Gender attraction in sentence comprehension

Article

Published Version

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.
Published version at: https://www.glossa-journal.org/articles/10.5334/gjgl.1300/
To link to this article DOI: http://dx.doi.org/10.5334/gjgl.1300

Publisher: Ubiquity Press

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

Gender attraction in sentence comprehension

JORGE GONZÁLEZ ALONSO (iD)

IAN CUNNINGS (iD)

HIROKI FUJITA (iD)

DAVID MILLER (iD)

JASON ROTHMAN (iD)

Author affiliations can be found in the back matter of this article

ABSTRACT

Agreement attraction, where ungrammatical sentences are perceived as grammatical (e.g., *The key to the cabinets were rusty*), has been influential in motivating models of memory access during language comprehension. It is contested, however, whether such effects arise due to a faulty representation of relevant morphosyntactic features, or as a result of memory retrieval. Existing studies of agreement attraction in comprehension have largely been limited to subject-verb number agreement, primarily in English, and while attraction in other agreement phenomena such as gender has been investigated in production, very few studies have focused on gender attraction in comprehension. We conducted five experiments investigating noun-adjective gender agreement during comprehension in Spanish. Our results indicate attraction effects during online sentence processing that are consistent with approaches ascribing attraction to interference during memory retrieval, rather than to a faulty representation of agreement features. We interpret our findings as consistent with the predictions of cue-based parsing.

CORRESPONDING AUTHOR:
Jorge González Alonso

UiT The Arctic University of Norway, NO

jorge.gonzalez.alonso@uit.no

1 INTRODUCTION

When engaged in any form of language comprehension, individuals must integrate, in real time, different types of semantic, syntactic, phonological and pragmatic information, as well as experiential knowledge and sensory stimuli. Syntactic dependencies, which make up a large part of the grammatical information in an utterance and are crucial to comprehension, involve two or more elements (words or phrases) that should be processed or interpreted together (e.g., a subject and a verb, a noun and an adjective, a pronoun and its antecedent). Morphological cues play an important role in signalling syntactic dependencies. Across languages, morphosyntactic agreement morphology may help a comprehender to keep track of the constituents that must enter into a syntactic dependency. In English, for example, subjects and verbs must agree in number (e.g., *The boy was tired* vs. **The boy were tired*).

In non-adjacent dependencies, intervening linguistic material increases the distance between the relevant elements. The processing of non-adjacent dependencies becomes particularly challenging when this intervening linguistic material contains elements that have similar characteristics to the elements involved in the actual dependency (Jäger, Engelmann & Vasishth 2017). A classic example of this is subject-verb agreement attraction (e.g., **The **money** for the **projects have** been well administered by the board*), where the verb's number morphology agrees with an intervening noun phrase (*the projects*) instead of the subject's head noun (*the money*). First described and experimentally elicited in production (e.g. Bock & Miller 1991; Bock & Cutting 1992), agreement attraction effects have also been extensively studied in language comprehension (e.g. Pearlmutter, Garnsey & Bock 1999; Wagers, Lau & Phillips 2009; Dillon et al. 2013; Tanner, Nicol & Brehm 2014; Lago, Shalom, Sigman, Lau, & Phillips 2015; Schlueter, Williams & Lau 2018; Lago, Gračanin-Yuksek, Şafak, Demir, Kırkıcı, & Felser 2019). The increased interest in attraction effects in comprehension has in part been motivated by the role that agreement attraction has played as primary evidence within a wider literature on the memory architecture that underlies language comprehension (e.g. Lewis & Vasishth 2005; Wagers et al. 2009; Dillon et al. 2013; Lago et al. 2015; Jäger et al. 2017; Parker, Shvartsman & Van Dyke 2017; Villata, Tabor & Franck 2018; Hammerly, Staub & Dillon 2019). Debate here has focused on whether subject-verb agreement attraction results from a faulty representation of the relevant morphosyntactic features on the sentence subject, or from an error-prone memory retrieval mechanism.

Most of the existing literature on attraction in comprehension has focused on subject-verb number agreement (Pearlmutter et al. 1999; Kaan 2002; Wagers et al. 2009; Tanner et al. 2014; Lago et al. 2015). While other types of morphosyntactic agreement, such as gender agreement, have received some attention in production (Vigliocco & Franck 1999; Vigliocco & Zilli 1999; Badecker & Kuminiak 2007; Slioussar & Malko 2016), surprisingly little research has examined gender agreement attraction in comprehension. Indeed, whereas subject-verb agreement has been widely studied in comprehension, other types of morphosyntactic agreement, such as noun-adjective agreement, have been less well studied. For example, we are aware of only one other existing study on noun-adjective gender agreement (Acuña-Fariña, Meseguer & Carreiras 2014).

To fill in these empirical gaps in the literature and to test the generalisability of different models of agreement attraction in comprehension, we report five experiments that examined noun-adjective gender agreement attraction in Spanish. By combining different offline and online experimental paradigms, our study sheds new light on the time-course of memory access during sentence processing. We begin by discussing existing research on different models of agreement attraction in comprehension, before discussing previous work on gender attraction.

1.1 NUMBER AGREEMENT ATTRACTION IN COMPREHENSION

Agreement attraction errors in spoken and written language production have featured in descriptive (Quirk, Greenbaum, Leech, & Svartvik 1972) and theoretical linguistic work (Kimball & Aissen 1971; Francis 1986) for decades. This is illustrated in the classic example in (1). Out of the two noun phrases (NPs) highlighted in bold, only the singular *the key* is syntactically licensed as the head of the clause's subject, and therefore, as a valid controller of agreement on the verb. However, instead of taking the singular form *was*, the verb agrees in number with the immediately preceding NP—typically called an *attractor* or *distractor* in the literature—: the plural *the cabinets*.

(1)   Bock & Miller (1991)
         ***The key** to **the cabinets were** rusty.

In comprehension, agreement attraction is defined as differences in reading or listening times, as well as sentence acceptability, observed between versions of a sentence in which the head and attractor nouns display the same or different agreement marking which, in turn, matches or mismatches with marking on the agreement target (e.g., a verb). As discussed in more detail below, two broad accounts of attraction have been proposed. According to representation-based accounts (e.g. Pearlmutter et al. 1999; Hammerly et al. 2019), attraction occurs because of a faulty or ambiguous representation of the number properties of the sentence subject (*the key to the cabinets*) when the two nouns (*the key* and *the cabinets*) have different number specifications. Alternatively, according to retrieval-based accounts (e.g. Wagers et al. 2009; Jäger et al. 2017), attraction is dependent on the match between the number properties of the verb (*were*) and previous nouns in the sentence. A number of studies have attempted to tease these two accounts apart (e.g. Wagers et al. 2009; Hammerly et al. 2019).

Wagers et al. (2009) examined sentences such as those in (2), which manipulate number agreement between the auxiliary verb (*was/were*), the sentence subject (*the key*), and an intervening attractor noun (*the cell/s*):

(2)   Wagers et al. (2009)
         a.   The key to the cell unsurprisingly was rusty from many years of disuse.
         b.   The key to the cells unsurprisingly was rusty from many years of disuse.
         c.   *The key to the cells unsurprisingly were rusty from many years of disuse.
         d.   *The key to the cell unsurprisingly were rusty from many years of disuse.

Wagers et al. reported longer reading times for ungrammatical sentences (2c/d) than grammatical conditions (2a/b). Additionally, attraction effects were observed in ungrammatical sentences, such that reading times were reliably shorter in (2c), when the attractor matched the number properties of the verb, in comparison to (2d), when it did not. Following Jäger et al. (2017), we will refer to this as *facilitatory interference*, as reading times are shorter when the attractor matches the number properties of the verb. Wagers et al. (2009) did not observe any differences between the two grammatical conditions, even though there is attractor-verb matching in 2a as compared to 2b. This difference in attraction effects between ungrammatical and grammatical conditions was dubbed by Wagers et al. the *grammatical asymmetry* in agreement attraction.

The grammatical asymmetry has been observed in a number of studies (e.g. Dillon et al. 2013; Shen, Staub & Sanders 2013; Tanner et al. 2014; Lago et al. 2015; 2019; Tucker, Idrissi & Almeida 2015; Schlueter et al. 2018; but see Pearlmutter et al. 1999; Häussler 2009; Enochson & Culbertson 2015). This pattern of results has been argued to support a retrieval-based account of attraction couched within the framework of cue-based memory retrieval during sentence processing (e.g., Wagers et al. 2009). According to such models, language comprehension is subserved by a direct-access retrieval mechanism interacting with a content-addressable memory (e.g. McElree 2000; McElree, Foraker & Dyer 2003). In this framework, sentence constituents are encoded in memory as a sentence is processed. Every constituent that marks the tail-end of a long-distance dependency initiates retrieval of a potential antecedent that may function as the controller of that dependency. Retrieval is conducted through a query that specifies a set of criteria, or cues, that the antecedent must satisfy. The item in memory that provides the best match is subsequently retrieved. For subject-verb agreement as in (2), retrieval cues at the verb will include specification for an antecedent with particular structural and morphological features. For (2c/d), retrieval cues at the verb may include [+plural] and [+nominative]. Although no item in memory fully matches these cues in (2c/d), in (2c) the attractor (*cells*) returns a partial match with [+plural] (while *key* matches the case cue). Cue-based parsing predicts that this partial match will result in the attractor being retrieved as the agreement controller some proportion of the time, leading to facilitatory interference reflected in shorter reading times.

In their explanation of the lack of attraction in grammatical sentences, Wagers et al. (2009) argued that the retrieval mechanism is robust against interference from partially matching

attractors when the target fully matches the cues to retrieval. Since, in grammatical conditions, the head noun matches both the structural cue and the verb's morphological features, Wagers et al. reasoned that the attractor would be unable to produce significant interference, since it can only ever match the morphological cue. On the other hand, in ungrammatical sentences the head noun *only* matches the structural cue. Therefore, when the attractor matches the morphological specification of the verb, both the head noun and the attractor are tied in the number of retrieval cues they satisfy, which makes interference more likely. Although Wagers et al. (2009) did not observe attraction effects in grammatical sentences, most implementations of cue-based parsing (e.g. Lewis & Vasishth 2005; Lewis, Vasishth & Van Dyke 2006; Van Dyke & McElree 2011; Jäger et al. 2017) do predict differences between grammatical sentences, but in the *opposite* direction to effects found in ungrammatical sentences. For example, the Lewis and Vasishth (2005) implementation of this approach predicts competition during retrieval when multiple items in memory partially match the retrieval cues in grammatical sentences. This would predict longer reading times in (2a), since the head noun and attractor become activated by their match to the number properties of the verb, than (2b), when only the head noun matches this cue. Jäger et al. (2017) refer to this as *inhibitory interference*, as reading times for grammatical sentences are predicted to be longer when an attractor noun partially matches the cues at retrieval. However, while studies on other linguistic dependencies have reported inhibitory interference (e.g. Van Dyke 2007; Van Dyke & McElree 2011; for review see Jäger et al. 2017), some studies on subject-verb agreement attraction have failed to observe this effect (e.g. Wagers et al. 2009; Shen et al. 2013; Tanner et al. 2014; Lago et al. 2015). Nicenboim, Vasishth, Engelmann and Suckow (2018) argued that inhibitory effects in subject-verb number agreement may be numerically small and difficult to detect, and that the apparent lack of inhibitory interference in some studies may be due to lack of sufficient statistical power. Indeed, they estimated inhibitory effects in subject-verb agreement to be of the magnitude of 9 ms in a sample of 184 participants across two experiments.

