent Unit Hypothesis. *Vision (Basel, Switzerland)*, *3*(4), 50. https://doi.org/10.3390/vision3040050

Zhu, H. (2021). *kableExtra: Construct complex table with 'kable' and pipe syntax*. https://CRAN.R-project.org/package=kableExtra

Table 1

Descriptive statistics of the target words used in Experiment 1.

	Target Word		
	N	Mean Zipf	SD Zipf
Feminine			
4 letters	75	4.57	0.58
5 letters	75	4.64	0.48
6 letters	75	4.65	0.46
Masculine			
4 letters	75	4.47	0.57
5 letters	75	4.88	0.40
6 letters	75	4.80	0.45

Note: Word frequency is given in Zipf ($\log_{10}$(fpmw) + 3, where fpmw is the frequency per million words).

Table 2

Accuracy as well as means and standard deviations of correct response times (in s) by flanker condition and grammatical gender of the target word in Experiment 1.

		Accuracy	Correct responses		
Grammatical Gender	Flanker Condition	% correct	N	Mean RT	SD
Masculine	Control	94.6	7577	0.495	0.169
Masculine	Congruent	95.4	7680	0.498	0.168
Masculine	Incongruent	91.9	7401	0.547	0.196
Feminine	Control	92.1	7358	0.482	0.159
Feminine	Congruent	94.3	7557	0.485	0.165
Feminine	Incongruent	89.1	7155	0.535	0.196

Note: *N* = number of correct responses; RT = response time (in s); SD = standard deviation

Table 3

Coefficients and 95% credible intervals from a Bayesian linear mixed model fitted to correct reaction times in Experiment 1, with flanker condition, target word grammatical gender, and their interaction as predictors for both the μ and the β parameters of the ex-Gaussian distribution. Coefficients for which the 95% credible interval does not include 0 are shown in bold.

	Estimate		Credible Interval (b)	
	b	SE	2.5%	97.5%
Parameter: μ				
Intercept	**0.505**	**0.009**	**0.487**	**0.522**
Control vs. Congruent Flanker	**-0.003**	**0.001**	**-0.006**	**0.000**
Control/Congruent vs. Incongruent Flanker	**-0.051**	**0.003**	**-0.056**	**-0.046**
Grammatical Gender (Feminine vs. masculine)	**0.006**	**0.002**	**0.003**	**0.009**
Control vs. Congruent:Grammatical Gender	0.000	0.002	-0.003	0.003
Control/Congruent vs. Incongruent:Grammatical Gender	0.000	0.002	-0.003	0.003
Parameter: β				
Intercept	**-2.049**	**0.032**	**-2.114**	**-1.987**
Control vs. Congruent Flanker	-0.004	0.015	-0.033	0.025
Control/Congruent vs. Incongruent Flanker	**-0.253**	**0.013**	**-0.279**	**-0.227**
Grammatical Gender (Feminine vs. masculine)	**0.028**	**0.007**	**0.014**	**0.041**
Control vs. Congruent:Grammatical Gender	0.023	0.015	-0.007	0.053
Control/Congruent vs. Incongruent:Grammatical Gender)	0.018	0.013	-0.007	0.044

Note: *b* = coefficient estimate; *SE* = standard error; *2.5%* and *97.5%* = lower and upper bounds of the 95% credible interval. Coefficients for which the 95% credible interval does not include 0 are shown in **bold**.

Table 4

Coefficients and 95% credible intervals from a Bayesian generalized linear mixed model using the Bernoulli distribution to predict response accuracy in Experiment 1, with target word grammatical gender, and their interaction as predictors. Coefficients for which the 95% credible interval does not include 0 are shown in bold.

