PROCESSING GRAMMATICAL GENDER IN FRENCH

# PROCESSING GRAMMATICAL GENDER IN FRENCH: AN INDIVIDUAL DIFFERENCES STUDY

By MEAGAN NUCULAJ, B.A.

A Thesis Submitted to the School of Graduate Studies in the Partial Fulfillment of the Requirements for the Degree Master of Science

McMaster University MASTER OF SCIENCE (2023) Hamilton, Ontario (Cognitive Science of Language)

TITLE: Processing Grammatical Gender in French: An Individual Differences Study AUTHOR: Meagan Nuculaj, B.A. (McMaster University) SUPERVISOR: Dr. Ivona Kuçerová NUMBER OF PAGES: viii, 47

## Lay Abstract

Why is ‘pen’ masculine and ‘car’ feminine? Grammatical gender is a widespread feature of languages that comes naturally to native speakers and perplexes many second language learners. The assignment of gender seems to be random, but upon closer examination, patterns can be established. What do these differences mean for speakers of gendered languages? In the current study, we set out to determine how masculine and feminine grammatical gender is processed in French and how this is influenced by differences between individual speakers. Participants read French sentences that were either grammatical or contained a mismatch in gender between article and noun. Reading times were used to evaluate how speakers react when encountering an ungrammatical form with either masculine or feminine gender. Participants also completed tasks measuring response inhibition and verbal fluency to see how individuals with different cognitive and language skills react differently to unexpected forms.

**Abstract**

Past studies of grammatical gender have shown that native speakers encounter processing difficulties when encountering a form that does not agree in gender with previous words. However, the specific behavioral and neural responses to these difficulties have not been replicated across studies of the same type. This is in part due to different experimental designs and statistical analyses, but a crucial factor may be the lack of control between nouns of masculine and feminine gender in stimuli creation. Masculine and feminine gender show distinct distributional asymmetries and collapsing them into one condition diminishes the explanatory power of any study examining grammatical gender. We used reading times in a self-paced reading experiment to examine whether masculine and feminine gender violations differentially affect processing speeds. Fifty French speakers read sentences that were well-formed or contained a mismatch in gender between determiner and noun, half of which were masculine and half feminine. Following Beatty-Martínez et al. (2021), we added individual difference measures to determine how participant-specific factors modulate processing. Participants also completed a category verbal fluency task and the AX-CPT, a measure of cognitive control. They found that ERP components were modulated by these components for Spanish speakers and the modulation differed between masculine noun and feminine noun violations. We hypothesized that reading times would be similarly affected in French, a closely related language with the same gender categories. However, no conditions or interactions reached statistical significance. It is unclear whether this is due to the experimental manipulation or lack of control for participants' language background, as we had a high number of bilingual and multilingual participants. Regardless, elements of the procedure may provide insight on how to design future experiments that lay a groundwork in understanding the most basic elements of gender processing.

**Acknowledgements**

I would like to express my deepest gratitude to my supervisor, Dr. Ivona Kučerová, whose guidance shaped my understanding and nurtured my love for linguistic research. I would also like to sincerely thank Dr. Elisabet Service for her time, support, and invaluable help. Finally, I would like to voice my endless appreciation for Chelsea Whitwell, Ethan Stollar, and Steven Gonder, without whose programming expertise this experiment would not be possible.

**Table of Contents**

**List of Figures and Tables**

## 1. Introduction

Grammatical gender is a common noun classification system in which nouns are sorted into two or more distinct categories or genders that influence agreement on other parts of speech. Although these systems have a basis in conceptual or semantic gender, most nouns are assigned gender on an arbitrary basis (Corbett, 1991). Existing research says that the perception and production of grammatical gender is largely a matter of syntactic processing (Barber & Carreiras, 2005; Deutsch & Bentin, 2001; Gunter et al., 2000; Osterhout & Mobley, 1995), but lack of consistent and replicable studies call into question whether the answer is so straightforward.

Further, some languages may have transparent means (whether orthographic, phonological, morphological, or semantic) to predict the gender assignment of a given non-animate noun, others, such as French, are not so reliable (Lyster, 2006). This, along with distributional differences between masculine and feminine nouns within languages (Eddington, 2002) suggest that there may be important perceptional differences between genders in a single language. This study aims to understand these potential differences in French.

In section 2, I will recap current theories of gender assignment, with a focus on French. I will then go over the existing neural research in agreement processing as well as differences in experimental methodology. Finally, I will summarize a recent study that I chose to replicate based on my analysis of those previous topics. In section 3, I will lay out the methodology I utilized and summarize the results in section 4. Next, in section 5, I will discuss what the results can tell us about the split between masculine and feminine gender, individual differences in reading times, and sentence processing in general. Finally, I will make concluding remarks about the overall findings and future directions of this area of study in section 6.

## 2. Literature Review

### 2.1 Grammatical gender

First and foremost, it is important to establish a definition of grammatical gender. Corbett (1991) described grammatical gender as a way of sorting nouns into two or more classes reflected by agreement on other elements. This description separates grammatical gender from other forms of nominal classification, such as declension and classifiers. In this system, the most important factor in determining if a language has gender is the agreeing elements, such as verbs, adjectives, or determiners, even though gender itself is restricted to nouns. Kramer (2015) expanded on that, crucially adding a "semantic core generalization" to that definition, which distinguishes gender from other phi features, such as number and person. This generalization states that "grammatical gender is always assigned to at least a subset of nouns on the basis of animacy, humanness, and/or social gender for humans/biological sex for animals". These generalizations can capture all languages that are described to have grammatical gender. Some Germanic languages do not follow this generalization at first glance, such as Dutch or Swedish, which have the two genders "neuter" and "common", but this can be explained by the historic collapse of the masculine and feminine gender: words that belong to the common gender were once categorized as one of the two, with conceptual gender being the determiner of which category a noun would fall into when that distinction still existed (Trudgill & Lohndal, 2013 for Dutch; Van Epps et al. for Swedish and other Scandinavian languages). In contrast, the semantic basis of number is related to the cardinality of units denoted on the noun and the semantic basis of person involves the roles participants play in the relevant discourse. This semantic basis usually extends to all elements marked for these features, unlike gender where the mapping is more often arbitrary.

Out of 44% of documented world languages that have a gender system at all, 52% of those use some level of formal assignment, although these languages are mainly found in Eurasia and Africa, in the Indo-European, Afro-Asiatic and Niger-Congo families. Within the Indo- European family, most languages have the same two to three genders, which are masculine and feminine, plus neuter for the languages with three (Corbett, 2013).

In some ways, French is a classic example of a two-gendered language, but there are many things that make it unique, even among Romance languages. Like Spanish and Italian, gendered nouns in French can be broken down into four categories: semantic, morphological, referential, and syntactic. Semantic, sometimes referred to as inherent gender, is based on conceptual gender (e.g., un homme 'a man', une femme 'a woman'). Morphological gender arises from a single lexical entry, but bears two separate morphological forms (e.g., américain 'American-MSC',

américaine 'American-FEM'). Alternately, nouns with referential gender (also called épicènes) arise in a singular form, but can be used with either gender depending on the referent (e.g., un/une artiste 'an-MSC/an-FEM artist-MSC/FEM'). For all of these types of gender, there is a semantic component determining the gender of the word and a syntactic component that is involved in concord: all nouns must agree with relevant determiners and adjectives. Importantly, there is also syntactic gender, which is purely a relational gender between phrasal constituents where the mapping between noun form and gender is completely arbitrary. (Ayoun, 2010).

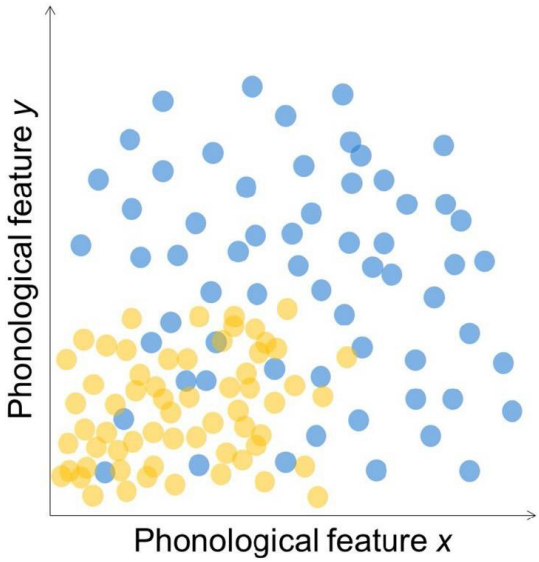
Unlike Spanish and Italian, French is very phonologically opaque, which makes it especially difficult to determine a word's gender by its orthography (Hull, 1968). In French, there are often many silent letters and the mapping between orthographic form and pronunciation is often complex, for example oeufs 'eggs' being pronounced with a single vowel sound. Even though graphemes do not reliably map to phonemes in French, the spelling of most words can predict the pronunciation if grouped correctly. There are usually many different written forms mapping to a single sound (Levitt, 1979). When looking at nouns with purely syntactic gender, there is more predictability when using phonological endings, though, with corpus analyses contending that the gender of 81% of all feminine nouns and 80% of all masculine nouns can be systematically predicted by their phonological endings (Lyster, 2006). This is considerably above chance, but relatively low compared to 98% for feminine nouns and 96% for masculine nouns in Spanish (Clegg, 2011).

Although statistically, most French nouns are gender predictable, with phonology being the most reliable predicter in syntactically gendered nouns, there are additional difficulties when mapping gender to French nouns when acquiring the language. Compared to many languages where gender is marked on both singular and plural determiners, French only marks the singular. Moreover, French has a property called "liason" which means that gender is not explicitly marked on determiners preceding words beginning with a vowel sound (e.g., le vélo 'the-MSC car-MSC compared to l'avion 'the airplane-FEM'). Surprisingly, this does not seem to make a significant difference for L1 learners, considering most children make almost no errors of gender attribution by the age of 3 (Clark, 1985), but the lack of explicitly gendered tokens does seem to affect L2 learners and could have an overall effect on processing.

As discussed above, both masculine and feminine nouns in each language have approximately equal predictive power towards their gender. Both Spanish and French also have about an equal split between masculine and feminine nouns (Bull, 1965; Roché, 1992). Despite this, there are key distributional differences. As illustrated in figure 1, feminine nouns in Spanish

have a closer range of variance in phonological form than masculine nouns do.

**Figure 1: Category structure**




Hypothetical illustration of feminine (yellow) and masculine (blue) category structure. Exemplars closer to the x- and y-intercept represent members whose phonemic makeup is more prototypical of the category, and those departing from it represent members sharing less phonological characteristics in common (Beatty-Martínez et al., 2021).

This difference in schematicity fits with a theory that masculine gender is the "default" in Spanish. This was suggested to be the case by Beatty-Martínez et al. (2019) based on the distributional data as well as behavioral evidence. For example, words of Greek origin that end in -a are usually masculine, despite this clashing with the form of regular Spanish words and nominalizations of nouns are masculine as well. Additionally, native Spanish speakers show faster reading times for feminine relative to masculine nouns in the processing of long-distance agreement dependencies (López Prego, 2015). This may suggest that speakers do not store the gender of a noun when it is default and retrieve the gender later on, when it becomes relevant in agreement relations. They are also less accurate at detecting gender violations in masculine nouns, as compared to feminine nouns when performing under increased processing demands (López Prego & Gabriele, 2014). Spanish children are more likely to attribute masculine gender to irregular nouns (Pérez-Pereira, 1991) and native speakers prefer masculine when assigning gender to loanwords (De la Cruz Cabanillas et al., 2007).