In sum, although inhibitory effects in grammatical sentences are difficult to observe without large samples or simply do not obtain, the key prediction of all cue-based parsing accounts is an asymmetry in attraction effects, with facilitatory interference predicted for ungrammatical sentences only. Since we do not attempt to tease apart different cue-based models here, we will refer to these accounts collectively as *retrieval-based accounts* of agreement attraction in comprehension.

In contrast to retrieval-based accounts of agreement attraction, other theories (e.g. Franck, Vigliocco & Nicol 2002; Franck, Vigliocco & Nicol 2008; Eberhard, Cutting & Bock 2005) place the locus of agreement attraction effects at the encoding/representation of agreement features of the grammatically licit controller of the dependency — in subject-verb number agreement, for example, this would be the subject of the main clause. Due to this understanding of attraction effects in terms of a faulty or ambiguous encoding of feature values rather than a retrieval error, such accounts have come to be known as *representational models* of agreement attraction (e.g., Wagers et al. 2009).

All representational theories were originally developed to account for attraction effects in production, yet several studies have sought to extend their predictions to comprehension (e.g., Pearlmutter et al. 1999; Hammerly et al. 2019). According to the feature-percolation account (e.g. Nicol, Forster & Veres 1997; Franck et al. 2002), agreement computation is achieved by a process that builds up syntactic hierarchies through a successive merging of smaller segments. Features of the smaller segments are passed (or percolate) upwards when these are unified into larger units. Occasionally, the number specification of the attractor noun may percolate upwards to the subject node, constituting the value against which number marking on the verb will be checked. This would predict that in sentences with a complex subject comprising a singular head noun followed by a second, plural NP (e.g., 2b/c), the number property of the subject NP (*the key to the cells*), which should follow that of its singular head noun (*the key*), will be mis-specified as [+plural] on some proportion of trials (Franck et al. 2002).

In contrast, the Marking and Morphing model (M&M; Eberhard et al. 2005; Bock & Middleton 2011) proposes that the number specification of the subject NP is not categorically singular or plural, but rather takes on a gradient value somewhere in between those two extremes. Bock and colleagues called this the *singular and plural value* (SAP). The two main contributors to the

SAP of an NP are, on the one hand, semantic or notional number (i.e., the notion of plurality or numerosity at a conceptual level), and on the other the number specification of morphemes in the phrase (i.e., number marking; Eberhard et al. 2005), including not only the head noun but also any other nouns present in complement or adjunct phrases within the NP. Because both of these contributing factors can be, at a minimum, singular or plural in various combinations and at various levels (e.g., notionally plural nouns that are morphologically singular), the SAP value of different NPs will present continuous variation along the singular-plural scale. In a complex subject headed by a singular noun, the presence of an embedded plural noun contributes to a less extreme overall SAP value on the subject node, leading to higher ambiguity between a singular and a plural valuation of the subject, as compared to a complex subject with two singular nouns.

While there are notable differences of implementation between the feature-percolation and M&M models, in both cases attraction is inherent to the agreement process, and a function of the tension between the different contributors to the number valuation of the agreement controller (i.e., the subject node). This predicts that success in computing the dependency is primarily dependent on how information is encoded or represented in memory. Importantly, these representation-based accounts predict a different pattern of attraction effects in comprehension to that of retrieval-based accounts, such that attractors whose feature (e.g., number) specification matches with that of the head noun should facilitate reading times in both ungrammatical *and* grammatical sentences. In other words, these models do not predict a grammatical asymmetry.

While some studies have reported results consistent with symmetrical attraction effects in comprehension (e.g., Pearlmutter et al. 1999; Häussler 2009), the grammatical asymmetry has been a consistent finding in several other studies (e.g., Wagers et al. 2009; Shen et al. 2013; Tanner et al. 2014; Lago et al. 2015). Wagers et al. in particular noted that previous studies supporting representation-based accounts may contain an important confound in that plural nouns are 'heavier' to process and take longer to read than singular nouns. This effect can also spill over to subsequent words. As such, when the verb and the attractor are adjacent (e.g., *The key to the cabinet(s) was rusty*), it is difficult to tease apart whether any differences in reading time at the verb are the result of attraction, even in grammatical sentences, or merely a reflection of the fact that plural nouns incur a reading time penalty in comparison to their singular counterparts.

In sum, number agreement attraction in comprehension has been accounted for both in terms of a faulty or ambiguous representation of number in the complex subject, or as a result of memory interference in the retrieval of an appropriate agreement controller for the number properties of the verb. The crucial contrast between these two groups of models is that representational theories predict symmetrical effects across grammatical and ungrammatical sentences, while retrieval-based accounts predict asymmetric effects, with facilitatory interference being restricted to ungrammatical sentences.

1.2 GENDER AGREEMENT ATTRACTION

While subject-verb number agreement attraction has been widely examined in both production and comprehension, fewer studies have investigated gender agreement attraction, especially in comprehension (although see Martin, Nieuwland & Carreiras 2012; 2014; Slioussar & Malko 2016; Paspali & Marinis 2020; Villata & Franck 2020). Many studies have manipulated gender *congruency* between an anaphor/pronoun and its antecedent to investigate memory retrieval during anaphora resolution (for review, see Jäger et al. 2017). However, we restrict our discussion here to studies on morphosyntactic gender *agreement* between sentence constituents, as this type of purely morphosyntactic agreement relationship is the focus of the current study.

While English has little morphosyntactic agreement beyond subject-verb agreement, other languages with richer morphology allow for testing of attraction effects in a wider range of morphosyntactic phenomena. These constitute important testing cases of the generalisability of both retrieval-based and representation-based accounts of attraction. In Romance languages like Spanish, elements in a sentence, such as determiners and adjectives, must

agree in grammatical gender with the nouns that they modify, as in (3). Animate nouns referring to human individuals, as well as some non-human animals, are assigned masculine or feminine gender in line with their biological sex. As such, we refer to this as *biological gender*. This property is then reflected in agreeing elements within the sentence, as in (3a/b). Inanimate nouns, where there is no sex to reflect, are also assigned an invariant gender, which is again reflected in agreeing words (determiners, adjectives), as in (3c/d). We will refer to this as *lexical gender*.

(3)   *Spanish*

　　a. **La**   niñ-**a**   estaba castigad-**a**.
　　 the(F) child-F was   grounded-F
　　 'The girl was grounded.'

　　b. **El**   niñ-**o**   estaba castigad-**o**.
　　 the(M) child-M was   grounded-M
　　 'The boy was grounded.'

　　c. **La**   **manzana** estaba sabros-**a**.
　　 the(F) apple(F)   was   tasty-F
　　 'The apple was tasty.'

　　d. **El**   **plátano** estaba sabros-**o**.
　　 the(M) banana(M) was   tasty-M
　　 'The banana was tasty.'

A number of studies have examined gender attraction in production (Vigliocco & Franck 1999; Vigliocco & Zilli 1999; Anton-Mendez, Nicol & Garrett 2002; Badecker & Kuminiak 2007; Franck et al. 2008). For example, Vigliocco and Franck (1999) reported four experiments on noun-adjective agreement in Italian and French that showed gender attraction effects for both biological and lexical gender nouns (see also Vigliocco, Butterworth & Semenza 1995). Further research on noun-adjective agreement between a subject and a predicative adjective in Spanish, Italian and French (Franck et al. 2008) replicated these general findings, which have subsequently been shown to extend beyond this particular domain, being reported in research on languages with three-gender systems (Slovak, Russian) examining subject-verb gender agreement (Badecker & Kuminiak 2007; Slioussar & Malko 2016, Exp. 1).

With regards to comprehension, Slioussar and Malko (2016; Experiments 2ab and 3) conducted a series of self-paced reading experiments on subject-verb agreement in Russian, where participants read grammatical and ungrammatical sentences in which the gender of a past-tense verb matched or mismatched the gender of the head noun in a complex subject containing an attractor noun. Gender (mis)match between the attractor and the head noun/verb was systematically manipulated. Their results showed no indication of attraction effects in grammatical sentences, for either combination of head and attractor noun. Instead, the results displayed an asymmetrical pattern of agreement attraction, with facilitatory interference in ungrammatical sentences only, consistent with cue-based parsing (see also Villata & Franck 2020, for self-paced reading evidence of gender attraction in French object-verb agreement).

Paspali and Marinis (2020) examined gender agreement attraction in Greek in two different grammatical contexts: adjectival predicates and object clitics. Using data from self-paced listening (Exp. 1 and 2) and both timed (Exp. 3) and untimed acceptability judgements (Exp. 4), they found evidence of attraction in both contexts and across both real-time and sentence-final measures (speeded judgements), although the untimed acceptability judgements of Experiment 4 did not show a significant grammaticality by attractor-match interaction (see also Fuchs, Polinsky & Scontras 2015 for similar disagreement between the results of timed and untimed judgements).

Martin et al. (2012) conducted an event-related potential (ERP) study investigating gender agreement in Spanish noun ellipsis, as in (4). Here, the determiner *otra/otro* 'another' must agree in gender with the feminine noun *la camiseta* 'the t-shirt'. In their stimuli, the determiner's gender matches the head noun in grammatical (4a) but mismatches in ungrammatical (4b) sentences. Crucially, gender (mis)match with an intervening attractor noun was simultaneously manipulated.

(4)   *Spanish* (Martin et al. 2012)
   a.   Marta   se   compró **la   camiseta** que estaba al   lado de **la**
        Marta   herself   bought the(F) t-shirt(F) that was   at.the side of the(F)
        falda/ del   vestido y   Miren cogió **otr-a**   para salir de fiesta.
        skirt(F)/ of.the(M) dress(M) and Miren took another-F to   go of party
        'Marta bought herself the t-shirt that was next to the skirt/the dress and Miren
        took another to go to the party.'

   b.   Marta   se   compró **la   camiseta** que estaba al   lado de **la**
        Marta   herself   bought the(F) t-shirt(F) that was   at.the side of the(F)
        falda/ del   vestido y   Miren cogió **otr-o**   para salir de fiesta.
        skirt(F)/ of.the(M) dress(M) and Miren took another-M to   go   of party
        'Marta bought herself the t-shirt that was next to the skirt/the dress and Miren
        took another to go to the party.'

In contrast to studies on subject-verb number agreement in English, which reported a P600 response to agreement violations (Osterhout & Mobley 1995; Xiang, Dillon & Phillips 2009; Shen et al. 2013; Tanner et al. 2014), Martin et al. (2012) reported a sustained negativity following the critical determiner in ungrammatical compared to grammatical sentences. There was also a significant interaction between grammaticality and attractor. For grammatical conditions, the ERP response was significantly more negative when the attractor mismatched the gender of the determiner compared to when it matched. Although the negativity to ungrammatical sentences was slightly reduced when the attractor matched the gender of the determiner, as predicted by all accounts of attraction, the comparison between ungrammatical conditions was not significant.