	Estimate		Credible Interval (b)	
	b	SE	2.5%	97.5%
Parameter: θ				
Intercept	**3.011**	**0.072**	**2.872**	**3.154**
Control vs. Congruent Flanker	**-0.219**	**0.067**	**-0.353**	**-0.090**
Control/Congruent vs. Incongruent Flanker	**0.702**	**0.060**	**0.588**	**0.820**
Grammatical Gender (Feminine vs. masculine)	**0.139**	**0.045**	**0.051**	**0.227**
Control vs. Congruent:Grammatical Gender	**0.125**	**0.058**	**0.012**	**0.241**
Control/Congruent vs. Incongruent:Grammatical Gender	-0.032	0.048	-0.128	0.062

Note: $\theta = \log\left(\frac{p}{1-p}\right)$, where p is the probability of a correct response; b = coefficient estimate, *SE* = standard error, *2.5%* and *97.5%* = lower and upper bounds of the 95% credible interval. Coefficients for which the 95% credible interval does not include 0 are shown in **bold**.

Table 5

Descriptive statistics of the target words used in Experiment 2.

	Target Word				
	Zipf		OLD20		
	Mean	SD	Mean	SD	N
Feminine					
4 letters	4.60	0.61	1.13	0.24	60
5 letters	4.52	0.40	1.38	0.24	60
Masculine					
4 letters	4.49	0.59	1.12	0.21	60
5 letters	4.86	0.40	1.51	0.24	60

Note: Word frequency is given in Zipf ($\log_{10}$(fpmw) + 3, where fpmw is the frequency per million words). OLD20 is the mean Levenshtein distance between the target word and the 20 nearest neighbors in the lexicon.

Table 6

Accuracy as well as means and standard deviations of correct response times (in s) by flanker condition and grammatical gender of words in Experiment 2.

		Accuracy	Correct responses		
Grammatical Gender	Flanker Condition	% correct	N	Mean RT	SD
Masculine	Congruent	93.1	2935	0.439	0.135
Masculine	Incongruent	91.6	3047	0.456	0.151
Feminine	Congruent	93.3	3049	0.444	0.140
Feminine	Incongruent	92.7	2945	0.451	0.144

Note: *N* = number of correct responses; RT = response time (in s); SD = standard deviation

Table 7

Accuracy as well as means and standard deviations of correct response times (in s) by flanker condition and apparent grammatical gender of ***pseudowords*** *in Experiment 2.*

		Accuracy	Correct responses		
Apparent Grammatical Gender	Flanker Condition	% correct	N	Mean RT	SD
Masculine	Congruent	92.2	2950	0.491	0.149
Masculine	Incongruent	93.2	3038	0.488	0.145
Feminine	Congruent	89.8	2972	0.499	0.152
Feminine	Incongruent	91.0	3011	0.493	0.147

Note: *N* = number of correct responses; RT = response time (in s); SD = standard deviation

Table 8

Coefficients and 95% credible intervals from a Bayesian linear mixed model fitted to correct reaction times on word *trials in Experiment 2, with flanker condition, target word grammatical gender, and their interaction as predictors for both the μ and the β parameters of the ex-Gaussian distribution. Coefficients for which the 95% credible interval does not include 0 are shown in bold.*

	Estimate		Credible Interval (b)	
	b	SE	2.5%	97.5%
Parameter: μ				
Intercept	**0.446**	**0.010**	**0.427**	**0.467**
Congruent vs. Incongruent Flanker	**-0.005**	**0.001**	**-0.007**	**-0.003**
Grammatical Gender (Feminine vs. masculine)	0.000	0.002	-0.003	0.003
Congruent vs. Incongruent:Grammatical Gender	0.001	0.001	-0.001	0.003
Parameter: β				
Intercept	**-2.348**	**0.050**	**-2.446**	**-2.253**
Congruent vs. Incongruent Flanker	**-0.044**	**0.014**	**-0.071**	**-0.018**
Grammatical Gender (Feminine vs. masculine)	-0.001	0.016	-0.031	0.030
Congruent vs. Incongruent:Grammatical Gender)	0.017	0.013	-0.008	0.042

Note: *b* = coefficient estimate; *SE* = standard error; *2.5%* and *97.5%* = lower and upper bounds of the 95% credible interval. Coefficients for which the 95% credible interval does not include 0 are shown in **bold**.