Masculine nouns in French also appear to have more variance than their feminine counterparts. The opposite has been argued in the past, citing the fact that there are 30 predominantly feminine suffixes compared to only 14 masculine ones (Surridge, 1986). However, looking at phonological data shows a pattern much closer to that seen in Spanish. Tables 1 and 2 (from Ayoun, 2010; compiled from Tucker et al. 1977) show final phones in French, corresponding with their multiple orthographic representations and the percentage of tokens that

coincide with the more commonly realized gender for that final phone. Of the 25 phones that predict greater than chance (61% and above), 15 are more likely to be masculine. Consonantal final phones are more evenly split, with 8 final phones (associated with 27 orthographic endings) predicting feminine and 6 (25 orthographic endings) predicting masculine. However, the consonantal final phones with the highest predictive power are still mostly masculine. Vocalic final phones are more reliable predictors as a whole and they overwhelmingly lean masculine, with 9 final phones (associated with 91 orthographic endings) predicting masculine and only 2 (16 orthographic endings) predicting feminine.

**Table 1: Consonantal final phones**

| phone | spelling | #tokens | M | F | % |
|---|---|---|---|---|---|
| [ʒ] | -j, -ge, -ges | 1453 | *1368* | 85 | 94% |
| [m] | -m, -me, -mes | 1406 | *1292* | 114 | 92% |
| [z] | -se, -ze | 612 | 61 | *551* | 90% |
| [r] | c+re, +res, v+[r]+c | 512 | *417* | 95 | 81% |
| [f] | -f, -fs, -fe, -fes, -phe, -phes | 131 | *101* | 30 | 77% |
| [g] | -g, -gs, -c, -gue, -gues | 235 | *172* | 63 | 73% |
| [v] | -v, -ve, -ves | 143 | 45 | 98 | 69% |
| [j] | -v+il, ille, illes | 352 | 114 | *238* | 68% |
| [n] | -n, -ne, -nes | 1135 | 358 | *777* | 68% |
| [d] | -d, -de, -des | 668 | 227 | *441* | 66% |
| [ʃ] | -ch, -sh, -che, -ches | 290 | 99 | *191* | 66% |
| [b] | -b, -be, -bes, -bbe, -bbes | 129 | *84* | 45 | 65% |
| [ɲ] | -gne, -gnes | 69 | 27 | *42* | 61% |
| [s] | -s, -ss, -x, -ce, -se, -xe, -ces | 1379 | 531 | *848* | 61% |
| [l] | -l, -ls, -le, -les, -lle, -lles | 1126 | 474 | *652* | 58% |
| [k] | -c, -cs, -ch, -chs, -ck, cks, -q, -que, -ques | 609 | *333* | 276 | 55% |
| [p] | -p, -pe, -pes | 214 | 104 | *110* | 51% |
| [t] | -t, -te, -tte, -the, -tes, -ttes, -thes | 2269 | *1162* | 1107 | 51% |

**Table 2: Vocalic final phones**

| phone | spelling | #tokens | M | F | % |
|---|---|---|---|---|---|
| [œ̃] | -um, -un, uns, -unt | 17 | 17 | 0 | 100% |
| [ɑ̃] | -an, -anc, -and, -anf, -ang, -aon, -amp, -ans, -ancs, -amps, -ant, -end, -eng, -ens, -ends, -ems, -empt, -ent, -ents, -ants | 1963 | *1949* | 14 | 99% |
| [ɛ̃] | -aim, -ym, -én, -ien, -en, -ain, -ein, -in, -aing, -oing, -eing, -éens, -iens, -ains, -eins, -ins, -inct, -ingt, -ient, -aint, -eint, -int | 938 | *929* | 9 | 99% |
| [ø] | -oeud, -eue, oeufs, -eut, -eu, -eux | 189 | *184* | 5 | 99% |
| [o] | -o, -oc, -op, -os, -ots, -ot, -aud, -aut, -ault, -au, -aux, -eaux, -aulx | 865 | *841* | 24 | 97% |
| [y] | -u, -ul, -us, -uts, -ut, -ux, -ue, -ues | 201 | *195* | 6 | 95% |
| [u] | -ou, -ouc, -oul, -oo, -oup, -ous, -out, -ouls, -oux, -oue, -oues | 171 | *150* | 21 | 88% |
| [wa] | -oi, -oids, -ois, -oigt, -oit, -oix, -oie, -oies, oye | 179 | *153* | 26 | 85% |
| [a] | -a, -ac, -ap, -ats, -as, -at | 791 | *648* | 143 | 82% |
| [i] | -ic, -id, -il, -is, -it, -ix, -i, -y, -ys, -ie, -ies, -ye | 2337 | 575 | *1762* | 75% |
| [õ] | -on, -om, -on+c, -om+c | *2668* | 794 | 1874 | 70% |
| [e] | -é, -és, -ée, -ées, -er, -ers, -ez, -ai, -ais, -ait, -aits, - aix, -aie, -aies, -ay, -et, -êt, -ès, -ect, -ects, -ey, -egs | 3416 | *1962* | 1454 | 57% |

Nelson (2005) also used theoretical evidence to propose a default gender in French, positing that all nouns in French are masculine unless given reason not to be. Certain morphemes, phones, and semantic categories can be strongly associated with feminine gender or masculine gender, while others do not reliably map to either. Of those strong associations, each is theorized to have a value (+m for masculine denoting, +f for feminine denoting). If the factors that are +f outweigh those that are +m, the noun will overwhelmingly be feminine. Alternately, if they are equal, predominantly +m, or not present at all, the noun will most likely be masculine. For example, categories such as bones, books, and shapes are considered to be masculine, as the vast majority of category members carry masculine gender, whereas forms of precipitation, phonetic terms, and lines gestalt are considered to be feminine for the same but opposite reason. There are also several consonantal phonological endings that are considered to be feminine, similar to the ones outlined in table 1. They use a principle of economy to argue that phonological endings that map more frequently with masculine are not considered to be a masculine marker by this analysis, but rather an absence of marking. This patterns with the discrepancy in variance between masculine and feminine gender, noted above.

When a word belongs to a category associated with a gender and has no competing factors, it will be that gender. However, when a word has neither category nor form that is associated with either gender, it is masculine. When a noun has competing factors, such as a masculine category, but feminine marking, it is also masculine. See (1)-(4) for examples.

(1) *tuba* 'tuba' [no rules apply] 0m 0f= m
(2) *chant* 'song' [superordinate =m] lm 0f= m
(3) *trajectoire* 'trajectory' [line gestalt= f] 0m 1f = f
(4) *dividende* 'dividend' [finance= m; nasal+ consonant= f] lm lf=m

The complete set of rules can account for between 90 to 98% of nouns in French. Although that is a convincing number, areas of the reasoning are unconvincing. There is a degree of creativity in assigning nouns to certain categories that does not seem to have a strong empirical basis. For example, one of the rules states that superordinate nouns are +m, while nouns of specificity are +f. However, it is unclear why *but* 'aim, goal' (defined as: anything that one strives to reach) is classified as a superordinate (although an "improper" one, as its subtypes admittedly cannot be defined in an objective manner) and given a +m value, while *cible* 'target' (defined as: an object that one shoots at with a bow and arrow or firearm) is classified specific noun. Meanwhile, *secours* 'aid' (defined as: anything given to help a person) is not considered specific or superordinate, but *aumone* 'alms, charity' (anything given to help a poor person) is considered specific. It is possible that certain category descriptions and judgments as to which category nouns belong could be influenced by the rest of the category and the urge to make things fit.

This could be cleared up by behavioral evidence. Like Spanish, loan words and nominalizations of verbs in French are generally masculine. Unlike Spanish, there is not much else evidence because less studies are done in French. Given no theoretical or behavioral evidence that feminine could be the default, it is most likely that masculine is the default, but it is also possible that French has no default at all.

When a certain gender has a default status, it is considered the unmarked form, whereas the non-default can be considered marked. It is important to note that evidence suggests that gender is stored in the lexical entry of a noun (Corbett, 2000). That means that once a noun is learned, its gender is stored with it and phonological and morphological cues are not necessary in agreement processing. However, speakers may still have prototypical forms stored that may become relevant when encountering a novel word or unexpected agreement patterns. This has important implications

on processing, which I will expand on at the end of the next section.

### 2.2: ERP methods to understand agreement processing

Past behavioral studies have suggested that grammatical gender must be acquired for individual lexical items (Ayoun, 2017) and that determiners play a large role in setting up expectations for a succeeding noun (Jakubowicz & Faussart, 1998). Despite this, there are not many recent experiments on L1 processing and much of the evidence is not conclusive. It may be the case that behavioral methods alone do not provide enough information to strongly validate claims of processing strategies, but this has not yet been investigated thoroughly.

In current research, electroencephalograms (EEGs) are used to provide the temporal resolution necessary to distinguish different levels of processing as well as determine whether different agreement patterns show any differences in neural activation. To interpret EEG data, it is crucial to pay attention to specific, well-studied ERP components. The components that are most relevant to the present topic are the N400, LAN, and P600. The N400 is a centro-parietal electrical negativity that occurs 400ms after stimulus onset and is said to reflect processing of semantic errors or general lack of predictability, possibly due to difficulty in integrating a particular word in the preceding linguistic context (Kutas & Hillyard, 1980). On the other hand, syntactic errors are thought to be represented by both the LAN and the P600, often in conjunction. The LAN (late anterior negativity) is generally elicited by elements that turn correct sentences into incorrect ones and can vary in timing depending on the type of error present, from word category to morphosyntax, whereas the P600 is a late centro-parietal positivity elicited by both syntactically incorrect and syntactically unusual sequences and is generally preceded by the LAN only in the incorrect cases. This suggests that LAN reflects the actual detection of a morphosyntactic mismatch, while P600 represents reprocessing and repair (Friederici, 1995). Furthermore, the P600 could also reflect more controlled processes, as it is shown to be affected by a broad range of task variables including semantics and working memory (Gunter et al., 1997).

The most common paradigm used to elicit these components in sentence processing experiments is the rapid serial visual presentation (RSVP) paradigm. In this paradigm, participants are rapidly presented sentences one word at a time, with the word being centered in the middle of the screen, often separated by fixation crosses. This has the drawback of being an unnatural way to read sentences, as there are usually no real-world scenarios where information is presented in this matter. However, it performs the crucial function of minimizing eye movements, which create large interferences in EEG data. It also makes it easy to focus analysis on a single target word.

(Raymond et al., 1992).

Past RSVP studies looking into agreement errors have elicited a host of ERP components; P600 seems to be the most reliable (Molinaro et al., 2011), but some studies have also elicited the N400 (Barber & Carreiras, 2003; Bhattamishra et al., 2019; Deutsch & Bentin, 2001; Guajardo & Wicha, 2014; and others) or the LAN (Alemán Bañón & Rothman, 2014; Beatty-Martinez et al., 2021; Caffarra et al., 2015; and others), with a biphasic LAN-P600 response being the most common.

Additionally, there has not been a consensus regarding the differences between number, gender, and person errors or combined errors (Alèman Bañón & Rothman, 2014; Barber & Carreiras, 2003; Nevins et al, 2007; and others). Grammatical agreement processing is mainly believed to be a syntactically driven operation (Frazier & Clifton, 1996). However, there is a very clear semantic basis as number, person, and gender all have real-world meanings that transcend strictly grammatical categories, even if this is not reflected on every word. Given that information, it is surprising that there is not a consistent, replicable component that can be associated with agreement errors, but the results seem to vary and not in a wholly predictable way. This may be due to differences in the design and purposes of existing experiments examining agreement. A1 in the appendix directly compares results across multiple studies which I will expand upon below.