In a follow-up study, Martin et al. (2014) investigated noun ellipsis in constructions like (5). Again, sentences contained either grammatical or ungrammatical ellipsis, and the gender of an attractor was manipulated in tandem. This study differed from their previous one in that the head noun, *el colgante* 'the necklace' in (5), was the object of the subsequent relative clause rather than the subject. Martin et al.'s argument is that this might alter the relative prominence of the target and attractor nouns in (4) as compared to (5), with the target being less prominent in (5) because it is an object in both the main and relative clause, whereas in (4) it was the relative clause's subject.

(5)   *Spanish* (Martin et al. 2014)
        Rafaela perdió **el   colgante** que junto   con **el   anillo/ la   sortija**
        Rafaela lost   the(M) necklace(M) that together with the(M) ring(M) the(F) ring(F)
        siempre llevaba y   Mónica recuperó **otr-o/   otr-a**   que había perdido
        always wore   and Mónica recovered another-M another-F that had   lost
        años atrás.
        years back
        'Rafaela lost the necklace that together with the ring she always wore and Mónica
        recovered another that she had lost years before.'

Martin et al. (2014) reported an increased early negativity (100–400 post onset of the determiner) in grammatical sentences when the attractor matched as compared to when it mismatched the gender of the determiner. No differences emerged in this period between the ungrammatical conditions as a function of attractor-determiner (mis)match. In a later time window, they observed a grammaticality effect in the form of an increased positivity (P600) in response to ungrammatical as compared to grammatical sentences. This effect was qualified by a grammaticality by attractor (mis)match interaction. In close inspection, the interaction was driven by differences between responses to the two grammatical conditions, rather than the ungrammatical ones.

In sum, the results of these two studies suggest that attractor nouns influence ellipsis resolution. Although a consistent pattern of results was not observed across the two studies, this might be due to the attractor being in different syntactic positions in the two experiments (Martin et al. 2014). While these studies do suggest that the gender of an attractor may influence gender agreement during comprehension, whether the precise pattern of results is more compatible with cue-based retrieval rather than representational accounts of attraction is less

clear. Additionally, although an asymmetry in attraction effects was found in grammatical and ungrammatical sentences, the precise pattern of results was different to other studies on agreement attraction. It might be that gender plays a different role during memory access for ellipsis constructions than other types of agreement. Note also that, in both studies, other constituents in the sentence, such as *Miren* in (4) and *Mónica* in (5), along with the pronouns in (4), also carry gender marking that may have interfered in how the ellipsis was resolved. Further research on noun ellipsis in Spanish that controls for these issues would be warranted in this case.

To our knowledge, only one existing study has investigated noun-adjective gender agreement in Spanish comprehension. Acuña-Fariña et al. (2014) tested sentences such as (6) during reading while participants' eye-movements were monitored. They manipulated gender and number agreement, testing grammatical sentences only.

(6)   *Spanish* (Acuña-Fariña et al. 2014)
   a.   El   nombre   del   niñ-o   (de los niñ-o-s) / de la   niñ-a   (de
        the   name(M).SG of.the child-M.SG of the child-M-PL of the child-F.SG of
        las   niñ-a-s)   era   alemán [...]
        the   child-F-PL was   German.M
        'The name of the child(ren) was German [...]'

   b.   Los   nombre-s del   niñ-o   (de los niñ-o-s) / de la   niñ-a
        the   name(M)-PL of.the child-M.SG of the child-M -PL of the child-F.SG
        (de   las   niñ-a-s)   eran   alemán [...]
        of   the   child-F-PL were   German.M
        'The names of the child(ren) were German [...]'

For number agreement, Acuña-Fariña et al. reported longer reading times when the attractor mismatched the number of the adjective compared to when it matched. Importantly, because their stimuli consisted of grammatical sentences only, this means that the attractor also mismatched the number of the subject's head noun in these conditions. While these results might be taken to support representation-based accounts of attraction, note that this effect was most noticeable at the verb before the adjective (*era* 'was'), rather than at the adjective itself. As such, these results may be confounded with the plural penalty reported in previous studies (e.g., Wagers et al. 2009).

For gender agreement, Acuña-Fariña et al. reported significantly more regressions out of the adjective when the attractor mismatched the gender of the adjective compared to when it matched. This effect may again be considered more consistent with representational accounts of attraction rather than retrieval-based accounts. Note, however, that in their inferential statistical analysis of gender attraction Acuña-Fariña et al. collapsed their data across number matching and mismatching conditions. As such, it may be difficult to establish with enough certainty that their estimates for effects of gender attraction are not influenced by effects of number attraction across the different gender conditions. Indeed, cue-based parsing would predict less interference from attractor nouns that match in gender but not number agreement with an adjective than from attractors that match in both features, but it is not possible to draw any conclusions on this issue based on the reported analyses. Note also that, since this study only tested grammatical sentences, it obviously cannot provide insight into a potential *asymmetry* in gender agreement attraction in grammatical and ungrammatical sentences, which is crucial when teasing apart different accounts of agreement attraction in comprehension.

2 THE CURRENT STUDY

The current study investigated agreement attraction in Spanish noun-adjective gender agreement during comprehension. By examining this understudied domain of grammatical agreement, we aimed to tease apart representational and retrieval-based accounts of attraction in comprehension. In particular, we tested for a grammatical asymmetry in attraction effects, as predicted by cue-based parsing, across five experiments. Experiment 1a investigated nouns with biological gender in an untimed acceptability judgement task, while Experiment

1b investigated the time-course of processing with the same materials in an eye-movement study. Experiments 2a and 2b attempted to generalize our findings from Experiments 1a and 1b, testing nouns with lexical gender instead. Finally, Experiment 3 compared both biological and lexical gender within a single speeded judgement paradigm. The design in each experiment directly compared attraction in grammatical and ungrammatical sentences, to tease apart retrieval and representation-based accounts of attraction in comprehension.

2.1 EXPERIMENT 1A

Experiment 1a was an offline acceptability judgement task with four conditions as in (7).

(7)   *Spanish*
Grammatical, Attractor Match
a.   La   compañer-a de la   carter-a   parecía muy content-a de poder
        the partner-F   of the postwoman-F seemed very happy-F   to able
        ayudar a   repartir.
        help   to deliver
        'The colleague of the postwoman seemed very happy to help with the delivery.'

Grammatical, Attractor Mismatch
b.   La compañer-a del   carter-o   parecía muy content-a de poder
        the partner-F   of.the postman-M seemed very happy-F   to able
        ayudar a   repartir.
        help   to deliver
        'The colleague of the postman seemed very happy to help with the delivery.'

Ungrammatical, Attractor Match
c.   *El   compañer-o de la   carter-a   parecía muy content-a de poder
        the partner-M   of the postwoman-F seemed very happy-F   to able
        ayudar a   repartir.
        help   to deliver
        'The colleague of the postwoman seemed very happy to help with the delivery.'

Ungrammatical, Attractor Mismatch
d.   *El   compañer-o del   carter-o   parecía muy content-a de poder
        the partner-M   of.the postman-M seemed very happy-F   to able
        ayudar a   repartir.
        help   to deliver
        'The colleague of the postman seemed very happy to help with the delivery.'

In the example sentences above, (7a/b) are both grammatical as the subject's head noun (*compañera*) matches in gender with the adjective *contenta*, while (7c/d) are ungrammatical as the head noun (*compañero*) mismatches the gender of the adjective (*contento* would be grammatical). Gender match between the attractor and adjective is also manipulated. Following conventions in the comprehension literature on attraction, we use the terms *match/mismatch* to refer to agreement congruency between the adjective and the attractor noun, rather than the feature-(mis)match between the head noun and the attractor. As such, in (7a/c) the attractor matches the gender of the adjective, while in (7b/d) it mismatches.

The main aim of Experiment 1a was to ensure that the materials to be tested in our eye-movement study, Experiment 1b, displayed the intended range of acceptability. As such, we expected participants to rate ungrammatical sentences as less acceptable than grammatical sentences. Some studies have however reported attraction in untimed offline judgements, both in number and gender agreement, in at least English, German and Spanish (e.g. Häussler 2009; Xiang, Grove & Giannakidou 2013; Fuchs et al. 2015; Scontras, Polinsky & Fuchs 2018). As such, if attraction influences offline sentence judgements for Spanish gender agreement, we should observe facilitatory interference, such that (7c) should be rated as more acceptable than (7d). Cue-based parsing would predict this effect in ungrammatical sentences only. If attraction is a result of a faulty representation of the gender of the attractor, we should also find evidence of attraction in grammatical sentences, such that (7b), where the attractor mismatches the gender of the adjective, should be rated as less acceptable than (7a), where

the attractor matches. Alternatively, if attraction effects in gender agreement do not persist to offline judgements, we may observe a significant main effect of sentence grammaticality only.

2.1.1 Participants

32 native Spanish speakers (12 males, mean age 29), who were recruited via the internet and were originally from Spain, voluntarily took part in the experiment. Across our five experiments, criteria for selection included being a native speaker of some variety of Spanish, and having been born and raised in a Spanish-speaking country.

2.1.2 Materials

32 experimental items as in (7) were created (see *https://osf.io/hm8yn/* for a full list). Each consisted of a single sentence containing a predicative adjective, and which factorially manipulated sentence grammaticality and attractor (mis)match. The critical adjective was always modified by an adverb to increase distance between it and the attractor, such that any potential spillover effects from the different attractors across conditions could be minimised during processing in Experiment 1b. Gender agreement was always manipulated using the biological gender of the sentence subject's head and attractor noun. Half of the items contained an adjective marked with masculine gender and half with feminine gender.[1]

In addition to the experimental items, 32 filler sentences were also constructed that consisted of a variety of different syntactic structures. Half of the filler sentences were grammatical, and half were ungrammatical.