Table 9

Coefficients and 95% credible intervals from a Bayesian linear mixed model fitted to correct reaction times on pseudoword *trials in Experiment 2, with flanker condition, pseudoword apparent grammatical gender, and their interaction as predictors for both the μ and the β parameters of the ex-Gaussian distribution. Coefficients for which the 95% credible interval does not include 0 are shown in bold.*

	Estimate		Credible Interval (b)	
	b	SE	2.5%	97.5%
Parameter: μ				
Intercept	**0.493**	**0.011**	**0.472**	**0.514**
Congruent vs. Incongruent Flanker	**0.003**	**0.001**	**0.001**	**0.005**
Apparent Grammatical Gender (F vs. M)	0.004	0.002	0.000	0.008
Congruent vs. Incongruent:Apparent Grammatical Gender	0.000	0.001	-0.002	0.002
Parameter: β				
Intercept	**-2.342**	**0.050**	**-2.440**	**-2.243**
Congruent vs. Incongruent Flanker	**0.042**	**0.013**	**0.017**	**0.068**
Apparent Grammatical Gender (F vs. M)	0.014	0.017	-0.019	0.048
Congruent vs. Incongruent:Apparent Grammatical Gender	-0.012	0.015	-0.042	0.017

Note: *b* = coefficient estimate; *SE* = standard error; *2.5%* and *97.5%* = lower and upper bounds of the 95% credible interval. Coefficients for which the 95% credible interval does not include 0 are shown in **bold**.

Table 10

Coefficients and 95% credible intervals from a Bayesian generalized linear mixed model using the Bernoulli distribution to predict response accuracy in Experiment 2, with target word grammatical gender, and their interaction as predictors. Coefficients for which the 95% credible interval does not include 0 are shown in bold.

	Estimate		Credible Interval (b)	
	b	SE	2.5%	97.5%
Parameter: θ				
Intercept	**2.905**	**0.091**	**2.731**	**3.087**
Congruent vs. Incongruent Flanker	0.066	0.048	-0.030	0.158
Grammatical Gender (Feminine vs. masculine)	0.045	0.065	-0.082	0.173
Congruent vs. Incongruent:Grammatical Gender	-0.033	0.040	-0.112	0.043

Note: $\theta = \log\left(\frac{p}{1-p}\right)$, where p is the probability of a correct response; b = coefficient estimate, *SE* = standard error, *2.5%* and *97.5%* = lower and upper bounds of the 95% credible interval. Coefficients for which the 95% credible interval does not include 0 are shown in **bold**.

Table 11

Coefficients and 95% credible intervals from a Bayesian generalized linear mixed model using the Bernoulli distribution to predict response accuracy for pseudowords in Experiment 2, with flanker condition, apparent pseudoword word grammatical gender, and their interaction as predictors. Coefficients for which the 95% credible interval does not include 0 are shown in bold.

	Estimate		Credible Interval (b)	
	b	SE	2.5%	97.5%
Parameter: θ				
Intercept	**2.907**	**0.110**	**2.694**	**3.123**
Congruent vs. Incongruent Flanker	-0.065	0.046	-0.154	0.028
Apparent grammatical Gender (F vs. M)	**-0.157**	**0.075**	**-0.307**	**-0.012**
Congruent vs. Incongruent:Apparent grammatical Gender	0.011	0.040	-0.066	0.092

Note: $\theta = \log\left(\frac{p}{1-p}\right)$, where p is the probability of a correct response; b = coefficient estimate, *SE* = standard error, *2.5%* and *97.5%* = lower and upper bounds of the 95% credible interval. Coefficients for which the 95% credible interval does not include 0 are shown in **bold**.

Table 12

Descriptive statistics of the target (non-animal) words used in Experiment 3.