Crucially, morphological marking of grammatical gender is not something that exists in isolation. Neurological studies of gender processing vary largely in their treatment and control of multiple factors. Firstly, there is a great number of ways to manipulate the stimuli, such as differences in the syntactic category of the feature being manipulated, the syntactic distance between the trigger word and the target word, and the presence or absence as well as location of the surrounding structure. There are also features more directly tied to gender marking that affect agreement processing, such as regularity, phonological transparency, animacy, and markedness. I will go over each of these and how they may possibly affect the results of a study. Languages differ in which elements are marked for gender, as well as their word order. For this reason, experiments on gender may choose to manipulate the determiner, adjective (post- or pre- nominal), or verb and consequently measure the ERP on either the noun, post-nominal adjective, or verb. These linguistic elements also differ in their syntactic position and distribution, so it is possible that this could influence processing. Of particular importance, the gender of a noun is inherent and fixed, whereas other elements are assigned their gender based on agreement with the noun. For this reason, it is most reliable to measure agreement on the noun (Beatty-Martínez et al., 2021).

Past studies have also looked at the difference in word pairs versus full sentences and found

increased responsivity to sentences (Barber & Carreiras, 2005), suggesting a role being played by sentence-level syntactic factors. Differences have also been found in relation to a target word's position within a sentence, with violations showing greater P600 amplitude to errors in the middle of a sentence compared to the beginning (Barber & Carreiras, 2005) and greater N400 amplitude at the end of the sentence compared to the middle (Hagoort, 2003). There is also a difference in processing for agreement pairs that are within phrases compared to those across phrases (Aléman Bañon, 2010). This is likely due to the amount of context available at different points of processing. A sentence context is intrinsically richer than a word context and expectancies tend to build up towards the end of sentences, as more information is added and there are less options on how to fill the gap. Overall, there are many sentence-spanning effects that must be considered before drawing conclusions based of manipulations of agreement elements.

As for the elements affecting agreement directly, one important factor is regularity. Many languages such as Italian, and Portuguese have reliable patterns of noun-endings that correlate with certain grammatical genders. If the majority of nouns of a certain gender carry affixes of a certain form, it can be considered the 'regular' form. This is true of nouns and adjectives. Anything that does not map with the morphological form of the majority could be considered irregular. It would theoretically create a large expectation to see that morphological form after (a) encountering a determiner signalling that form or (b) encountering a form such as a noun (or adjective – depending which comes first in said language) that is already morphologically marked. For this reason, it is easy to assume that regular and irregular forms would show differential processing latencies or amplitudes from each other. When learning a language, regular forms are considered easier to acquire (Pinker, 1995), although this does not necessarily map to its ease of processing. One study in Italian (Caffarra et al, 2015) compared regularly and irregularly marked gendered words and their agreement with determiners. They found an increased amplitude of the P600 and LAN in a violation condition with regular nouns. However, a similar study on Brazilian Portuguese compared irregular forms in both determiner-noun pairs and noun-adjective pairs and contrastingly found that the P600 was higher for irregular words, but only in the noun-adjective condition (de Resende et al., 2019). This could be in part due to the fact that the unusual word, in this case the irregular noun, showed up before the disagreeing element, in this case the adjective. The noun-adjective condition was also preceded by a determiner, which agreed with the noun, so this likely sets up a higher expectation than sentences with only one gendered element. It is still unclear why there was no result for the determiner-noun condition when it was seen in Italian. Here, there could be a difference between the languages themselves and their own level of regularity. Nevertheless,

it is the start of a pattern of entangled responses that are found in studies of gender agreement.

Related to regularity is the concept of phonological transparency. In some languages, notably French, a word such as an adjective may be marked orthographically for agreement, but the phonological realization does not convey this marking. Behavioral studies have shown a clear difference in processing between phonologically transparent or opaque agreement elements (Largy & Fayol, 2001; Frisson & Sandra, 2002), with shorter reaction times for written errors associated with oral cues. An ERP study also showed larger P600 components to such violations in French (Frenck- Mestre et al., 2009). These studies were all performed using binary stimuli, but there may be more of a continuum of both reliability and noticeability of phonological gender markers in certain languages.

Next, there is the factor of animacy – both conceptual and grammatical. Historically, grammatical gender in Proto-Indo-European languages is believed to be based in systems of animacy, where animate noun classes split into feminine and masculine (Brugmann, 1891; István, 1959). Some languages such as Russian, Polish, and Czech still have animacy divisions encoded into their grammatical gender systems (Riley, 1999). These divisions further challenge the view of gender agreement as a purely syntactic beast, as the assignment of otherwise arbitrary gender is based on real-world semantic information. ERP studies using these languages are sparse, but even outside of languages where it is grammaticalized, animacy seems to affect agreement processing. In Hebrew, a higher N400 response has been observed in violation conditions involving animate nouns (Deutsch & Bentin, 2001). In Hindi, N400 responses were found only in violation conditions involving inanimate nouns, with P600 responses being observed in violation conditions with animate ones (Bhattamishra et al, 2019), although the N400 in this study was notably later than usual. It is also worth noting that both of these studies only used masculine trigger nouns. Regardless, there appears to be some sort of semantic modulation occurring to agreement syntax coming from the notion of animacy. Understanding the processing of languages that use animacy as a grammatical element may prove crucial in the understanding of that modulation as a whole.

Finally, there is the factor of markedness, as discussed in section 2.1. Certain subsets of features can be considered the default form. They are often considered syntactically less complex and as such, simpler to process. In contrast, forms that differ from the default carry more information and are considered marked. In section 2.1, I provided some rationale for masculine being the default and unmarked form, at least in Spanish. More reliably, singular nouns can be considered unmarked when compared to plural nouns because they are often not inflected at all (Deutsch & Bentin, 2001). Like phonological cues, morphological cues may differ in their

reliability of associations and therefore modulate ERP components elicited in response to unexpected or unacceptable forms. In Hebrew, Deutsch & Bentin (2001) found larger P600 responses to number-marked predicates compared to unmarked forms in violation conditions. Alemán Bañón & Rothman (2016) found that the P600 occurred earlier in violation conditions for all marked elements in Spanish, but only increased in amplitude for number-marked elements, possibly due to lack of inflection on the singular. These differences prove problematic while analyzing data from past studies, as gender is often collapsed as a uniform construct where either masculine and feminine stimuli are examined together or only masculine triggers are used as a form of "control" (Beatty-Martinez & Dussias, 2019). This may be an essential component contributing to the lack of consistency across gender violation studies, especially considering the distributional differences discussed in section 2.1. As of writing, very few researchers have split their stimuli between masculine and feminine forms.

### 2.3 Brain potentials reveal differential processing of masculine and feminine grammatical gender in native Spanish speakers

One study by Beatty-Martínez et al. (2021) in Spanish examined masculine and feminine errors separately. Their experimental design could hold a key to the clearer understanding of gender processing. They focused their study first and foremost on gender processing, which eliminated many of the confounds in section 2.2 that resulted from manipulating components other than gender. They also employed individual difference measures to overcome possible statistical errors from past studies and gain a more nuanced understanding of native speakers' variable strategies when encountering gendered nouns. Recent studies have shown that the attested LAN-P600 response may in some instances be due to a statistical error when examining grand average waveforms. Some individuals display centrally distributed N400 effects and others display right-hemisphere dominant P600 effects. When averaged together across multiple participants, that appears as a biphasic LAN-P600 response (Tanner, 2015). When examining waveforms on an individual level, it appears that native speakers' responses fall on a continuum from preponderant-negative to preponderant-positive (Kim et al., 2018; Qi et al., 2017; Tanner, 2019). This variability has been associated with individual differences in both linguistic and cognitive domains (Batterink & Neville, 2013; Kim et al., 2018; Morgan-Short et al., 2012; Pakulak & Neville, 2010; Pélissier, 2020; Qi et al., 2017; Zirnstein et al., 2018).

To understand this relationship, Beatty-Martínez et al. included three tasks in their experiment. Their main task was a sentence reading task designed to measure agreement

expectancies. They also included two individual difference tasks: one measuring category verbal fluency and another measuring cognitive control.

For the category verbal fluency task, participants were asked to verbally generate as many exemplars of a given semantic category (e.g. musical instruments, vegetables) as they could in 30 seconds. This was intended to measure vocabulary size under the theory that individuals who could name more exemplars would likely have greater experience with correspondences between gender and noun form, as individuals who have a larger vocabulary are more likely to have greater depth of knowledge of those words (Schmitt, 2014). These participants would theoretically have a more defined schematicity between the two gender categories compared to participants who did not perform as well on the task. However, it could also reflect categorical organization of semantic memory. If this is the case, it is possible that an individual's ability to mentally categorize tokens of semantic meaning could reflect their ability to categorize other aspects of language, such as grammatical gender.

For the cognitive control measure, they used a distractor version of the AX-continuous performance task. In this task, participants were shown strings of five letters. First, a cue, which was either 'A' or 'B', followed by three distractor letters, and finally a probe, which was either 'X' or 'Y'. They were instructed to press 'no' after every cue or distractor letter. At the probe, they were instructed to only press 'yes' if it was an 'X' probe following an 'A' cue. 70% of all trials followed the A-X sequence, with the remaining trials being evenly split between the other three conditions, as shown in Table 3. All letters were shown for 300ms with 1000ms interstimulus interval.

**Table 3: AX-CPT task**

| Cue | Example Distractors | Probe | Correct Answer at Probe | Frequency |
|---|---|---|---|---|
| A | C G P | X | Yes | 70% |
| B | W H F | Y | No | 10% |
| A | M L D | X | No | 10% |
| B | V C S | Y | No | 10% |

The uneven frequency of correct trials creates an expectation of the appearance of the X-probe followed by an A-cue, which creates a false expectancy when a Y-probe follows the A-cue. This is said to pit effects of reactive response inhibition against the effects of proactive goal maintenance

and conflict monitoring. Participants who have longer response times and higher error rates are thought to heavily rely on cue information. This is analogous to the expectancies set up by gendered determiners in the agreement expectancy task.

For the agreement expectancy task, they constructed 240 semantically low-constraint sentences using masculine-feminine noun pairs. All sentences followed the pattern of DETERMINER - SUBJECT NOUN - VERB - DETERMINER 2 - OBJECT NOUN - PREPOSITIONAL PHRASE. They split these sentences into four conditions: Congruent Masculine, Incongruent Masculine, Congruent Feminine, and Incongruent Feminine. In the congruent conditions, the second determiner agreed in gender with the object noun. The gender of the determiner was switched in incongruent trials, creating an ungrammatical sentence. Table 4 shows an example quartet, with the target nouns bolded.

**Table 4: Stimuli**

| Condition | Determiner | Noun | Sample Trial |
|---|---|---|---|
| Congruent Masculine | Masculine | Masculine | La señora compró el **vestido** en la tienda. |
| Incongruent Masculine | Feminine | Masculine | La señora compró la **vestido** en la tienda. |
| Congruent Feminine | Feminine | Feminine | La señora compró la **cartera** en la tienda. |
| Incongruent Feminine | Masculine | Feminine | La señora compró el **cartera** en la tienda. |

The sentences were split into four experimental lists and presented, along with fillers, to participants using the rapid serial visual presentation (RSVP) paradigm, where sentences were quickly displayed one word at a time at the centre of the screen. Participants were instructed to read for comprehension and asked a picture relatedness question at the end of every sentence. ERP components were measured on the target noun.