2.1.3 Procedure and Data Analysis

Experimental and filler items were pseudo-randomised across four lists in a Latin-square design, with a different order being presented to each participant. The experiment was administered online using Google Forms, with participants completing the experiment in their own time at a location of their choosing. Participants were instructed to simply read each sentence and rate it on a scale from 1–5, with one meaning 'completely unacceptable' and five meaning 'completely acceptable'. No reference was made to *grammaticality* in the instructions, which always referred to the *acceptability* or *validity* of the sentences, and did not convey the existence of precisely right or wrong answers. In each case, the sentence appeared onscreen with the rating scale appearing below. All sentences appeared on a single webpage which the participant scrolled down. Participants were able to revisit and undo answers.[2]

Before analysis, acceptability ratings were z-score transformed. Z-scores were calculated for each participant separately based on all sentences, both experimental items and fillers, in the experiment. Analysis was conducted using a linear mixed-effects model with crossed random effects for subjects and items (Baayen, Davidson & Bates 2008). The model included sum coded (-1/1) fixed main effects of sentence grammaticality (grammatical vs. ungrammatical), attractor (match vs. mismatch), and their interaction. The model was fit using the 'maximal' random effects structure (Barr, Levy, Scheepers & Tily 2013) that converged. In each experiment reported here, we first fit a maximal model with by-subject and by-item random intercepts, random slopes for all repeated measures for subjects and items, and random correlation parameters. If this model did not converge, we refit the model without random correlation parameters. If this model still did not converge, we iteratively removed the random effect that accounted for the least variance until convergence was achieved. For each fixed effect, *p* values

---

[1]   Some studies on gender agreement attraction have queried whether there is a markedness asymmetry between masculine/feminine gender (e.g. Vigliocco, Butterworth & Garrett 1996; Badecker & Kuminiak 2007; Slioussar & Malko 2016), as has been argued for number agreement (e.g. Kimball & Aissen 1971; Eberhard 1997; Pearlmutter et al. 1999). Although the markedness asymmetry appears robust for number (e.g., Pearlmutter et al. 1999; Wagers et al. 2009), at least in production (Bock & Miller 1991; Bock & Cutting 1992; Eberhard 1997), it has not been robustly observed in gender agreement (e.g., Vigliocco et al. 1996; cf. Badecker & Kuminiak 2007; Scontras et al. 2018). As our research questions were related to attraction in general rather than any potential markedness asymmetry, we adopted a design with equal numbers of masculine/feminine adjectives that ensured the critical adjective region was identical across conditions. For this reason, we do not discuss the potential gender markedness asymmetry any further.

[2]   As an anonymous reviewer points out, this allowed participants to potentially violate item order and re-rate a sentence based on subsequent exposure. We are unsure as to how often this type of revision might have occurred in our study, however, given the absence of explicit feedback. While we do not expect this to have changed the overall patterns of the data, we acknowledge this limitation.

were calculated using the Satterthwaite approximation (see Luke 2017). Data and analysis scripts for all experiments reported in this paper can be found at the second author's Open Science Framework webpage (*https://osf.io/hm8yn/*).

2.1.4 Results and Discussion

Mean acceptability ratings are shown in *Table 1*. Analysis revealed a significant main effect of grammaticality (estimate = –0.798, SE = 0.038, t = –20.94, p < .001), with ungrammatical sentences being rated as less acceptable than grammatical sentences. Neither the main effect of attractor (estimate = –0.005, SE = 0.016, t = –0.32, p = .753), nor its interaction with grammaticality was significant (estimate = –0.014, SE = 0.014, t = –0.98, p = .337).

	EXPERIMENT 1A	EXPERIMENT 2A	EXPERIMENT 3	
			BIOLOGICAL GENDER	LEXICAL GENDER
Grammatical, Attractor Match	4.07 (0.07)	4.08 (0.07)	0.90 (0.01)	0.90 (0.01)
Grammatical, Attractor Mismatch	4.10 (0.07)	4.11 (0.07)	0.91 (0.01)	0.88 (0.01)
Ungrammatical, Attractor Match	1.36 (0.04)	1.64 (0.06)	0.87 (0.01)	0.90 (0.01)
Ungrammatical, Attractor Mismatch	1.30 (0.04)	1.57 (0.06)	0.93 (0.01)	0.93 (0.01)

Table 1 Mean acceptability ratings for Experiments 1a and 2a (1 = totally unacceptable, 5 = totally acceptable) and mean proportions correct in Experiment 3 (standard errors in parentheses).

These results indicate clear grammaticality effects but no evidence of agreement attraction influencing acceptability ratings in this untimed task. We return to this lack of attraction in our offline task, especially in comparison to existing English data on subject-verb agreement (Xiang et al. 2013), in the General Discussion. Experiment 1b investigated whether gender agreement attraction could be found during processing with the same materials by monitoring participants' eye-movements during reading.

2.2 EXPERIMENT 1B

In Experiment 1b participants read the 32 sentences in four conditions as in (7), while their eye-movements were monitored. Retrieval and representation-based accounts of attraction make different predictions regarding the time-course of processing. Both accounts would predict longer reading times for ungrammatical sentences like (7c/d) than grammatical sentences like (7a/b). Retrieval based accounts would then predict a grammatical asymmetry, with facilitatory interference in ungrammatical sentences only. Specifically, this would predict shorter reading times in (7c) compared to (7d). Inhibitory interference, with longer reading times in (7a) than (7b), would be predicted for grammatical sentences, though we note that this effect may be small and difficult to detect (Nicenboim et al. 2018). Representational theories would predict attraction in both grammatical and ungrammatical sentences. This would result in shorter reading times for (7c) compared to (7d), and also (7a) compared to (7b). Thus, the crucial difference between the two accounts is whether the attractor manipulation emerges as a main effect or an interaction with grammaticality. Grammaticality by attractor interactions are most clearly explained by retrieval-based models of attraction during comprehension, but would be unexpected from the perspective of representation-based accounts.

2.2.1 Participants

32 native Spanish speakers (9 males, mean age 30) from Spain or Latin America, none of whom had completed Experiment 1a, were paid to take part in Experiment 1b. Participants were recruited from the University of Reading and surrounding community.

2.2.2 Materials

The same 32 experimental items from Experiment 1a were used in Experiment 1b. In addition to these 32 experimental items, we also constructed 96 grammatical filler texts, some of which comprised more than one sentence, that contained a variety of different syntactic structures. Experimental items were always displayed across a single line of text onscreen, while fillers took up between one and three lines. Comprehension questions requiring a yes/no response

were asked after all experimental trials and half of the fillers, to make sure that participants were reading for meaning. Half of the comprehension questions required a 'yes' response, and half 'no'. To avoid bringing attention to our experimental manipulations, comprehension questions on experimental trials never probed the critical dependency (e.g., *El entrevistador del candidato era bastante misterioso y preguntaba muy poco* 'The interviewer of the candidate was quite mysterious and asked very little [=few questions].' Question: *¿Hubo menos preguntas de lo normal en la entrevista?* 'Were there fewer questions than usual in the interview?').

2.2.3 Procedure and Data Analysis

Items were pseudo-randomised across four lists in a Latin-square design, and a different order was presented to each participant. Eye movements were recorded with an EyeLink 1000 eye-tracker, sampling at 1000Hz. While viewing was binocular, eye-movements were recorded from the right eye only. Each session began with calibration on a nine-point grid, and recalibration between trials was conducted if required. Four practice trials were presented to participants before the 32 experimental and 96 filler trials. Before each trial, participants fixated on a marker above the first word of the upcoming trial. Upon fixation on this marker, the text appeared. Participants read each text silently, pressing a button on a control pad once completed, and comprehension questions were answered by pressing one of two buttons on the control pad. The entire experiment lasted 30–40 minutes.

Reading times were calculated for two regions of text. The *critical region* consisted of the predicative adjective (*contenta*, in example 7), while the *spillover region* consisted of the rest of the sentence up to but not including the final two words (*de poder ayudar*). We excluded the final two words from analysis to prevent any end of trial artefacts from affecting reading times. Three reading time measures were calculated at each region. *First-pass time* sums fixation durations within a region when it is first entered from the left up until it is exited to either the left or right, while *regression path time* sums fixation durations starting from when it is first entered from the left, up until but not including the first fixation of a region to the right. We also calculated *total viewing time*, the sum of all fixations within a region. Regions that were initially skipped during reading were treated as missing data in first-pass and regression path times, and regions that received no fixations at all were treated as missing data in total viewing times. Short fixations of 80 ms or below within one degree of visual arc of another fixation were merged. All other fixations of 80 ms or below, as well as those above 800 ms, were removed before analysis. Less than 1% of trials were removed due to track loss.

Analysis was conducted using mixed-effects models fit with the maximal random effects structure that converged. The models were fit to log-transformed reading times, to remove skew and to normalise model residuals (Vasishth & Nicenboim 2016). As in Experiment 1a, analyses included sum-coded fixed effects for grammaticality (grammatical vs. ungrammatical) and attractor (match vs. mismatch). To minimise the number of separate statistical tests conducted at different regions, we conducted one analysis for each measure including *region* (critical vs. spillover) as a (sum-coded) fixed effect. Thus, the maximal model included by-subject and by-item random intercepts and slopes for grammaticality, attractor, region and their interactions. As this analysis involves having two non-independent datapoints from a single trial, we additionally included random intercepts of *trial*, defined as the unique subject and item pairing that constituted an individual trial. As region is the only repeated measure at the level of the trial, we included by-trial random slopes for region only (for further discussion of using region as a fixed effect, see Cunnings & Sturt 2018).

Given that the lexical material across regions differs, we do not discuss main effects of region below. Other main effects and interactions however are indicative of the time-course of grammaticality and interference/attraction effects. In the case of grammaticality by attractor interactions, planned comparisons were conducted using nested contrasts to examine attractor effects at the two levels of sentence grammaticality.

2.2.4 Results

Average question accuracy across all trials was 87% correct (all above 76%), indicating that participants paid attention to the content of the sentences. Summaries of the reading times and statistical analyses are provided in *Tables 2* and *3*.

	FIRST PASS TIME		REGRESSION PATH TIME		TOTAL VIEWING TIME	
Critical Region						
Grammatical, Attractor Match	306	(8)	444	(29)	446	(16)
Grammatical, Attractor Mismatch	311	(8)	398	(26)	466	(18)
Ungrammatical, Attractor Match	328	(11)	521	(37)	565	(23)
Ungrammatical, Attractor Mismatch	332	(10)	555	(44)	608	(27)
Spillover Region						
Grammatical, Attractor Match	379	(14)	462	(32)	599	(25)
Grammatical, Attractor Mismatch	384	(12)	476	(24)	581	(21)
Ungrammatical, Attractor Match	387	(13)	593	(43)	679	(30)
Ungrammatical, Attractor Mismatch	382	(16)	770	(59)	680	(27)

Table 2 Summary of eye-movement measures in Experiment 1b (in milliseconds, standard errors in parentheses).

	MODEL ESTIMATE (SE)		t	p
First Pass Time				
Region	0.062	(0.025)	2.46	.018
Grammaticality	0.010	(0.010)	1.05	.309
Attractor	0.004	(0.010)	0.47	.645
Region * Grammaticality	–0.015	(0.010)	–1.47	.155
Region * Attractor	–0.008	(0.010)	–0.84	.413
Grammaticality * Attractor	–0.006	(0.010)	–0.57	.575
Region * Grammaticality * Attractor	–0.011	(0.009)	–1.15	.261
Regression Path Time				
Region	0.073	(0.036)	2.03	.048
Grammaticality	0.089	(0.013)	7.12	<.001
Attractor	0.024	(0.012)	1.98	.048
Region * Grammaticality	0.018	(0.013)	1.34	.181
Region * Attractor	0.024	(0.016)	1.49	.151
Grammaticality * Attractor	0.023	(0.012)	1.87	.071
Region * Grammaticality * Attractor	–0.012	(0.015)	–0.76	.455
Total Viewing Time				
Region	0.092	(0.033)	2.74	.009
Grammaticality	0.084	(0.016)	5.26	<.001
Attractor	0.014	(0.012)	1.24	.215
Region * Grammaticality	–0.021	(0.009)	–2.29	.035
Region * Attractor	–0.010	(0.010)	–1.00	.331
Grammaticality * Attractor	0.004	(0.012)	0.36	.717
Region * Grammaticality * Attractor	–0.005	(0.010)	–0.53	.604

Table 3 Summary of the statistical analysis in Experiment 1b.