	Target Word				
	Zipf		OLD20		
	Mean	SD	Mean	SD	N
Feminine					
4 letters	4.58	0.61	1.13	0.24	70
5 letters	4.68	0.48	1.41	0.25	69
6 letters	4.68	0.46	1.67	0.27	69
Masculine					
4 letters	4.49	0.59	1.17	0.26	66
5 letters	4.89	0.41	1.52	0.23	70
6 letters	4.82	0.45	1.70	0.19	72

Note: Word frequency is given in Zipf ($\log_{10}$(fpmw) + 3, where fpmw is the frequency per million words). OLD20 is the mean Levenshtein distance between the target word and the 20 nearest neighbors in the lexicon.

Table 13

Accuracy as well as means and standard deviations of correct response times (in s) by flanker condition and grammatical gender of non-animal target words in Experiment 3.

		Accuracy	Correct responses		
Grammatical Gender	Flanker Condition	% correct	N	Mean RT	SD
Masculine	Congruent	97.3	5576	0.409	0.136
Masculine	Incongruent	97.6	5626	0.409	0.130
Feminine	Congruent	96.7	5619	0.417	0.137
Feminine	Incongruent	97.0	5629	0.417	0.137

Note: *N* = number of correct responses; RT = response time (in s); SD = standard deviation

Table 14

Coefficients and 95% credible intervals from a Bayesian linear mixed model fitted to correct reaction times on non-animal trials in Experiment 3, with flanker condition, target word grammatical gender, and their interaction as predictors for both the μ and the β parameters of the ex-Gaussian distribution. Coefficients for which the 95% credible interval does not include 0 are shown in bold.

	Estimate		Credible Interval (b)	
	b	SE	2.5%	97.5%
Parameter: μ				
Intercept	**0.410**	**0.008**	**0.395**	**0.425**
Congruent vs. Incongruent Flanker	0.000	0.001	-0.001	0.001
Grammatical Gender (Feminine vs. masculine)	**0.004**	**0.001**	**0.002**	**0.006**
Congruent vs. Incongruent:Grammatical Gender	0.001	0.001	-0.001	0.002
Parameter: β				
Intercept	**-2.317**	**0.048**	**-2.412**	**-2.225**
Congruent vs. Incongruent Flanker	0.007	0.009	-0.010	0.024
Grammatical Gender (Feminine vs. masculine)	**0.039**	**0.010**	**0.019**	**0.058**
Congruent vs. Incongruent:Grammatical Gender)	0.011	0.009	-0.007	0.029

Note: *b* = coefficient estimate; *SE* = standard error; *2.5%* and *97.5%* = lower and upper bounds of the 95% credible interval. Coefficients for which the 95% credible interval does not include 0 are shown in **bold**.

Table 15

Coefficients and 95% credible intervals from a Bayesian generalized linear mixed model using the Bernoulli distribution to predict response accuracy on non-animal trials in Experiment 3, with flanker condition, target word grammatical gender, and their interaction as predictors. Coefficients for which the 95% credible interval does not include 0 are shown in bold.

	Estimate		Credible Interval (b)	
	b	SE	2.5%	97.5%
Parameter: θ				
Intercept	**2.907**	**0.110**	**2.694**	**3.123**
Congruent vs. Incongruent Flanker	-0.065	0.046	-0.154	0.028
Grammatical Gender (F vs. M)	**-0.157**	**0.075**	**-0.307**	**-0.012**
Congruent vs. Incongruent:Grammatical Gender	0.011	0.040	-0.066	0.092

Note: $\theta = \log\left(\frac{p}{1-p}\right)$, where p is the probability of a correct response; b = coefficient estimate, *SE* = standard error, *2.5%* and *97.5%* = lower and upper bounds of the 95% credible interval. Coefficients for which the 95% credible interval does not include 0 are shown in **bold**.

Appendix

Appendix A: Experimental materials

Experiment 1

Experiment 2

Experiment 3