The researchers predicted that the incongruent masculine condition would show a more reliable LAN effect than the incongruent feminine condition, because the feminine determiner is a more marked cue, leading to higher prediction error. They also predicted that differences between

participants would be modulated by the individual differences measures. The category verbal fluency task measured linguistic knowledge and experience, which could affect reanalysis and integration. The AX-CPT reflects reliance on cues, which could affect an individual's response when encountering an unexpected form after a cue of varying reliability.

Both predictions proved correct. Masculine and feminine violations showed LAN and P600 responses across individuals, but the LAN effect was larger for masculine nouns and the P600 was more broadly distributed for feminine nouns. The smaller LAN for feminine nouns may suggest that feminine gender cues preceding masculine-noun violations may have induced greater prediction error, possibly due to the markedness of the feminine determiner. This patterns with the greater P600 in the incongruent feminine condition, as it has been shown to be more sensitive to errors with more robust cues (Mehravari et al., 2015).

Individual responses ranged on a continuum between negative-dominant, positive-dominant and biphasic, with an inverse relation between the negative and positive (participants who exhibited a strong LAN response had a weaker P600 response, and vice versa). There was also an inverse relation between LAN responses between the two gender conditions as well as a positive correlation between the LAN effect for masculine nouns and the P600 effect for feminine nouns. However, a P600 response for masculine nouns did not reliably predict positivity or negativity for feminine nouns.

As for the individual difference measures, participants with higher verbal category fluency were more likely to have a P600-dominant response pattern to unexpected feminine nouns. Conversely, participants with longer reaction times on the AX-CPT showed a LAN-dominant response to unexpected masculine nouns.

They concluded that the high levels of individual differences between participants and the lack of standard LAN-P600 patterns challenge the "native norm" used to compare monolingual speakers to bilinguals and second language learners. Further, they suggest that the P600 can be viewed as an index of category exemplar strength, as there are key differences in representational strength between the two gender categories, as shown by the modulation of the P600 by verbal fluency score on feminine noun errors only. They also suggest that participants with better inhibitory control ability also experienced less difficulty associated with encountering a prediction error. Importantly, they emphasize the importance on future language processing research examining masculine and feminine gender categories as separate entities.

## 2.4 Aims of the Present Study

The prior study raises questions about previous theories of the neural processing of grammatical gender, both in questioning the standard of the LAN-P600 pattern and introducing a new possibility for the P600. As for theory, it provides neural evidence of the difference between feminine and masculine in processing and offers possible usage-based differences to account for differences between speakers. It introduces novel experimental methods that can potentially greaten the understanding of so-called "native norms" in sentence processing. It has yet to be proven if these individual differences extend to the behavioral realm or pattern similarly in other languages. To tackle both of those questions, I conducted a behavioral replication of this study in French. French is a good candidate language because it is typologically very similar to Spanish and the stimuli can largely be translated using the same word order. However, there are also notable differences, particularly weaker cue reliability as discussed in section 2.1. As French speakers encounter less noun tokens where the gender is visibly or audibly marked on the determiner, they may not have as strong of a mapping between nouns and their genders.

The two most common behavioral methods for studying sentence processing are eye-tracking and self-paced reading. Eye-tracking studies measure the gaze and movements of the eyes when a participant is presented with a chunk of text (Popa et al., 2015). Whereas self-paced reading displays sentences one word at a time and allows participants to move through words at their own speed. Eye-tracking is more natural, as it allows participants to process sentences using similar strategies that they would use in everyday reading. It also allows researchers to examine both online processing by measuring things like time of first fixation, which occurs immediately when you see a word, and offline processing by measuring things like looking back at a word after moving past it in the passage, which corresponds to late processing of a word (Liversedge et al., 1998). Self-paced reading studies are more limited to online processing, but have other benefits. For example, in eye-tracking studies, it is common for participants to skip past small words or functional words (Drieghe et al., 2005). This can be particularly detrimental to outcomes when looking at two-letter determiners, as is common in studies of grammatical gender in Romance languages. Self-paced reading studies also closely mirror the RSVP paradigm that is needed in EEG studies, making it possible to use the same stimuli between behavioral and neural studies. This allows researchers to directly compare results without as many added confounds.

I employed the same verbal fluency and AX-CPT tasks as well as a self-paced reading task. This mirrored the RSVP task in Beatty-Martínez et al. (2021) except participants were

measured on reaction time, as they were moving through the words at their own pace. Although this would not have correlates to the negative- dominant and positive-dominant responses seen in the ERP study, it could be used to evaluate the differences in expectations for masculine nouns and feminine nouns and whether they are similarly modulated by semantic organization, accessibility of lexical knowledge, and cognitive control.

I expected to see longer reading times in incongruent trials compared to congruent ones, as participants might take longer to process an unexpected form. I also expected to see longer reading times in the masculine incongruent time compared to the feminine incongruent conditions, because the feminine determiner would set up a higher expectation, leading to the masculine noun being especially unexpected. For the individual differences measures, I expected participants who produced more words on the category verbal fluency task (which from here on will be referred to as the semantic word fluency) to have longer reading times on the incongruent feminine trials compared to other participants, because they would have more knowledge of noun genders. On the other hand, participants who had higher error rates and longer response times on the AY condition of the AX-CPT task were expected to show longer reading times on all incongruent trials because they would experience more prediction error from the determiner. I expected this difference to be more pronounced in masculine incongruent trials, because the feminine determiner is a more marked cue.

This hypothesis is formed under the assumption that French speakers process gender in the same way that Spanish speakers do. Although there is a possible difference in cue reliability, I do not expect it to make a significant difference. My rationale for this is that native French speakers have so many tokens of experience mapping gender to noun forms through other agreeing elements that the percentage of tokens where gender is not explicitly marked become irrelevant. Further, once the gender of a word is learned, it should be stored in the lexicon and speakers should not need to rely on marking on agreeing elements to access that gender information. Most native adult speakers of both French and Spanish perform at ceiling when quizzed on the gender of known nouns (Kupisch et al., 2013 for French; Montrul et al., 2008 for Spanish), so there is likely no significant difference in lexical storage.

Overall, my study should complement the findings of Beatty-Martínez et al. (2021) in determining what may be universals in language processing and what elements are highly variable between participants and languages.

## 3. Method

### 3.1 Participants

50 participants were recruited through Prolific, an online recruitment platform based in the United Kingdom. They were paid £6 for an estimated 40 minutes of their time. All participants were screened through Prolific and listed French as a fluent language, primary language, and earliest language learned in life. Due to platform limitations, only 4 participants were monolingual French speakers. 18 participants were bilingual, listing French and English as fluent languages and 26 were multilingual. The participants were split between 16 females and 34 males with ages ranging from 18 to 68 (M = 30.2). 10 submissions had to be discarded due to poor language proficiency or inattention, leaving 40 submissions to be analyzed. All participants gave informed consent, and the procedures were approved by the McMaster Research Ethics Board (project ID 6599).

### 3.2 Stimuli and Tasks

All tasks were programmed using PsychoPy (Peirce et al., 2019).

#### 3.2.1 Semantic Word Fluency Task

Participants were shown category names in French and were asked to type as many exemplars of the category as they could in 30 seconds. Each participant saw four out of a possible eight category words in random order. The categories were animals, furniture, fruit, vegetables, colours, musical instruments, or body parts.

#### 3.2.2 Cognitive Control Task

Participants saw a cue letter in uppercase which was either ‘A’ or ‘B’ and instructed to press the space bar to move on to the next letter, which would be a sequence of three lowercase distractor letters, which could be any letter other than ‘A’, ‘B’, ‘X’, or ‘Y’. Finally, they saw a probe letter which was either ‘X’ or ‘Y’. There were 30 trials, preceded by 5 practice trials where the participant saw the correct answer after answering. 70% of trials were AX, with the remaining 10% split between AY, BX, and BY.

#### 3.2.3 Self-Paced Reading Task

Eighty semantically low constraint sentences were constructed using masculine-feminine noun pairs, following the same structure as used by Beatty-Martínez et al. (2021), as discussed in section 2.3. The gender of the determiners was likewise alternated to create the same four conditions.

Determiners were always singular and were equally balanced between definite and indefinite across the list. Definiteness was always kept the same across noun pairs. The target nouns were controlled for orthographic length (masculine nouns: M = 6.60; feminine nouns: M = 7.24; t(39) = 1.3727, p = .178), phonological length (masculine nouns: M = 5.23; feminine nouns: M = 5.33; t(39) = 0.2510, p = .803), and frequency (masculine nouns: M = 35.17; feminine nouns: M = 30.96; t(39) = 0.3375, p = .738) using data from the CLEARPOND database (Marian et al., 2012). See Table 5 for an example experimental quartet, with the target nouns bolded.

**Table 5: Main Task Stimuli**

| Condition | Sentence | Translation |
|---|---|---|
| Congruent Masculine | Le jardinier a planté un **arbre** dans la cour. | The gardener planted a.MSC tree.MSC in the yard. |
| Incongruent Masculine | Le jardinier a planté une **arbre** dans la cour. | The gardener planted a.FEM tree.MSC in the yard. |
| Congruent Feminine | Le jardinier a planté une **graine** dans la cour. | The gardener planted a.FEM seed.FEM in the yard. |
| Incongruent Feminine | Le jardinier a planté un **graine** dans la cour. | The gardener planted a.MSC seed.FEM in the yard. |

Each participant saw 10 trials per condition, resulting in 40 unique trial quartets which were sorted into four experimental lists in a Latin square design, such that no participant saw more than one version of each quartet. Experimental sentences were pseudo-randomized along with 80 filler sentences, 40 of which were grammatically correct and 40 of which contained other types of agreement violations. A full list of experimental stimuli can be found in A5 of the appendix.

### 3.3 Design and Analysis

#### 3.3.1 Semantic Word Fluency Task

This was an individual difference measure. The dependent variable was the number of unique, semantically appropriate words each participant produced for each category name. The average across all four categories was compared to the main task results.

#### 3.3.2 Cognitive Control Task

This was also an individual difference measure. The independent variables for this task were the

participant number and the sequence type (AX, AY, BX, or BY) and dependent variables were response time and error rate. The condition of interest was specifically the AY condition, as longer response times and increased error rates on this condition can reflect cognitive control. The BY condition can be used as a control, as it has the same probe, but no expectancy set up by the cue.

### 3.3.3 Self-Paced Reading Task

The independent variables for this task were grammaticality (incorrect or correct) and gender (feminine or masculine). The dependent variable was response time, which was measured at every target noun.

## 3.4 Procedure

All tasks were completed online on the participants' personal computers on the experiment hosting site Pavlovia. The median time of completion was 22m10s.

### 3.4.1 Cognitive Control Task

The participants were instructed to find sequences that started with 'A' and ended with 'X'. They pressed the space bar to move through the cue and distractor letters until they reached the probe letter in uppercase. They would press 'o' for OUI for AX trials or 'n' for NON for all other trials. Letters remained on the screen until the participant pressed a button to move onto the next screen. Trials were separated by a fixation cross.

### 3.4.2. Semantic Word Fluency Task

Participants saw 4 out of the 8 possible semantic categories, one at a time. Each category name appeared in the middle of the screen for 30 seconds with a textbox below. Participants were instructed to type as many unique exemplars of each category as they could think of in the text box, pressing the enter key after each word. They were also instructed to avoid repeating words or using proper nouns. Once the time was up for a category, they immediately moved on to the next category.