No significant effects were observed in first-pass times. In regression path times, there was a significant main effect of grammaticality, with longer reading times in ungrammatical sentences, and a significant main effect of attractor, with longer reading times when the attractor mismatched. These two main effects were however modulated by a marginal grammaticality by attractor interaction.

To examine this interaction, nested contrasts were conducted at the two levels of grammaticality. These revealed that there were no significant differences between the two grammatical

conditions (estimate = 0.001, SE = 0.017, *t* = 0.06, *p* = .949), but that regression path times were significantly shorter for ungrammatical sentences when the attractor matched the gender of the adjective (estimate = 0.047, SE = 0.017, *t* = 2.64, *p* = .013). This pattern of results is shown in *Figure 1*, and is indicative of facilitatory interference in ungrammatical sentences only, similar to what has previously been observed in subject-verb number agreement (e.g., Wagers et al. 2009; Lago et al. 2015).
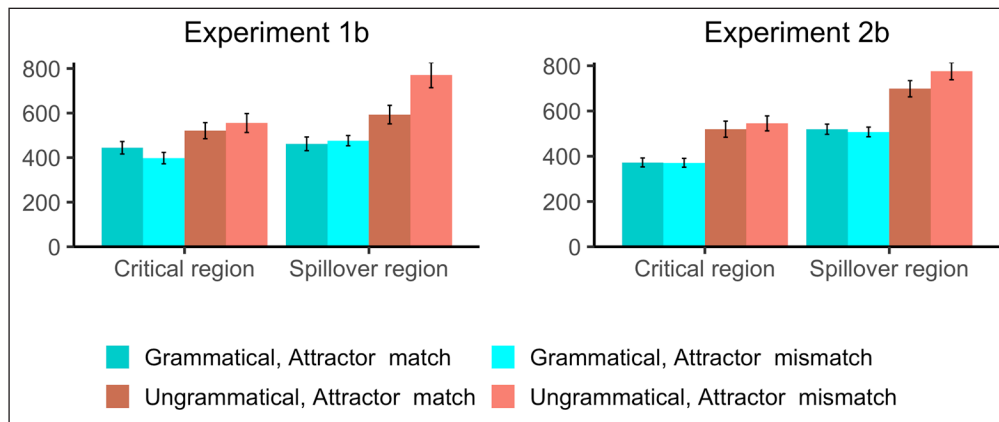
In total viewing time, we observed a significant main effect of grammaticality, with longer total viewing times for ungrammatical sentences. There was also a significant grammaticality by region interaction, as the grammaticality effect was larger at the critical region than the spillover region.

2.2.5 Discussion

The results of Experiment 1b revealed grammaticality effects in both regression path and total viewing times. Additionally, we found suggestive evidence of attraction, with shorter regression path times for ungrammatical sentences when the attractor matched the gender of the adjective. However, the attractor did not significantly influence reading times in grammatical sentences in any measure. This suggests facilitatory interference in ungrammatical sentences only. This pattern of results, and the asymmetry of attraction effects in ungrammatical vs. grammatical sentences, is more compatible with retrieval-based accounts such as cue-based parsing (Lewis et al. 2006; see, e.g., Wagers et al. 2009; Dillon et al. 2013; Lago et al. 2015) than with representation-based accounts.

Together, the results of Experiments 1a and 1b suggest that retrieval interference may influence online sentence processing, but that such effects appear not to persist to untimed, offline judgements. These results however are limited to nouns with biological gender. We thus devised Experiments 2a and 2b with the primary aim of acting as a conceptual replication of Experiments 1a and 1b, to test nouns marked with lexical rather than biological gender.

2.3 EXPERIMENT 2A

Experiment 2a was an acceptability judgement task as in Experiment 1a, but manipulated lexical gender, in four conditions as in (8).

(8)   *Spanish*
         Grammatical, Attractor Match
         a.   La   madera de la   puerta   era   realmente dur-a   y   aguantaba sin
              the wood(F) of the door(F) was   really   hard-F and resisted   without
              problemas.
              problems
              'The wood of the door was really hard and resisted without problems.'

         Grammatical, Attractor Mismatch
         b.   La   madera del   cuadro   era   realmente dur-a   y   aguantaba
              the wood(F) of.the painting(M) was   really   hard-F and resisted
              sin   problemas.
              without problems
              'The wood of the painting was really hard and resisted without problems.'

Ungrammatical, Attractor Match

c. *El   marco   de la puerta era realmente dur-a y   aguantaba
    the frame(M) of the door(F) was really   hard-F and resisted
    sin   problemas.
    without problems
    'The door frame was really hard and resisted without problems.'

Ungrammatical, Attractor Mismatch

d. *El   marco   del   cuadro   era realmente dur-a y   aguantaba
    the frame(M) of.the painting(M) was really   hard-F and resisted
    sin   problemas.
    without problems
    'The frame of the painting was really hard and resisted without problems.

As in Experiment 1a, conditions (8a/b), where the feminine sentence subject's head noun (*madera*) matches the gender of the adjective (*dura*), are grammatical, while conditions (8c/d), where the subject's head noun (*marco*) is masculine, are ungrammatical. Additionally, an attractor (*puerta*) matches the gender of the adjective in (8a/c) but mismatches (*cuadro*) in (8b/d). In each case, the sentence subject and attractor were nouns in which lexical gender was manipulated. Because nouns with lexical gender do not alternate between masculine and feminine forms, as is the case with nouns with biological gender, each item uses two pairs of lexical items, instead of the single pair in Experiments 1a and 1b. This is an inescapable requirement of manipulations of lexical gender alone (see Antón-Méndez et al. 2002, for discussion; see also Franck et al. 2008, for similar manipulations). As in Experiment 1a, we expected ungrammatical sentences to receive lower ratings than grammatical sentences. Any interactions between sentence grammaticality and the attractor would be indicative of agreement attraction.

2.3.1 Participants

32 native Spanish speakers (11 males, mean age 35), who were recruited via the internet as in Experiment 1a and were from Spain or Latin America, voluntarily took part in Experiment 2a.

2.3.2 Materials

32 experimental items like (8) were created (see *https://osf.io/hm8yn/* for a full list). As in Experiment 1a, each consisted of a single sentence containing a predicative adjective, always modified by an adverb, with four versions that factorially manipulated sentence grammaticality and attractor gender (mis)match. Gender agreement was always manipulated using lexical gender alone. Half of the items contained an adjective marked with masculine gender and half with feminine gender. As different nouns were used across the masculine/feminine manipulations, each noun was used twice, once with a masculine adjective and once with a feminine adjective. In addition to the experimental items, 32 filler sentences were also constructed as in Experiment 1a.

To ensure that the naturalness of the different preambles across the four conditions of a given item did not induce different reading times in the pre-critical regions, we conducted an online norming study with 163 native speakers of Spanish. Participants provided a 1 to 5 Likert-scale judgement (1 worst, 5 best) on the naturalness/plausibility of the complex NPs that constitute our sentence preambles, which were mixed with (more) clearly unnatural/implausible NPs (e.g., *the musician under the lion* or *the pacifist with the nuclear weapon*). Descriptive statistics by condition, shown in **Table 4**, strongly suggest that there are barely any differences in the perception of naturalness or plausibility across the complex subjects of our four conditions:

CONDITION	EXAMPLE NOUN PHRASE	MEAN	MEDIAN	*SD*
Grammatical, Attractor Match	*El castillo del bosque*	3.85	3.93	0.86
Grammatical, Attractor Mismatch	*El castillo de la colina*	3.92	4.11	0.78
Ungrammatical, Attractor Match	*La iglesia del bosque*	3.94	3.96	0.74
Ungrammatical, Attractor Mismatch	*La iglesia de la colina*	3.97	4.29	0.82

Table 4 Descriptive statistics (mean 1–5 rating, median and standard deviations) per condition in the online norming study.

Note: "Condition" refers here to the experimental conditions used in Experiments 2a/b, but these manipulations were not part of the norming study.

2.3.3 Procedure and Data Analysis

The procedure of the experiment and data analysis was the same as in Experiment 1a.

2.3.4 Results and Discussion

The results are presented in *Table 1*. We observed a significant main effect of grammaticality (estimate = –0.747, SE = 0.036, t = –20.49, p < .001), in the absence of any other significant effects (for the main effect of attractor, estimate = –0.008, SE = –0.019, t = 0.43, p = .672; for the interaction, estimate = –0.017, SE = 0.019, t = –0.90, p = .375). As in Experiment 1a, the results of Experiment 2a indicate that ungrammatical sentences were rated as significantly less acceptable than grammatical sentences, but we again found no evidence of attraction in this untimed offline task. Experiment 2b tested for gender agreement attraction during online sentence processing.

2.4 EXPERIMENT 2B

In Experiment 2b, participants read 32 sentences as in (8), plus 96 fillers, while their eye-movements were monitored. From the perspective of cue-based parsing, any retrieval cues derived from the critical predicative adjective in our experimental sentences should influence retrieval processes for sentences containing biological and lexical gender similarly. In both cases, for example, the retrieval cues will include both syntactic constraints (e.g., [+head]) and cues generated from the number and gender properties of the adjective (e.g., [+masculine]). From this perspective, the predictions for Experiment 2b should be the same as Experiment 1b.

2.4.1 Participants

32 native Spanish speakers (10 males, mean age 28) from Spain or Latin America from the University of Reading and surrounding community were paid to take part in Experiment 2b. None had taken part in any of the previously reported experiments.

2.4.2 Materials

The 32 experimental items from Experiment 2a were used in Experiment 2b. The same 96 fillers from Experiment 1b were also used in Experiment 2b. All other aspects of the materials were the same as Experiment 1b.

2.4.3 Procedure and Data Analysis

The procedure and data analysis were identical to Experiment 1b. No trials were removed due to track loss.

2.4.4 Results

Average question accuracy was 89% correct (all above 73%). Summaries of the reading times and statistical analysis are provided in *Tables 5* and *6*.