### 3.4.3 Self-Paced Reading Task

Sentences were presented one word at a time, centered on the screen, with a blank screen between sentences. The subjects were instructed to read the sentences at their own pace and press the space bar when they were ready to move from word to word. 18 sentences at random spots throughout the presentation were followed by a comprehension task which served as an attention check. The

comprehension task was a single word that appeared on the screen and the participant was instructed to determine whether or not the word had appeared in the preceding sentence, pressing “o” for oui if it had and “n” for non if it had not. The words were evenly split between having and having not been present in the preceding sentence. Each of the six types of sentence (the four experimental conditions, correct fillers, and incorrect fillers) had three sentences that were followed by this task. The participants completed six practice rounds before moving on to the main task, in which each sentence was followed by a comprehension question and they were presented the correct answer after responding.

## 4. Results

### 4.1 Statistical considerations

All analyses were run in jamovi (The jamovi project, 2023, an interface for R (R Core Team, 2022). Linear mixed models were made using the jamovi module GAMlj (Gallucci, 2019).

### 4.2 Semantic Word Fluency Results

The average number of exemplars per category was $M = 8$, $SD = 2.59$.

### 4.3 Cognitive Control Results

Table 6 summarizes the mean reaction times and error rates per condition across participants. In total, there were 47 errors out of 1200 trials. They were contained to the same 11 participants, with most participants answering correctly on every trial. There was not a significant difference in error rate between any of the trials. Since the accuracy data was at ceiling, it was not further analyzed.

The AY condition had the slowest reaction times but did not significantly differ from the control condition, BY ($t(39) = 1.67$, $p = 0.1023$). However, there was a significant difference between reaction times between the AY and AX conditions ($t(39) = 5.54$, $p<0.01$).

**Table 6: AXCPT Results**

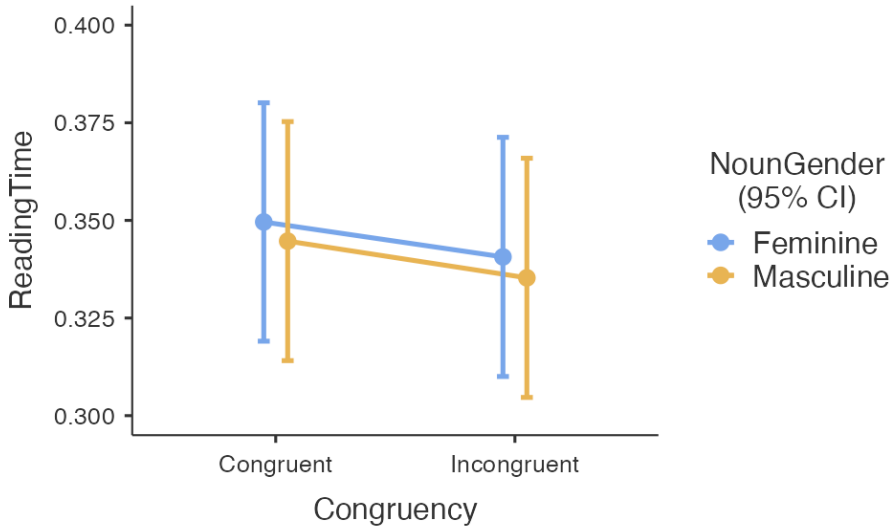
| **Condition** | AX | AY | BX | BY |
|---|---|---|---|---|
| **Reaction Time (ms)** | $M = 776.3$, $SD = 283.49$ | $M = 965.73$, $SD = 352.03$ | $M = 821.25$, $SD = 335.42$ | $M = 862.33$, $SD = 492.36$ |
| **Error Rate** | 0.03 | 0.04 | 0.1 | 0.05 |

### 4.4 Self-paced Reading Results

A repeated measures ANOVA analysis was run to examine the effect of gender and congruency on reading time. Participants with average reading times under 200 ms or over 700 ms were filtered out. Neither main effects nor their interaction was significant. A2 and A3 in the appendix display the full ANOVA results.

Next, a linear mixed model was run to re-examine the task effects and control for both variance between sentences and participants as random effects. Trials that were under 200 ms or over 700 ms were filtered out. Again, no significant results were found. See Figure 2 below for a plot of the results. A4 in the appendix displays the full mixed model results.

**Figure 2: Reading Times**




Mean reading time in seconds for each condition, after filtering, with a 95% confidence interval.

### 4.4 Interactions

For the verbal fluency task, the mean number of exemplars per category was calculated for every participant. For the AXCPT task, the difference in reading time between the AY condition and the AX condition was calculated. These were used as covariates.

The linear mixed model was re-run with word level control variables: frequency, orthographic length, and phonological length. No significant effects were found.

## 5. Discussion

### 5.1. Results

For the main task, participants were expected to have initial difficulty integrating the incongruent forms, resulting in longer reading times in incongruent conditions than congruent ones. This difference was expected to be particularly pronounced in the masculine incongruent condition, as masculine can be considered less marked. Therefore, seeing a feminine determiner can be expected to set up more defined expectations, which would make the transition from a feminine determiner to a masculine noun more disruptive.

Verbal category fluency can range by person. A large range in results was expected for this task. However, it was predicted that participants with a high verbal category fluency score would have a more defined schematicity between the forms of masculine and feminine nouns. As feminine nouns hold a tighter range of variance, it was expected that participants with a more defined schematicity between categories would have longer reading times on the incongruent feminine trial than other participants.

For the AX-CPT task, a large expectancy is built up by seeing predominantly AX sequences, so I expected to see a slower reaction time and higher error rates in the AY condition compared to both the AX and the BY. Like A-cues, feminine determiners are also believed to set up a higher expectancy and cause a greater prediction error when that expectancy is not met. Engaging in cognitive control can aid in mediating this conflict, therefore shorter reaction times in the AY condition compared to the BY and AX condition relative to other participants was expected in participants with greater inhibitory control. This was expected to translate to reading times, seeing longer reading times in incongruent conditions for participants with higher response times in the AY condition. This difference was expected to be more pronounced in the masculine incongruent condition, due to the higher expectancy set up by feminine determiners.

Most surprisingly, there was no significant difference between congruent and incongruent conditions in the main reading task, regardless of noun gender. This could simply be a result of poor filtering of subjects. Most participants included in the data of this study were multilingual, with up to five languages listed as earliest language in life and up to ten languages listed as fluent languages. Eight participants were excluded before data analysis because they either answered in English, nonsensical French, or gibberish to the verbal category prompts. As this study offered a financial reward, it is possible that participants had incentive to misrepresent their level of fluency in French. Non-fluent speakers of French would be expected to have much lower sensitivity to grammatical errors of any kind. In future studies, a fluency test may be necessary to filter out speakers with low proficiency if it is necessary to recruit bilingual or multilingual participants.

Further, the second largest group of participants were bilingual French-English speakers. There is still a question of fluency here, but also of the processing differences between bilingual and monolingual speakers. Many bilinguals have a dominant language and it is theorized that bilinguals utilize higher syntactic structures from their stronger language during processing in their second language, resulting in performance and processing differences in comparison to monolingual speakers (Bernardini & Schlyter, 2004). As English does not have grammatical gender or any features that agree between noun and adjective, it is unclear exactly what the parallel structures would be, but it is likely that dominant English speakers may have a less defined underlying structure for grammatical gender agreement, even if they are perfectly proficient in French. In fact, past research has suggested that English speakers are less sensitive to agreement cues even within their own language compared to speakers of other languages (MacWhinney et al., 1984).

Even among early, balanced bilinguals, bilingual speakers may have longer reaction times to grammaticality judgment tasks, suggesting a possible difference in processing speed (Foursha et al., 2006). For multilingual speakers who speak another gendered language, interference between the gender of a noun between those languages can slow processing as well, with easier processing to be seen for languages in the same gender between both languages and different gender between languages (Lemhöfer et al, 2008). This interference would have been especially hard to account for as participants spoke a variety of other gendered languages, including Portuguese, Spanish, Arabic, Greek, German, and Dutch. In future studies, it would be most informative to recruit only monolingual speakers to get defensible results. If a consistent pattern for monolingual speakers is identified, it can then be compared to patterns from bilingual speakers from different languages to gain an understanding of any differences that may exist and control for them in studies that have a mix of monolingual and bilingual speakers. Ideally, all studies that are examining the underlying representation of grammatical structures would be done on monolingual speakers as a form of control, but this is unfortunately very challenging in many geographic regions and on many experimental recruitment sites.

There is also the possibility that the repair and reanalysis process that occurs when encountering incongruent gender forms in French is relatively minor when compared to other languages, like Spanish. Self-paced reading measures may not have the temporal sensitivity to capture these processes, especially compared to methods like EEG. French speakers may be particularly insensitive to determiner-noun mismatches, as a large number of word forms they encounter in regular reading (plurals and liaison nouns) do not show any morphological marking for gender agreement. This suggests that it may not be a strictly necessary part of processing of noun phrases and the information can essentially be skipped over if it is not overtly available or if it is not reliably matched. It is also possible that the effect is present, but too minor to influence reading times.

Although French and Spanish speakers both perform at ceiling in gender assignment tasks, it is possible that there could be more minor differences in the strength of their knowledge and how strong the association is between noun and gender. This could be studied using a design where speakers of both Spanish and French are asked to assign gender to words while they are performing other tasks with increasing cognitive demand.

Within incongruent conditions, there was also no difference between phrases with feminine nouns and phrases with masculine nouns. As there was no effect of congruency to begin with, it is possible that whichever reason for the lack of effect there carried over to gender. If the reason for the lack of the results in incongruent conditions was a factor of fluency or the absence, interference, or weakness of gender schematicity in the first place, it seems unlikely that there would be a difference between genders. However, if the supposed difference in markedness was sufficiently strong, it could have resulted in significant results in only one condition, likely the masculine noun violation, as there would be a large expectation that was violated in that mismatch.

It is also possible that there is no default gender in French or less likely, that it is feminine. More background research could be conducted on the form of the French language and the behavior of French speakers to develop a stronger understanding of the difference between masculine and feminine. Firstly, there has not yet been a comprehensive analysis of the differences in feminine nouns and masculine nouns. Corpus data could be used to create more extensive frameworks detailing how orthographic, phonological, morphological, and semantic properties pattern between the two genders. This could be used to determine levels of category adherence of specific words within the language and examine whether one gender has a more defined schematicity than the other.

It could also be beneficial to conduct studies on gender assignment of novel words or error rates in children in existing words that display known markers of gender of varying predictive power. This could further define the difference in mental schematicity at multiple levels of representation. Additionally, studies could be done on adult native speakers to determine how gender processing and recognition is affected in long distance dependencies or tasks with increased processing demands. Although similar studies have already been conducted in other gendered languages, mainly Spanish, it is unclear if and how French speakers would differ.

There is also the possibility that determiner and noun pairs are too close together to modulate reading time. It may be beneficial to conduct a study with distance between the noun and agreeing element. It may also be informative to see how gendered elements which agree in gender because of co-reference behave. In French, pronouns take gendered form for inanimate objects. If feminine is marked, it is possible that the gender information is held onto longer after encountering a feminine noun compared to a masculine noun. It

is also possible that gender information is not processed at all in unmarked forms. If this is the case, you would expect to see longer processing times when encountering a masculine pronoun, as a reader may retroactively access gender information for previous nouns.

Another unexpected result was the lack of effect of the verbal fluency results on sentence reading speeds. In native speakers, this task was theorized to be an index of a larger vocabulary and likely more experience with the patterns related to mapping gender on nouns. This task could also be said to measure semantic categorization abilities. People with higher scores on this task were hypothesized to take longer when encountering a feminine noun violation, but this was not the case. This could again be an effect of fluency, bilingualism, or the lack of sufficient resolution to see differences in gender processing. This measure could also be used to distinguish between people with low and high fluency, in which case you would expect fluent speakers to have a more pronounced response to all incongruent conditions, but this was also not true. The nature of this task had one large confound, which was typing speed. Different individuals may have different typing speeds and it is likely more related to experiential factors such as age or education than linguistic factors. Further, this task may not be descriptively powerful at all, as it is arguable what it is measuring. If there had been significant results of this study, it would have been easy to re-interpret what this task reflects to fit with any given theory or any surprising results found post hoc. Future studies may need to develop a task that is less ambiguous in what it measures.