	FIRST PASS TIME		REGRESSION PATH TIME		TOTAL VIEWING TIME	
Critical Region						
Grammatical, Attractor Match	305	(11)	373	(21)	422	(19)
Grammatical, Attractor Mismatch	284	(10)	371	(21)	389	(18)
Ungrammatical, Attractor Match	305	(11)	520	(37)	553	(25)
Ungrammatical, Attractor Mismatch	333	(14)	545	(34)	563	(23)
Spillover Region						
Grammatical, Attractor Match	482	(20)	519	(23)	685	(31)
Grammatical, Attractor Mismatch	452	(17)	508	(21)	663	(29)
Ungrammatical, Attractor Match	484	(20)	698	(36)	794	(31)
Ungrammatical, Attractor Mismatch	462	(18)	776	(39)	779	(29)

Table 5 Summary of eye-movement measures in Experiment 2b (in milliseconds, standard errors in parentheses)

	MODEL ESTIMATE (SE)		t	p
First Pass Time				
Region	0.198	(0.033)	5.94	<.001
Grammaticality	0.011	(0.010)	1.17	.244
Attractor	–0.012	(0.010)	–1.23	.218
Region * Grammaticality	–0.021	(0.012)	–1.79	.084
Region * Attractor	–0.011	(0.011)	–0.96	.346
Grammaticality * Attractor	0.014	(0.013)	1.12	.275
Region * Grammaticality * Attractor	–0.015	(0.010)	–1.55	.131
Regression Path Time				
Region	0.185	(0.039)	4.73	<.001
Grammaticality	0.131	(0.015)	8.60	<.001
Attractor	0.013	(0.012)	1.11	.275
Region * Grammaticality	0.006	(0.014)	0.41	.687
Region * Attractor	0.007	(0.016)	0.42	.682
Grammaticality * Attractor	0.025	(0.013)	1.96	.064
Region * Grammaticality * Attractor	0.004	(0.012)	0.36	.721
Total Viewing Time				
Region	0.225	(0.038)	5.92	<.001
Grammaticality	0.119	(0.017)	7.17	<.001
Attractor	–0.008	(0.012)	–0.63	.531
Region * Grammaticality	–0.031	(0.009)	–3.58	<.001
Region * Attractor	0.007	(0.010)	0.70	.493
Grammaticality * Attractor	0.018	(0.012)	1.55	.133
Region * Grammaticality * Attractor	–0.014	(0.009)	–1.58	.115

Table 6 Summary of the statistical analysis in Experiment 2b

In first-pass reading times, there was a marginal grammaticality by region interaction, indicative of grammaticality effects, with longer reading times in ungrammatical sentences, at the critical region. In regression path times, there were significant main effects of grammaticality and attractor, that were modulated by a marginal grammaticality by attractor interaction. Nested contrasts revealed no significant differences between the two grammatical conditions (estimate = –0.011, SE = 0.015, t = –0.73, p = .469). For the ungrammatical conditions, reading times tended to be shorter when the attractor matched the gender of the adjective, with the contrast between the ungrammatical conditions being marginally significant (estimate = 0.038, SE = 0.020, t = 1.90, p = .071).

In total viewing time, there was a significant main effect of grammaticality, with longer reading times in ungrammatical conditions. The grammaticality by region interaction was also significant, with a larger grammaticality effect at the critical region.

2.4.5 Discussion

Experiment 2b revealed significant main effects of grammaticality, with reliably longer reading times in ungrammatical sentences. As in Experiment 1b, regression path times suggested facilitatory interference in ungrammatical sentences only. This is illustrated in *Figure 1*.

Given the marginal interactions in Experiments 1b and 2b in regression path times, to maximise statistical power across our two eye-tracking experiments, we conducted an additional analysis that compared the regression path results from the two experiments. This analysis included fixed effects of region, grammaticality and attractor, and additionally a sum-coded fixed effect of experiment, plus all interactions. Experiment was treated as a between subjects

and between items manipulation, and the model was fit with the maximal random effects structure that converged.

The results of this model revealed significant main effects of region (estimate = 0.129, SE = 0.027, t = 4.84, p < .001), grammaticality (estimate = 0.110, SE = 0.010, t = 11.24, p < .001) and attractor (estimate = 0.018, SE = 0.008, t = 2.30, p = .022). There was also a significant experiment by grammaticality interaction (estimate = 0.021, SE = 0.010, t = 2.11, p =.041), with larger grammaticality effects for items with lexical gender. Importantly, there was a significant grammaticality by attractor interaction (estimate = 0.024, SE = 0.008, t = 2.80, p = .007), in the absence of any further significant main effects or interactions (all t < 1.34, all p > .188). Nested contrasts indicated that there were no significant differences between the two grammatical conditions (estimate = –0.005, SE = 0.011, t = –0.44, p = .663), while for ungrammatical sentences regression path times were significantly shorter when the attractor matched the gender of the adjective compared to when it mismatched (estimate = 0.042, SE = 0.013, t = 3.21, p = .002). These results suggest facilitatory interference in ungrammatical sentences only, across both eye-tracking experiments.

In sum, this combined analysis, which maximises our statistical power, revealed attraction effects in ungrammatical but not grammatical sentences. Thus, we maintain that our eye-tracking results for both biological and lexical gender indicate facilitatory interference in ungrammatical sentences, as predicted by cue-based parsing. Given that this finding is based on an additional analysis of both Experiments 1b and 2b combined, to replicate our main finding of a grammaticality by attractor interaction, we conducted a final experiment that tested biological and lexical gender in a single experiment, using a speeded judgement task.

2.5 EXPERIMENT 3

In Experiment 3, to maximise statistical power, we tested all of the experimental items from the previous experiments within a single speeded judgement task. Although our two untimed judgement experiments did not reveal any significant attraction effects, previous studies using a speeded forced choice judgement paradigm have reported attraction in speeded judgement tasks (e.g., Wagers et al. 2009; Schlueter et al. 2018). In Experiment 3, participants made a speeded judgement as to whether each sentence was grammatical or ungrammatical.

One notable change with respect to our previous experiments was that in Experiment 3 the critical sentences were truncated to end at the adjective. This allows a judgement very soon after presentation of the critical word itself. Some previous speeded judgement studies on attraction have involved judgements occurring some time after the critical word (Wagers et al. 2009; Schlueter et al. 2018). This is likely imposed by their manipulation of agreement attraction on auxiliary verbs, which, in canonical sentences, require subsequent material (e.g., 'The key to the cabinets was/were rusty after many years of disuse', from Wagers et al. 2009). However, with the predicative adjectives used in the current study, we are able to test for attraction in the speeded judgement paradigm at the critical word itself.

2.5.1 Participants

76 native Spanish speakers (30 males, mean age 32), originally from Spain or Latin America, participated in the experiment. Participants were either recruited via social media/email and took part voluntarily, or were paid a small sum to complete the experiment via the online participant recruitment platform Prolific.

2.5.2 Materials

64 experimental items were used in Experiment 3, consisting of the two sets of 32 experimental items from Experiments 1 and 2. The experimental sentences were truncated so that they ended at the adjective with no additional spillover text. Some items were reworded slightly to ensure all sentences ended felicitously at the adjective. In addition to the 64 experimental items, 128 filler sentences were also constructed consisting of a variety of different syntactic structures. Half of the fillers were grammatical, and half ungrammatical. The ungrammatical fillers were designed to reduce the saliency of the gender violation in the critical items, and so contained different types of errors (e.g., word order violations, other morphosyntactic errors, word substitutions) that would yield the sentences unacceptable.

2.5.3 Procedure and Data Analysis

The experimental and filler sentences were pseudo-randomised across four different lists in a Latin-square design, with a different random order being presented to each participant. The experiment was implemented in Ibex Farm (*www.spellout.net/ibexfarm*), a web-based platform for psycholinguistic experiments. Participants were instructed that they would read a series of sentences one word at a time, and that they had to decide whether they thought each sentence was grammatical or ungrammatical as quickly and as accurately as possible. Each sentence was presented to participants word by word, with each new word replacing the previous one in the centre of the screen. The pacing was 400 ms per word, with no blank screens between words. After the last word in a sentence, a question mark appeared onscreen signalling that the participant had to provide a response as quickly as possible from that moment onwards. A timeout of 1500 ms was used to encourage fast responses. Feedback was given if the participant missed this timeout, but no feedback was given regarding the correctness of each response. At the beginning of each trial, a cross appeared onscreen and participants pressed a button to begin viewing the sentence. The entire experiment lasted approximately 20–30 minutes.

Timeouts, amounting to less than 3% of the experimental data, were removed before further analysis. As the dependent variable was a binary response, coded as correct/incorrect, analysis was conducted using a generalised linear mixed-effects model (Jaeger 2008). The analysis was similar to the previous experiments, with fixed effects of grammaticality and attractor, but additionally included *gender* (biological vs. lexical) as a within-subject but between-item sum-coded fixed effect, plus all interactions. The model was fit with the maximal random effects structure that converged.

2.5.4 Results

The results are shown in **Table 3**. Analysis revealed a significant main effect of attractor (estimate $= 0.14$, SE $= 0.06$, $t = 2.39$, $p = .017$) that was modulated by a significant grammaticality by attractor interaction (estimate $= 0.18$, SE $= 0.06$, $t = 3.02$, $p = .003$) in the absence of any further significant effects (all $t < 1.66$, all $p > .098$). Planned comparisons indicated no significant differences between the two grammatical conditions (estimate $= -0.03$, SE $= 0.11$, $t = -0.30$, $p = .765$), but that for ungrammatical sentences, accuracy was significantly lower when the attractor matched the gender of the adjective compared to when it mismatched (estimate $= 0.36$, SE $= 0.12$, $t = 2.92$, $p = .004$). This pattern of results suggests attraction in ungrammatical sentences. Although the size of this effect in percentage terms is very small, it is compatible with cue-based parsing. We discuss these results, along with the rest of our data, in more detail below.

3 GENERAL DISCUSSION

This study had two aims. The first was to test the generalisability of attraction effects previously observed during comprehension in subject-verb number agreement to noun-adjective gender agreement. The second was to compare competing accounts of agreement attraction, especially in their contrasting predictions with respect to asymmetries in attraction effects. We discuss how our results bear on these two aims in turn below.

3.1 GENDER ATTRACTION IN COMPREHENSION

Our results demonstrate noun-adjective gender agreement attraction during online sentence comprehension in Spanish (see also Paspali & Marinis 2020, for self-paced listening evidence for gender attraction in adjectival predicates in Greek). We observed attraction in participants' eye-movements during reading, as well as in accuracy rates in a speeded judgement task. During reading, evidence of attraction was found in the form of significantly shorter reading times in ungrammatical sentences when the attractor matched the gender of the adjective. In our speeded judgement task, accuracy in the ungrammatical conditions was significantly lower when the attractor noun matched the gender of the adjective compared to when it mismatched. In both our eye-movement experiments and speeded judgement task, attraction effects were restricted to ungrammatical sentences, mirroring the grammatical asymmetry found in previous studies of subject-verb agreement (e.g., Wagers et al. 2009; Shen et al. 2013;

Tanner et al. 2014; Slioussar & Malko 2016; Schlueter et al. 2018). As discussed in more detail below, we argue that this pattern of results is more consistent with retrieval than representation-based accounts of attraction.

We did not replicate the results reported by Acuña-Fariña et al. (2014) in grammatical sentences. An important difference between their design and the one in the eye-tracking experiments of the present study should be mentioned here, however. Acuña-Fariña et al. simultaneously manipulated number and gender (mis)match between the head noun, the attractor, and the verb, whereas our experiment kept (singular) number constant and focused exclusively on gender manipulations. Because Acuña-Fariña et al.'s results are effectively collapsed across conditions that manipulated both number and gender agreement, we argue that our results provide a clearer estimate of gender attraction. Likewise, Martin et al. (2012; 2014) reported a different pattern of effects in their ERP studies of gender attraction in ellipsis. As noted in the introduction, ellipsis resolution may differ from noun-adjective gender agreement. Future research more directly comparing ellipsis and noun-adjective agreement, ideally using the same experimental paradigm, may be a useful avenue to elucidate these varied results.