The AX-CPT task had the most predictable results. As expected, the AY condition had the longest response time, which was a significant difference from the AX condition. However, the difference was not significant from the BY condition. This could mean that the difference seen between the AX and AY condition is a joint effect of the frequency of the AX condition, which made up 70% of trials, facilitating processing speed and the unexpectedness of the AY condition inhibiting it. This tracks with the BX and BY condition showing reaction times that were faster than the AX, but slower than the AY, as neither frequency nor unexpectedness should affect reaction times in those conditions, although notably, these were also not significant differences.

The error rates were also somewhat unexpected. The highest error rate in any condition was BX, as opposed to AY. This could be a measure of attention or memory, as participants are pressing yes on the correct probe, but either did not register or did not remember which cue preceded it. If the issue was lack of attention, this could also explain unexpected findings in other tasks. It is also possible that it was a factor of fluency, as certain participants were responding ‘yes’ on every BX trial, which could indicate a failure to understand the written instructions in French. However, these are not conclusions that can be reliably made, as none of the error rate differences reached significance between conditions.

Both the discrepancies in reaction time and error rates from Beatty-Martínez et al. (2021) could be a

shortcoming in experimental design, as participants went through the cue and distractors at their own pace, as opposed to only seeing letters for 300ms in the original experiment. They could also be an effect of the order of stimuli, as conditions were distributed in a pseudorandom style. The AY conditions were 5th, 8th, and 22nd out of 30, not including practice sessions. In the earlier trials, the participants may not have built up a strong expectancy yet. More trials may also have been necessary to obtain statistically significant results.

The AX-CPT task results also failed to modulate the main task results in any significant way. This could be a result of statistical power and ordering of the AX-CPT task in conjunction with the lack of normal results in the main task. As this study used a different platform and methodologies than previous studies using AX-CPT, it may prove beneficial to conduct norming studies that compare results from this specific set up (or an improved version of it) to other measures of cognitive control, such as a flanker task or Stroop task. Only when both the main task and AX-CPT task show consistent and replicable results on their own will it be possible to reliably compare the two.

Overall, it is hard to determine whether the lack of results in this study represent evidence against the hypothesis or are a symptom of drawbacks in screening and experimental design. Thus, it is necessary to gain a greater understanding of French gender distribution, create tasks that accurately show individual differences that may pertain to processing, and control for participant-level confounds in order to conduct a study that has sufficient explanatory power of differential gender processing in French. Most importantly, a study needs to be conducted using only monolingual speakers.

### 5.2 General Discussion

This study was a reminder of the importance of control in scientific studies. Every aspect of experimental design can have a result-altering effect on the data and once it is done, it is near impossible to separate results that are relevant in proving or disproving a hypothesis from those that are a side effect of the tasks, participants, or presentation.

Despite its multiple tasks, the main motivation of this study was simple: to understand how speakers of languages with grammatical gender use those features to process language. To clearly examine this, it is essential to conduct studies where different genders are split between conditions and other variables are not being manipulated within the target sentences. There are many other areas of study that interact with grammatical gender. In order to use grammatical gender to study other areas, it is essential to first have a theory of gender that is as clear as it could possibly be. For this, it is important to conduct simple and replicable studies in multiple languages, ideally using the same methodology, so they can easily and accurately be compared. It is also important that theoretical accounts of gender inform experimental ones and vice versa. It is only when gender effects are consistent and paired with well-informed theory that we can use

them to understand other features of language and processing.

Once a solid foundation is built in understanding the processing of grammatical gender in its simplest form, it allows for future research to use it as a tool to examine other aspects of language. Namely, the connection between conceptual and grammatical gender. It is still unclear how the semantic element of gender ties into the uses that seem purely structural, but comparing processing of conceptual and grammatical gender in languages that have both could be key to understanding how seemingly arbitrary gender assignment came to be how languages process gender, grammatical or otherwise.

Another important area to be explored is the differences in acquisition and processing between monolingual native speakers, bilingual native speakers, and learners of gendered languages. Grammatical gender is a notoriously difficult part of language for new speakers to acquire, especially at a native-like level. Having a deep understanding of monolingual processing is essential in understanding the nuances that are seen in other types of speakers. Connecting this with individual difference measures could hold a key to understanding acquisition and improving teaching and learning methods.

Finally, understanding of the processing of classic feminine-masculine gender systems creates a baseline to compare between other nominal classification systems. Even within what is considered gender based on biological sex or conceptual gender, there are many systems that act very differently from each other. Comparing languages that have either gained or lost genders over the years could provide insights into why languages have grammatical gender at all. There is also very little experimental research on systems of gender that are not based on biological sex or conceptual gender and comparing how speakers process an animacy-based or humanness-based system could broaden that understanding by an even greater scope. It would also be very valuable to compare gendered languages to languages with declensions and classifiers, which have been distinguished on a theoretical basis, but may share similarities (as well as important differences) with grammatical gender on the level of processing.

## 6. Conclusion

We investigated asymmetries in processing grammatical gender agreement in French and the possibility of individual differences modulating these. It is still unclear if French speakers process masculine and feminine grammatical gender separately or which individual differences may modulate processing of unexpected forms. Overall, theoretical, psycholinguistic, and neurolinguistic accounts of gender are not yet clear enough to either make confident assumptions about gender processing or use gender as a manipulation when addressing other research questions. The potential is clear and with more closely controlled, simple studies, it may be possible to build on a groundwork of understanding that could greatly influence many areas of linguistic inquiry. This experimental design may have the capability to contribute to this foundation of understanding and produce psycholinguistic results that are directly comparable to neurolinguistic research. However, the study would need to be repeated in person with a more carefully selected group of participants.

**References**

Alemán Bañón, J. (2010). *The processing of number and gender agreement in Spanish: An ERP investigation* (Doctoral dissertation, University of Kansas).

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

Ayoun, D. (2010). Corpus data: Shedding the light on French grammatical gender… or not. *Eurosla Yearbook*, *10*(1), 119-141.

Ayoun, D. (2018). Grammatical gender assignment in French: Dispelling the native speaker myth. *Journal of French Language Studies*, *28*(1), 113-148.

Barber, H., & Carreiras, M. (2003). Integrating gender and number information in Spanish word pairs: An ERP study. *Cortex*, *39*(3), 465-482.

Barber, H., & Carreiras, M. (2005). Grammatical gender and number agreement in Spanish: An ERP comparison. *Journal of cognitive neuroscience*, *17*(1), 137-153.

Batterink L, & Neville H (2013). Implicit and explicit second language training recruit common neural mechanisms for syntactic processing. *Journal of Cognitive Neuroscience*, 25, 936 951. 10.1162/jocn

Beatty-Martínez, A. L., Bruni, M. R., Bajo, M. T., & Dussias, P. E. (2021). Brain potentials reveal differential processing of masculine and feminine grammatical gender in native Spanish speakers. *Psychophysiology*, *58*(3), e13737.

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.

Bernardini, P., & Schlyter, S. (2004). Growing syntactic structure and code-mixing in the weaker language: The Ivy Hypothesis. *Bilingualism: language and cognition*, *7*(1), 49-69.

Bhattamishra, S., Muralikrishnan, R., & Choudhary, K. K. (2019). Animacy modulates gender

agreement computation in Hindi: An ERP study. *bioRxiv*, 764811.

Bull W (1965). *Spanish for teachers: Applied Linguistics*. Ronald Press.

Caffarra, S., Siyanova-Chanturia, A., Pesciarelli, F., Vespignani, F., & Cacciari, C. (2015).

Clark, Eve. 1985. "The acquisition of Romance, with special reference to French." The crosslinguistic study of language acquisition. Vol. I. Ed. Dan I. Slobin. Hillsdale, N.J.: Lawrence Erlbaum Associates. Pp. 705-707,

Clegg, J. H. (2011). A frequency-based analysis of the norms for Spanish noun gender. *Hispania*, 303-319.

Corbett, Greville G. 2013. *Systems of Gender Assignment*. In: Dryer, Matthew S. & Haspelmath, Martin (eds.) WALS Online (v2020.3) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.7385533 (Available online at http://wals.info/chapter/32, Accessed on 2023-10-20.)

Corbett, Greville G. *Gender*. Cambridge University Press, 1991.

De la Cruz Cabanillas, I., Martínez, C. T., Prados, M. D., & Redondo, E. C. (2007). English loanwords in Spanish computer language. *English for Specific Purposes*, *26*(1), 52-78.

De Resende, N. C. A., Mota, M. B., & Seuren, P. (2019). The processing of grammatical gender agreement in Brazilian Portuguese: erp evidence in favor of a single route. *Journal of psycholinguistic research*, *48*(1), 181-198.

Drieghe, D., Rayner, K., & Pollatsek, A. (2005). Eye movements and word skipping during reading revisited. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(5), 954.

Deutsch, A., & Bentin, S. (2001). Syntactic and semantic factors in processing gender agreement in Hebrew: Evidence from ERPs and eye movements. *Journal of memory and language*, *45*(2), 200-224.

Eddington, D. (2002). Spanish gender assignment in an analogical framework. *Journal of Quantitative Linguistics*, *9*(1), 49-75.

Foursha, C. D., Austin, J. B., & Van de Walle, G. (2006). Is language processing identical in monolinguals and early, balanced bilinguals. In *Proceedings Supplement of the 30th Boston University Conference on Language Development*. Boston, MA: Boston University Linguistics.

Frenck-Mestre, C., Foucart, A., Carrasco-Ortiz, H., & Herschensohn, J. (2009). Processing of grammatical gender in French as a first and second language: Evidence from ERPs. *Eurosla yearbook*, *9*(1), 76-106.

Friederici, A. D. (1995). The time course of syntactic activation during language processing: A model based on neuropsychological and neurophysiological data. *Brain and language*, *50*(3), 259-281.

Frisson, S., & Sandra, D. (2002). Homophonic forms of regularly inflected verbs have their own orthographic representations: A developmental perspective on spelling errors. *Brain and Language*, *81*(1-3), 545-554.

Gallucci, M. (2019). *GAMLj: General analyses for linear models*. [jamovi module]. Retrieved from https://gamlj.github.io/.

Guajardo, L. F., & Wicha, N. Y. (2014). Morphosyntax can modulate the N400 component: Event related potentials to gender-marked post-nominal adjectives. *NeuroImage*, *91*, 262- 272

Gunter, T. C., Friederici, A. D., & Schriefers, H. (2000). Syntactic gender and semantic expectancy: ERPs reveal early autonomy and late interaction. *Journal of cognitive neuroscience*, *12*(4), 556-568

Gunter, T. C., Stowe, L. A., & Mulder, G. (1997). When syntax meets semantics. *Psychophysiology*, *34*(6), 660-676

Hagoort, P. (2003). Interplay between syntax and semantics during sentence comprehension: ERP effects of combining syntactic and semantic violations. *Journal of cognitive neuroscience*, *15*(6), 883-899.

Hagoort, P., & Brown, C. M. (1999). Gender electrified: ERP evidence on the syntactic nature of gender processing. *Journal of psycholinguistic research*, *28*(6), 715-728.

Hull, A. (1968). The origins of New World French phonology. *Word*, *24*(1-3), 255-269.

Idrissi, A., Mustafawi, E., Khwaileh, T., & Muralikrishnan, R. (2021). A neurophysiological study of noun-adjective agreement in Arabic: The impact of animacy and diglossia on the dynamics of language processing. *Journal of Neurolinguistics*, *58*, 100964.

Jakubowicz, C., & Faussart, C. (1998). Gender agreement in the processing of spoken French. *Journal of Psycholinguistic Research*, *27*, 597-617.

Jegerski, J. (2013). Self-paced reading. In *Research methods in second language psycholinguistics* (pp. 20-49). Routledge.

Kim AE, Oines L, & Miyake A (2018). Individual differences in verbal working memory underlie a tradeoff between semantic and structural processing difficulty during language comprehension: An ERP investigation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 44, 406–420. 10.1037/xlm0000394

Kramer, R. (2015). *The morphosyntax of gender* (Vol. 58). Oxford University Press.

Kupisch, Tania, Deniz Akpınar, and Antje Stöhr. "Gender assignment and gender agreement in adult bilinguals and second language learners of French." *Linguistic Approaches to Bilingualism* 3, no. 2 (2013): 150-179.

Kutas, M., & Hillyard, S. A. (1980). Reading between the lines: Event-related brain potentials during natural sentence processing. *Brain and language*, *11*(2), 354-373.

Kutas, M., & Hillyard, S. A. (1983). Event-related brain potentials to grammatical errors and semantic anomalies. *Memory & cognition*, *11*(5), 539-550.

Largy, P., & Fayol, M. (2001). Oral cues improve subject-verb agreement in written French. *International journal of psychology*, *36*(2), 121-131.

Lemhöfer, K., Spalek, K., & Schriefers, H. (2008). Cross-language effects of grammatical gender in bilingual word recognition and production. *Journal of Memory and Language*, *59*(3), 312-330.''

Levitt, J. (1978). The influence of orthography on phonology: a comparative study (English, French, Spanish, Italian, German)

Liversedge, S. P., Paterson, K. B., & Pickering, M. J. (1998). Eye movements and measures of reading time. In *Eye guidance in reading and scene perception* (pp. 55-75). Elsevier Science Ltd.

López Prego, B. (2015). *The online use of markedness information in L1 and L2 Spanish gender agreement* (Doctoral dissertation, University of Kansas).

López Prego, B., & Gabriele, A. (2014). Examining the impact of task demands on morphological variability in native and non-native Spanish. *Linguistic Approaches to Bilingualism*, *4*(2), 192-221.

Lyster, R. (2006). Predictability in French gender attribution: A corpus analysis. *Journal of French Language Studies*, *16*(1), 69-92.

MacWhinney, B., Bates, E., & Kliegl, R. (1984). Cue validity and sentence interpretation in English, German, and Italian. *Journal of verbal learning and verbal behavior*, *23*(2), 127 150.

Marian, V., Bartolotti, J., Chabal, S., & Shook, A. (2012). CLEARPOND: Cross-linguisti easy access resource for phonological and orthographic neighborhood densities.

Martín-Loeches, M., Nigbur, R., Casado, P., Hohlfeld, A., & Sommer, W. (2006). Semantics prevalence over syntax during sentence processing: A brain potential study of noun adjective agreement in Spanish. *Brain Research*, *1093*(1), 178-189.

Mehravari, A. S., Tanner, D., Wampler, E. K., Valentine, G. D., & Osterhout, L. (2015). Effects of grammaticality and morphological complexity on the P600 event-related potential component. *PLoS One*,

Molinaro, N., Barber, H. A., & Carreiras, M. (2011). Grammatical agreement processing in reading: ERP findings and future directions. *cortex*, *47*(8), 908-930.

Molinaro, N., Vespignani, F., & Job, R. (2008). A deeper reanalysis of a superficial feature: An ERP study on agreement violations. *Brain research*, *1228*, 161-176.

Montrul, S., Foote, R., & Perpiñán, S. (2008). Gender agreement in adult second language learners and Spanish heritage speakers: The effects of age and context of acquisition. *Language learning*, *58*(3), 503-553.

Morgan-Short K, Steinhauer K, Sanz C, & Ullman MT (2012). Explicit and implicit second language training differentially affect the achievement of native-like brain activation patterns. *Journal of Cognitive Neuroscience*, 24, 933–947. 10.1162/jocn_a_00119

Nelson, D. (2005). French gender assignment revisited. *Word*, *56*(1), 19-38.

Nevins, A., Dillon, B., Malhotra, S., & Phillips, C. (2007). The role of feature-number and feature-type in processing Hindi verb agreement violations. *Brain research*, *1164*, 81-94.

Osterhout, L., & Mobley, L. A. (1995). Event-related brain potentials elicited by failure to agree. *Journal of Memory and language*, *34*(6), 739-773.

Pakulak E, & Neville HJ (2010). Proficiency differences in syntactic processing of monolingual native speakers indexed by event-related potentials. *Journal of Cognitive Neuroscience*, 12, 2728–2744. 10.1162/jocn.2009.21393

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., Kastman, E., Lindeløv, J. (2019). PsychoPy2: experiments in behavior made easy. *Behavior Research Methods.* 10.3758/s13428-018-01193-y

Pélissier M (2020). Comparing ERPs between native speakers and second language learners: Dealing with individual variability. In Edmonds A, Leclercq P, & Gudmestad A (Eds.), *Interpreting language-learning data* (pp. 39–69). Language Science Press. 10.5281/zenodo.4032282

Pérez-Pereira, M. (1991). The acquisition of gender: What Spanish children tell us. *Journal of*

*child language*, *18*(3), 571-590.

Pinker, S. (1995). Why the child holded the baby rabbits: A case study in language acquisition. *An invitation to cognitive science*, *1*, 107-133.

Popa, L., Selejan, O., Scott, A., Mureşanu, D. F., Balea, M., & Rafila, A. (2015). Reading beyond the glance: eye tracking in neurosciences. *Neurological Sciences*, *36*(5), 683-688.

Popov, S., & Bastiaanse, R. (2018). Processes underpinning gender and number disagreement in Dutch: An ERP study. *Journal of Neurolinguistics*, *46*, 109-121.

Qi Z, Beach SD, Finn AS, Minas J, Goetz C, Chan B, & Gabrieli JDE (2017). Native-language N400 and P600 predict dissociable language-learning abilities in adults. *Neuropsychologia*, 98, 177–191. 10.1016/j.neuropsychologia.2016.10.005

R Core Team (2022). *R: A Language and environment for statistical computing*. (Version 4.1) [Computer software]. Retrieved from https://cran.r-project.org. (R packages retrieved from CRAN snapshot 2023-04-07).

Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: An attentional blink?. *Journal of experimental psychology: Human perception and performance*, *18*(3), 849.

Riley, T. G. (1999). *It's alive! Grammatical animacy in Russian, Polish, and Czech*. University of Washington.

Roché, M. (1992). Le masculin est-il plus productif que le féminin? Langue Française, 96: 113 124

Schmitt, N. (2014). Size and depth of vocabulary knowledge: What the research shows. *Language learning*, *64*(4), 913-951.

Silva-Pereyra, J., Gutierrez-Sigut, E., & Carreiras, M. (2012). An ERP study of coreference in Spanish: semantic and grammatical gender cues. *Psychophysiology*, *49*(10), 1401-1411.

Tanner, D. (2015). On the left anterior negativity (LAN) in electrophysiological studies of morphosyntactic agreement: A Commentary on "Grammatical agreement processing in reading: ERP findings and future directions" by Molinaro et al., 2014. *Cortex*, *66*, 149 155.

Tanner D (2019). Robust neurocognitive individual differences in grammatical agreement processing: A latent variable approach. *Cortex*, 111, 210–237. 10.1016/j.cortex.2018.10.011

The jamovi project (2023). *jamovi*. (Version 2.4) [Computer Software]. Retrieved from https://www.jamovi.org.

Trudgill, P., & Lohndal, T. (2013). Gender maintenance and loss in Totenmålet, English, and other major Germanic varieties. *Search of Universal Grammar*, 77-108.

Tucker, G. R., Genesee, F., & Lambert, W. E. (1977). Communicational effectiveness of English children in French schools. *Canadian Journal of Education/Revue canadienne de l'education*, 15-24.

Van Epps, B., Carling, G., & Sapir, Y. (2021). Gender assignment in six North Scandinavian languages: Patterns of variation and change. *Journal of Germanic Linguistics*, *33*(3), 264-315.

Wicha, N. Y., Moreno, E. M., & Kutas, M. (2004). Anticipating words and their gender: An event-related brain potential study of semantic integration, gender expectancy

Zirnstein M, van Hell JG, & Kroll JF (2018). Cognitive control ability mediates prediction costs in monolinguals and bilinguals. *Cognition*, 176, 87–106. 10.1016/j.cognition.2018.03.001

## Appendix

**A1:** Neural responses across studies

| **Article** | **Language** | **Features manipulated** | **Part of syntactic structure where ERP measured** | **Neural response to agreement error** | **Other elements manipulated** |
|---|---|---|---|---|---|
| Alemán Bañón, J. (2010).<br><br>Experiment 1 (within phrases) | Spanish | Gender and number | Adjective | P600 | Within or. across phrases |
| Alemán Bañón, J., & Rothman, J. (2016). | Spanish | Gender and number | Adjective | LAN<br>P600 | Markededness |
| Barber, H., & Carreiras, M. (2003). | Spanish | Gender and number | Adjective | N400<br>P300* | |
| Barber, H., & Carreiras, M. (2005). | Spanish | Gender and number | Adjective/noun | N400 in pairs<br>LAN in full sentences and article noun pairs<br>P300 in pairs<br>P600 in sentences | Determiner-noun or noun-adjective<br><br>Pairs or full sentences |
| Beatty-Martínez, A. L., Bruni, M. R., Bajo, M. T., & Dussias, P. E. (2021). | Spanish | Gender | Noun | LAN<br>P600 | Markedness |
| Bhattamishra, S., Muralikrishnan, R., & Choudhary, K. K. (2019). | Hindi | Gender | Verb | N400 for animate<br>P600 for inanimate | Animacy |
| Caffarra, S., Siyanova-Chanturia, A., Pesciarelli, F., | Italian | Gender | Noun | LAN<br>P600 | Regularity |
| Vespignani, F., & Cacciari, C. (2015). | | | | | |

| | | | | | |
|---|---|---|---|---|---|
| De Resende, N. C. A., Mota, M. B., & Seuren, P. (2019). | Brazilian Portuguese | Gender | Noun | LAN<br>P600 | Regularity |
| Deutsch, A., & Bentin, S. (2001). | Hebrew | Gender | Predicate of sentential object complement clause | ELAN<br>N400<br>P600 | Markedness<br>Animacy |
| Frenck-Mestre, C., Foucart, A., Carrasco-Ortiz, H., & Herschensohn, J. (2009). | French | Gender | Adjective | P600 | German or English as L1<br>Phonetic variability |
| Guajardo, L. F., & Wicha, N. Y. (2014). | Spanish | Gender | Adjective | N400<br>LPCb | Semantic congruency |
| Gunter, T. C., Friederici, A. D., & Schriefers, H. (2000). | German | Gender | Determiner and noun | LAN<br>P600 only for high-cloze | Cloze probability |
| Hagoort, P. (2003). | Dutch | Gender and number | Noun | P600 | Semantic congruency<br>Sentence position |
| Hagoort, P., & Brown, C. M. (1999). | Dutch | Gender | Noun | P600<br>N400 for sentence-final position only | Sentence position |
| Idrissi, A., Mustafawi, E., Khwaileh, T., & Muralikrishnan, R. (2021). | Arabic | Gender | Verb | N400 only for human<br>P600 only for human | Full agreement (human) or deflected agreement (non-animate human) |
| Kutas, M., & Hillyard, S. A. (1983). | English | Number | Verb | LAN<br>Possible P600 | |
| Martín-Loeches, M., Nigbur, R., Casado, P., | Spanish | Gender and number | Adjective | LAN<br>P600 | Combined violations with semantics |
| Hohlfeld, A., & Sommer, W. (2006). | | | | | |

| | | | | | |
|---|---|---|---|---|---|
| Molinaro, N., Vespignani, F., & Job, R. (2008). | Italian | Gender | Noun | LAN<br>P600 | Phontactic violations |
| Nevins, A., Dillon, B., Malhotra, S., & Phillips, C. (2007). | Hindi | Number, gender, and person | Verb | P600 | |
| Popov, S., & Bastiaanse, R. (2018). | Dutch | Gender and number | Noun | P600 | |
| Silva-Pereyra, J., Gutierrez-Sigut, E., & Carreiras, M. (2012). | Spanish | Gender | Pronoun | N400<br>P600 | Ambiguity<br>Animacy |
| Wicha, N. Y., Moreno, E. M., & Kutas, M. (2004). | Spanish | Gender | Determiner or noun | P600 | Semantic expectancy<br>Semantic congruency |

*This study likely got an earlier positivity because they used word pairs instead of full sentences.