One difference between our eye-tracking experiments and some previous studies is that we did not observe significant attraction effects (as evidence of facilitatory interference) in total viewing times, instead observing the clearest evidence of this effect in the combined analysis of Experiments 1b and 2b in regression path times. This is perhaps surprising given that facilitatory interference in ungrammatical sentences has most consistently been found in total viewing times (Jäger et al. 2017; 2020). We note, however, that such effects have been observed in regression path times for some linguistic dependencies (e.g., Parker & Phillips 2017). We do not draw any strong conclusions about why we did not observe attraction effects in total viewing times, but note that further research into the time-course of attraction for gender agreement is required here.

Attraction effects were observed for both biological and lexical gender nouns. Indeed, we did not find any significant differences in the size of attraction effects between biological and lexical gender in our comparison of Experiments 1b and 2b, nor in our speeded judgement task. This is expected if attraction effects arise due to interference during retrieval. From the perspective of cue-based parsing, any interference arises as a result of the set of cues utilized at retrieval. In our study, the relevant gender cues marked on adjectives are the same irrespective of whether or not the sentence contains nouns marked with biological or lexical gender. From this perspective then, it naturally follows that attraction should be observed for sentences containing either biological or lexical gendered nouns.

Some differences between biological and lexical gender have been reported in production. For example, Vigliocco and Franck (1999) reported that biological gender head nouns were less prone to attraction from a lexical gender attractor, than head nouns with lexical gender (for discussion, see Bock 2003). Applying this to comprehension, differences between nouns marked with biological and lexical gender could potentially arise if the gender properties of these two types of nouns are represented differently. It might be, for example, that one type of noun leads to a stronger representation of the relevant gender feature in memory, which may then affect attraction. To tease these issues apart however, we would need to factorially manipulate head noun and attractor gender type (i.e. have biological gender head nouns with lexical gender attractors and vice versa), but this was not the aim of the present study. Further research is required to test for potential differences between how biological and lexical gender may affect processing during comprehension. While we as such cannot conclude that attraction effects for biological and lexical gender are identical based on the null effects reported here, our results do nonetheless indicate that the computation of agreement with both gender types is indeed susceptible to attraction in comprehension. Similarly, while we note that comparable attraction effects for both types of gender are expected from a memory retrieval perspective, other accounts (including representational ones) cannot be ruled out based on this finding.

The attraction effect we observed in the speeded judgement task was numerically small, and considerably smaller than previous speeded judgement studies on subject-verb number agreement. For example, the difference between ungrammatical conditions was approximately 5% in accuracy in our speeded judgement task, while Schlueter et al. (2018) reported an effect of approximately 30% in their speeded judgement experiments on English subject-verb number agreement. This is particularly striking given that Schlueter et al. tested for attraction several

words after the critical verb in their study, whereas we tested for effects at the critical adjective itself in our experiment. This might suggest that gender attraction effects in comprehension are smaller than number attraction effects. Some production work has discussed the possibility that these differences stem, among other things, from the fact that gender is lexically specified, whereas number is semantically determined for each utterance (e.g. Eberhard et al. 2005; Lorimor, Bock, Zalkind, Sheyman & Beard 2008). Alternatively, it might be that noun-adjective agreement is less susceptible to attraction than subject-verb agreement. One reason for this might be that retrieval of the sentence subject at the verb preceding the adjective in our study may have led to relatively high activation levels for the sentence subject, in turn leading to low levels of interference (see Dillon et al. 2013, for similar arguments about reflexives in English). This account would predict higher levels of interference for predicative adjectives in languages where the adjective precedes the verb. To our knowledge, however, this issue has not been previously tested. As such, it is difficult to tease these two possibilities apart based on our current data, but further research comparing gender and number attraction in both grammatical and ungrammatical sentences across these different dependencies would shed light on this issue.

Although we observed attraction in our speeded judgement task, no significant attraction effects were found in our two untimed judgement experiments. Here, ratings were significantly lower for ungrammatical than grammatical sentences, but this grammaticality effect was not influenced by the gender of the attractor. This might be unexpected given that attraction effects in subject-verb agreement have previously been observed in some untimed judgement tasks, at least in English and German (Häussler 2009; Xiang et al. 2013), and in noun-adjective agreement in Spanish (Fuchs et al. 2015; Scontras et al. 2018). One difference between our study and Xiang et al.'s study is that, in their study, participants made their judgement after a critical sentence was shown, while participants were free to make their judgement while also seeing the experimental sentences in our two untimed judgement tasks. Similarly, Fuchs et al. (2015) and Scontras et al. (2018) presented their sentences auditorily, which also makes it impossible for participants to revisit the input. The continuous availability of the sentence in our design might have made it easier for participants in our studies to notice the ungrammaticality in our critical stimuli, even when the attractor matched the gender properties of the critical adjective.

The question remains, however, as to why we observed attraction effects during reading and in speeded judgements, but not in our untimed judgement task. We suggest that these different results across tasks might indicate effects related to the time-course of gender attraction. Parker (2019) proposed an implementation of cue-based parsing that includes multiple retrievals over time. While initial retrievals are susceptible to interference, additional attempts at retrieval typically increase the probability of retrieving the target controller's representation rather than an attractor. This would predict less interference in tasks where participants have more time to provide a response, since different time constraints may yield outputs that reflect different stages in an iterative process of computation, where subsequent retrievals are triggered by the grammar until a grammatically licensed antecedent is retrieved (Parker 2019). Our results across timed and untimed experiments are in principle compatible with this proposal.[3]

3.2 REPRESENTATION VS. RETRIEVAL-BASED ACCOUNTS OF AGREEMENT ATTRACTION

The asymmetry in attraction effects that we observed is most directly compatible with the predictions of retrieval-based accounts of agreement attraction, such as cue-based parsing. We argued that the clearest evidence for representation-based accounts would be of attraction of a similar magnitude in both grammatical and ungrammatical sentences. The interactions between grammaticality and attractor that we observed were however not consistent with this prediction. Rather, the interactions that we observed, which signal a grammatical asymmetry in gender attraction effects, are more compatible with retrieval-based accounts.

---

3   Note that some studies (e.g. Nicenboim et al. 2018; Villata et al. 2018) have reported interference in offline measures in grammatical sentences. They too asked comprehension questions after the critical sentence however, with it no longer being available onscreen. Additionally, the offline measures used in these studies were content questions that tapped sentence interpretation rather than acceptability judgements. How the time-course of interference may be influenced by different offline measures that tap interpretation and acceptability is an avenue for future research. Our conclusions here are clearly restricted to measures of sentence acceptability, rather than interpretation.

While we observed facilitatory interference in ungrammatical sentences when the attractor matched the gender marking on the adjective, we did not find any significant differences, either facilitatory or inhibitory, in grammatical sentences in any of our experiments. Given concerns regarding the sample size that may be needed to observe inhibitory effects in grammatical sentences (Nicenboim et al. 2018), we are cautious in interpreting null effects here. Note that the model estimates in our combined analysis of regression path times for grammatical sentences in Experiments 1b and 2b was, though clearly not significant, in the direction of inhibitory interference, rather than the facilitatory effect that would be predicted by representation-based accounts. The numerical direction here may be consistent with the claim that inhibitory effects in agreement are small and difficult to detect without large sample sizes (Nicenboim et al. 2018).

Finally, we note that Hammerly et al. (2019) have recently argued that the grammatical asymmetry observed in forced-choice judgement tasks is an artefact of response bias. They argued that in previous forced-choice studies on agreement attraction, ungrammatical sentences were typically responded to less accurately than grammatical sentences. Once this response bias is accounted for, the grammatical asymmetry disappears, and a pattern of results emerges that is more consistent with representation rather than retrieval-based accounts of agreement attraction. While our study was not designed to test Hammerly et al.'s hypothesis, the results of our speeded judgement task (Experiment 3) provided no clear evidence of a response bias between grammatical and ungrammatical conditions. Indeed, accuracy was equally high overall for grammatical and ungrammatical sentences. Yet, we still observed an asymmetry in attraction effects across grammaticality conditions.

We should note, however, that Hammerly et al. predict a perfectly unbiased responder to show 5% more errors as a result of attraction in ungrammatical as compared to grammatical sentences. Numerically, the attraction effect in ungrammatical sentences in our Experiment 3 is of this magnitude, while the difference between grammatical sentences is very close to 0. In this sense, our results might be compatible with Hammerly et al.'s proposals. Hammerly et al. themselves also queried whether response bias can account for more implicit measures, such as eye-movements, where it is not clear whether an explicit judgement is made. Since we also observed results consistent with a grammatical asymmetry in our eye-movement data, and while Hammerly et al.'s proposal requires further investigation, we maintain that our results are most consistent with the predictions of retrieval-based accounts of agreement attraction, such as cue-based parsing.

Related to Hammerly et al.'s response bias proposal, one important difference between our study and the results of Acuña-Fariña et al. (2014) was that we tested both grammatical and ungrammatical stimuli, whereas Acuña-Fariña et al. tested only grammatical sentences. It might be that the relative amount of grammatical and ungrammatical sentences throughout an experiment influences attraction effects in grammatical sentences, even in implicit measures like eye-movements. In other words, it is possible that the absence of ungrammatical sentences in the experimental set effectively removes a tacit response bias introduced by ungrammatical stimuli in designs manipulating grammaticality in their critical stimuli. If true, this might explain the different results observed in grammatical sentences between our study and Acuña-Fariña et al. (2014) and would be compatible with Hammerly et al.'s (2019) account. However, there are reasons to exercise caution in entertaining this possibility. Other studies on subject-verb agreement that did not include ungrammatical sentences have also reported inhibitory, rather than facilitatory, interference in grammatical sentences, which is more in line with cue-based parsing (e.g., Nicenboim et al. 2018). Further research is required to examine the issue of how different tasks, and the relative ratio of grammatical to ungrammatical stimuli across an experiment, may modulate attraction effects.

4 CONCLUSION

We investigated agreement attraction in gender agreement between a noun and a predicative adjective during comprehension in Spanish. We observed attraction in ungrammatical sentences only during online sentence reading and in a speeded judgement task. These attraction effects were observed for nouns marked with either biological or lexical gender. No attraction was observed in an untimed judgement task, however. We take these results to

indicate that attraction effects in comprehension arise due to retrieval interference, rather than resulting from a faulty or ambiguous encoding of gender features in memory, and interpret our findings as being compatible with the predictions of cue-based parsing.