### A2: Repeated Measures ANOVA Within Subjects Effects

| | **Sum of Squares** | **df** | **Mean Square** | **F** | **p** | **$\eta^2_p$** |
|---|---|---|---|---|---|---|
| Gender | 19188 | 1 | 19188 | 0.6618 | 0.421 | 0.018 |
| Gender * AY-AX | 25683 | 1 | 25683 | 0.8859 | 0.353 | 0.023 |
| Gender * VFAvg | 11492 | 1 | 11492 | 0.3964 | 0.533 | 0.011 |
| Residual | 1.07e+6 | 37 | 28992 | | | |
| Congruency | 5382 | 1 | 5382 | 0.2328 | 0.632 | 0.006 |
| Congruency * AY-AX | 15195 | 1 | 15195 | 0.6572 | 0.423 | 0.017 |
| Congruency * VFAvg | 21817 | 1 | 21817 | 0.9436 | 0.338 | 0.025 |
| Residual | 855455 | 37 | 23120 | | | |
| Gender * Congruency | 264 | 1 | 264 | 0.0112 | 0.916 | 0.000 |
| Gender * Congruency | 604 | 1 | 604 | 0.0256 | 0.874 | 0.001 |

| | | | | | | |
|---|---|---|---|---|---|---|
| * AY-AX | | | | | | |
| Gender * Congruency * VFAvg | 2288 | 1 | 2288 | 0.0970 | 0.757 | 0.003 |
| Residual | 872578 | 37 | 23583 | | | |

**A3: Repeated Measures ANOVA Between Subjects Effects**

| | **Sum of Squares** | **df** | **Mean Square** | **F** | **p** | **$\eta^2_p$** |
|---|---|---|---|---|---|---|
| AY-AX | 7201 | 1 | 7201 | 0.0378 | 0.847 | 0.001 |
| VFAvg | 12547 | 1 | 12547 | 0.0658 | 0.799 | 0.002 |
| Residual | 7.05e+6 | 37 | 190621 | | | |

**A4: Mixed Model Fixed Effects Parameter Estimates**

| **Effect** | **Estimate** | **Standard Error** | **df** | **t** | **P** |
|---|---|---|---|---|---|
| (Intercept) | 0.34256 | 0.01414 | 40.1 | 24.2202 | < .001 |
| Incongruent - Congruent | -0.00917 | 0.00652 | 138.6 | -1.4055 | 0.162 |
| Masculine - Feminine | -0.00513 | 0.00677 | 137.4 | -0.7575 | 0.450 |
| Orthographic Length | -4.83e−4 | 0.00309 | 137.2 | -0.1564 | 0.876 |
| Phonological Length | 0.00264 | 0.00349 | 135.2 | 0.7563 | 0.451 |
| Frequency | 2.30e-5 | 5.89e-5 | 143.3 | 0.3904 | 0.697 |
| Incorrect – Correct * Masculine - Feminine | -4.61e−4 | 0.01304 | 137.8 | -0.0354 | 0.972 |

95% Confidence Interval

**A5: Full List of Experimental Stimuli**

1. Le couple a rénové la/*le terrasse de leur nouvelle maison.
Le couple a rénové le/*la balcon de leur nouvelle maison.

Translation: The couple renovated the terrace/balcony in their new house.

2. Le musicien a accordé la/*le clarinette avant le concert.
Le musicien a accordé le/*la trombone avant le concert.
Translation: The musician tuned the clarinet/trombone before the concert.

3. Le pirate a enterré la/*le bouteille sur la plage.
Le pirate a enterré le/*la trésor sur la plage.
Translation: The pirate buried the bottle/treasure on the beach.

4. Le cuisinier a utilisé une/*un pomme pour completer la recette.
Le cuisinier a utilisé un/*une œuf pour completer la recette.
Translation: The cook used an apple/egg to complete the recipe.

5. La fille a commencé une/*un diète pour surveiller sa santé.
La fille a commencé un/*une journal pour surveiller sa santé.
Translation: The girl started a diet/journal to monitor her health.

6. La dame a retourné une/*un écharpe au magasin.
La dame a retourné un/*une manteau au magasin.
Translation: The girl returned a coat/scarf to the store.

7. La famille a remplacé la/*le lampe dans la chambre.
La famille a remplacé le/*la divan dans la chambre.
Translation: The family replaced the lamp/couch in the room.

8. Le jardinier a planté une /*un graine dans la cour.
Le jardinier a planté un/*une arbre dans la cour.
Translation: The gardener planted a seed/tree in the court.

9. Le technicien a branché la/*le caméra à l'ordinateur.
Le technicien a branché le/*la moniteur à l'ordinateur.
Translation: The technician connected the camera to the computer/monitor.

10. Le facteur a annulé la/*le commande qui a été envoyé par la poste.
Le facteur a annulé le/*la colis qui a été envoyé par la poste.
Translation: The postman cancelled the order/package that was sent by post.

11. Le garçon a bu la/*le limonade pour se rafraîchir.
Le garçon a bu le/*la jus pour se rafraîchir.
Translation: The boy drank the lemonade/juice to cool down.

12. Le moine a visité la/*le chapelle pour prier.
Le moine a visité le/*la monastère pour prier.
Translation: The monk visited the chapel/monastery to pray.

13. Le maire a inauguré la/*le statue pour la communauté.
Le maire a inauguré le/*la parc pour la communauté.
Translation: The mayor inaugurated the statue/park for the community.

14. Le volontaire a trouvé une/*un église pour l'itinérante.
Le volontaire a trouvé un/*une refuge pour l'itinérante.
Translation: The volunteer found a church/shelter for the unhoused person.

15. Le bébé a apporté la/*le poupée dans la cuisine.
Le bébé a apporté le/*la jouet dans la cuisine.
Translation: The baby brought the doll/toy into the kitchen.

16, Le grand-père a raconté la/*le légende du dragon.
Le grand-père a raconté le/*la mythe du dragon.
Translation: The grandfather told the legend/myth of the dragon.

17. Le témoin a indiqué une/*un urgence à l'intersection.
Le témoin a indiqué un/*une accident à l'intersection.
Translation: The witness reported an emergency/accident at the intersection.

18. Le bénévole a organisé une/*un activité pour l'école.
Le bénévole a organisé un/*une événement pour l'école.
Translation: The volunteer organized an activity/event for the school.

19. Le charpentier a pris la/*le scie de la boîte.
Le charpentier a pris le/*la marteau de la boîte.
Translation: The carpenter took the saw/hammer from the box.

20. Le monsieur a perdu une/*un pièce dans le parc.
Le monsieur a perdu un/*une mouchoir dans le parc.
Translation: The gentleman lost a coin/handkerchief in the park.

21. La mère a préparé une/*un collation pour ses enfants.

La mère a préparé un/*une dessert pour ses enfants.
Translation: The mom prepared a snack/dessert for her kids.

22. La femme a attendu une/*un semaine pour l'appel.
La femme a attendu un/*une mois pour l'appel.
Translation: The woman waited a week/month for the call.

23. Le réalisateur a enregistré une/*un annonce à l'intersection.
Le réalisateur a enregistré un/*une film à l'intersection.
Translation: The director recorded an announcement/film at the intersection.

24. Le touriste a recommandé la boutique sur la plage.
Le touriste a recommandé le restaurant sur la plage.
Translation: The tourist recommended the boutique/restaurant on the beach.

25. Le styliste a appliqué la/*le lotion sur les cheveux.
Le styliste a appliqué le/*la revitalisant sur les cheveux.
Translation: The stylist applied the lotion/conditioner to the hair.

26. Le vendeur a laissé la/*le mallette à l'aéroport.
Le vendeur a laissé le/*la bagage à l'aéroport.
Translation: The salesman left the briefcase/baggage at the airport.
27. Le participant a accepté le/*la question lors du concours.
Le participant a accepté la/*le défi lors du concours.
Translation: The participant accepted the question/challenge during the competition.
28. La bonne a replacé la/*le vadrouille dans la garde-robe.
La bonne a replacé le/*la balai dans la garde-robe.
Translation: The maid put the mop/broom back in the wardrobe.
La bonne a replacé le vadrouille dans la garde-robe.
La bonne a replacé la balai dans la garde-robe.

29. Le président a confronté le/*la corruption dans son discours.
Le président a confronté la/*le terrorisme dans son discours.
Translation: The president addressed corruption/terrorism in his speech.
30. Le propriétaire a vendu la/*le copropriété à la ville.
Le propriétaire a vendu le/*la commerce à la ville.
Translation: The owner sold the condo/business to the city.

31. Le chevalier a choisi une/*un épée pour la bataille.
Le chevalier a choisi un/*une bouclier pour la bataille.
Translation: The knight chose a sword/shield for battle.
32. Le professeur a organisé une/*un réunion pour les élèves.
Le professeur a organisé un/*une atelier pour les élèves.
Translation: The professor organized a reunion/workshop for the students.

33. Le médecin a examiné la/*le cheville du gymnaste.
Le médecin a examiné le/*la pied du gymnaste.
Translation: The doctor examined the gymnast's ankle/foot.
34. La secrétaire a distribué une/*un brochure aux clients.
La secrétaire a distribué un/*une échantillon aux clients.
Translation: The secretary distributed a brochure/sample to clients.
35. Le nettoyeur a acheté un/*une détergent au magasin.
Le nettoyeur a acheté une/*un eponge au magasin.
Translation: The cleaner bought a detergent/sponge at the store.

36. Le garde a entendu une/*un alarme à la banque.
Le garde a entendu un/*une bruit à la banque.
Translation: The guard heard an alarm/sound at the bank.
37. Le scientifique a proposé une/*un analyse des données.
Le scientifique a proposé un/*une examen des données.
Translation: The scientist proposed an analysis/review of the data.

38. La ballerine a placé une/*un rosette sur son léotard.
La ballerine a placé un/*une ruban sur son léotard.
Translation: The ballerina placed a rosette/ribbon to her leotard.

39. Le journaliste a prévu une/*un conversation avec le president.
Le journaliste a prévu un/*une entretien avec le president.
Translation: The journalist planned a conversation/interview with the president.

40. Le contremaître a prêté une/*un échelle à l'ouvrier.
Le contremaître a prêté un/*une outil à l'ouvrier.
Translation: The foreman lent a ladder/tool to the worker.