ABBREVIATIONS

M = masculine, F = feminine, PL = plural, SG = singular

ADDITIONAL FILES

The additional files for this article can be found as follows:

- **Appendix A.** Experimental items from Experiments 1a and 1b. DOI: *https://doi.org/10.5334/gjgl.1300.s1*

- **Appendix B.** Experimental items from Experiments 2a and 2b. DOI: *https://doi.org/10.5334/gjgl.1300.s2*

ACKNOWLEDGEMENTS

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR AFFILIATIONS

Jorge González Alonso *orcid.org/0000-0001-5047-3226*
UiT The Arctic University of Norway, NO

Ian Cunnings *orcid.org/0000-0002-5318-0186*
University of Reading, GB

Hiroki Fujita *orcid.org/0000-0001-7649-9707*
University of Reading, GB

David Miller *orcid.org/0000-0002-1976-4487*
University of Illinois at Chicago, US

Jason Rothman *orcid.org/0000-0002-1895-561X*
UiT The Arctic University of Norway, NO;
Universidad Antonio de Nebrija, ES

REFERENCES

Acuña-Fariña, J. Carlos, Enrique Meseguer & Manuel Carreiras. 2014. Gender and number agreement in comprehension in Spanish. *Lingua* 143. 108–128. DOI: *https://doi.org/10.1016/j.lingua.2014.01.013*

Anton-Mendez, Ines, Janet L. Nicol & Merrill F. Garrett. 2002. The relation between gender and number agreement processing. *Syntax* 5(1). 1–25. DOI: *https://doi.org/10.1111/1467-9612.00045*

Baayen, R. Harald, Doug J. Davidson & Douglas M. Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59(4). 390–412. DOI: *https://doi.org/10.1016/j.jml.2007.12.005*

Badecker, William & Frantisek Kuminiak. 2007. Morphology, agreement and working memory retrieval in sentence production: Evidence from gender and case in Slovak. *Journal of Memory and Language* 56(1). 65–85. DOI: *https://doi.org/10.1016/j.jml.2006.08.004*

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

Bock, Kathryn. 2003. Psycholinguistically speaking: some matters of meaning, marking, and morphing. In *Psychology of Learning and Motivation* 44. 109–144. Elsevier. DOI: *https://doi.org/10.1016/S0079-7421(03)44004-8*

Bock, Kathryn & Carol A. Miller. 1991. Broken agreement. *Cognitive Psychology* 23(1). 45–93. DOI: *https://doi.org/10.1016/0010-0285(91)90003-7*

Bock, Kathryn & Erica L. Middleton. 2011. Reaching agreement. *Natural Language & Linguistic Theory* 29(4). 1033–1069. DOI: *https://doi.org/10.1007/s11049-011-9148-y*

Bock, Kathryn & J. Cooper Cutting. 1992. Regulating mental energy: performance units in language production. *Journal of Memory and Language* 31(1). 99–127. DOI: *https://doi.org/10.1016/0749-596X(92)90007-K*

Cunnings, Ian & Patrick Sturt. 2018. Retrieval interference and semantic interpretation. *Journal of Memory and Language* 102. 16–27. DOI: *https://doi.org/10.1016/j.jml.2018.05.001*

Dillon, Brian, Alan Mishler, Shayne Sloggett & Colin Phillips. 2013. Contrasting intrusion profiles for agreement and anaphora: Experimental and modeling evidence. *Journal of Memory and Language* 69(2). 85–103. DOI: *https://doi.org/10.1016/j.jml.2013.04.003*

Eberhard, Kathleen M. 1997. The marked effect of number on subject–verb agreement. *Journal of Memory and Language* 36(2). 147–164. DOI: *https://doi.org/10.1006/jmla.1996.2484*

Eberhard, Kathleen M., J. Cooper Cutting & Kathryn Bock. 2005. Making syntax of sense: number agreement in sentence production. *Psychological Review* 112(3). 531–559. DOI: *https://doi.org/10.1037/0033-295X.112.3.531*

Enochson, Kelly & Jennifer Culbertson. 2015. Collecting psycholinguistic response time data using Amazon Mechanical Turk. *PLOS ONE* 10(3). e0116946. DOI: *https://doi.org/10.1371/journal.pone.0116946*

Francis, W. Nelson. 1986. Proximity concord in English. *Journal of English Linguistics* 19(2). 309–317. DOI: *https://doi.org/10.1177/007542428601900212*

Franck, Julie, Gabriella Vigliocco & Janet Nicol. 2002. Subject-verb agreement errors in French and English: the role of syntactic hierarchy. *Language and Cognitive Processes* 17(4). 371–404. DOI: *https://doi.org/10.1080/01690960143000254*

Franck, Julie, Gabriella Vigliocco, Inés Antón-Méndez, Simona Collina & Ulrich H. Frauenfelder. 2008. The interplay of syntax and form in sentence production: a cross-linguistic study of form effects on agreement. *Language and Cognitive Processes* 23(3). 329–374. DOI: *https://doi.org/10.1080/01690960701467993*

Fuchs, Zuzanna, Maria Polinsky & Gregory Scontras. 2015. The differential representation of number and gender in Spanish. *The Linguistic Review* 32(4). DOI: *https://doi.org/10.1515/tlr-2015-0008*

Hammerly, Christopher, Adrian Staub & Brian Dillon. 2019. The grammaticality asymmetry in agreement attraction reflects response bias: experimental and modeling evidence. *Cognitive Psychology* 110. 70–104. DOI: *https://doi.org/10.1016/j.cogpsych.2019.01.001*

Häussler, Jana. 2009. *The emergence of attraction errors during sentence comprehension*. University of Konstanz dissertation.

Jaeger, T. Florian. 2008. Categorical data analysis: away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language* 59(4). 434–446. DOI: *https://doi.org/10.1016/j.jml.2007.11.007*

Jäger, Lena A., Daniela Mertzen, Julie A. Van Dyke & Shravan Vasishth. 2020. Interference patterns in subject-verb agreement and reflexives revisited: a large-sample study. *Journal of Memory and Language* 111. 104063. DOI: *https://doi.org/10.1016/j.jml.2019.104063*

Jäger, Lena A., Felix Engelmann & Shravan Vasishth. 2017. Similarity-based interference in sentence comprehension: literature review and Bayesian meta-analysis. *Journal of Memory and Language* 94. 316–339. DOI: *https://doi.org/10.1016/j.jml.2017.01.004*

Kaan, Edith. 2002. Investigating the effects of distance and number interference in processing subject-verb dependencies: an ERP study. *Journal of Psycholinguistic Research* 31(2). 165–193. DOI: *https://doi.org/10.1023/A:1014978917769*

Kimball, John & Judith Aissen. 1971. I think, you think, he think. *Linguistic Inquiry* 2(2). 241–246.

Lago, Sol, Diego E. Shalom, Mariano Sigman, Ellen F. Lau & Colin Phillips. 2015. Agreement attraction in Spanish comprehension. *Journal of Memory and Language* 82. 133–149. DOI: *https://doi.org/10.1016/j.jml.2015.02.002*

Lago, Sol, Martina Gračanin-Yuksek, Duygu Fatma Şafak, Orhan Demir, Bilal Kırkıcı & Claudia Felser. 2019. Straight from the horse's mouth: agreement attraction effects with Turkish possessors. *Linguistic Approaches to Bilingualism* 9(3). 398–426. DOI: *https://doi.org/10.1075/lab.17019.lag*

Lewis, Richard L. & Shravan Vasishth. 2005. An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science* 29(3). 375–419. DOI: *https://doi.org/10.1207/s15516709cog0000_25*

Lewis, Richard L., Shravan Vasishth & Julie A. Van Dyke. 2006. Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences* 10(10). 447–454. DOI: *https://doi.org/10.1016/j.tics.2006.08.007*

Lorimor, Heidi, Kathryn Bock, Ekaterina Zalkind, Alina Sheyman & Robert Beard. 2008. Agreement and attraction in Russian. *Language and Cognitive Processes* 23(6). 769–799. DOI: *https://doi.org/10.1080/01690960701774182*

Luke, Steven G. 2017. Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods* 49(4). 1494–1502. DOI: *https://doi.org/10.3758/s13428-016-0809-y*

Martin, Andrea E., Mante S. Nieuwland & Manuel Carreiras. 2012. Event-related brain potentials index cue-based retrieval interference during sentence comprehension. *NeuroImage* 59(2). 1859–1869. DOI: *https://doi.org/10.1016/j.neuroimage.2011.08.057*

Martin, Andrea E., Mante S. Nieuwland & Manuel Carreiras. 2014. Agreement attraction during comprehension of grammatical sentences: ERP evidence from ellipsis. *Brain and Language* 135. 42–51. DOI: *https://doi.org/10.1016/j.bandl.2014.05.001*

McElree, Brian. 2000. Sentence comprehension is mediated by content-addressable memory structures. *Journal of Psycholinguistic Research* 29(2). 111–123. DOI: *https://doi.org/10.1023/A:1005184709695*

McElree, Brian, Stephani Foraker & Lisbeth Dyer. 2003. Memory structures that subserve sentence comprehension. *Journal of Memory and Language* 48(1). 67–91. DOI: *https://doi.org/10.1016/S0749-596X(02)00515-6*

Nicenboim, Bruno, Shravan Vasishth, Felix Engelmann & Katja Suckow. 2018. Exploratory and confirmatory analyses in sentence processing: a case study of number interference in German. *Cognitive Science* 42. 1075–1100. DOI: *https://doi.org/10.1111/cogs.12589*

Nicol, Janet L., Kenneth I. Forster & Csaba Veres. 1997. Subject–verb agreement processes in comprehension. *Journal of Memory and Language* 36(4). 569–587. DOI: *https://doi.org/10.1006/jmla.1996.2497*

Osterhout, Lee & Linda A. Mobley. 1995. Event-related brain potentials elicited by failure to agree. *Journal of Memory and Language* 34(6). 739–773. DOI: *https://doi.org/10.1006/jmla.1995.1033*

Parker, Dan. 2019. Two minds are not always better than one: modeling evidence for a single sentence analyzer. *Glossa: A Journal of General Linguistics* 4(1). 64. DOI: *https://doi.org/10.5334/gjgl.766*

Parker, Dan & Colin Phillips. 2017. Reflexive attraction in comprehension is selective. *Journal of Memory and Language* 94. 272–290. DOI: *https://doi.org/10.1016/j.jml.2017.01.002*

Parker, Dan, Michael Shvartsman & Julie A. Van Dyke. 2017. The cue-based based retrieval theory of sentence comprehension: new findings and new challenges. In Linda Escobar, Vicenç Torrens, & Teresa Parodi (eds.), *Language Processing and Disorders*, 121–144. Newcastle: Cambridge Scholars Publishing.

Paspali, Anastasia & Theodoros Marinis. 2020. Gender agreement attraction in Greek comprehension. *Frontiers in Psychology* 11. 717. DOI: *https://doi.org/10.3389/fpsyg.2020.00717*

Pearlmutter, Neal J., Susan M. Garnsey & Kathryn Bock. 1999. Agreement processes in sentence comprehension. *Journal of Memory and Language* 41(3). 427–456. DOI: *https://doi.org/10.1006/jmla.1999.2653*

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech & Jan Svartvik. 1972. *A grammar of contemporary English*. London: Longman.

Schlueter, Zoe, Alexander Williams & Ellen Lau. 2018. Exploring the abstractness of number retrieval cues in the computation of subject-verb agreement in comprehension. *Journal of Memory and Language* 99. 74–89. DOI: *https://doi.org/10.1016/j.jml.2017.10.002*

Scontras, Gregory, Maria Polinsky & Zuzanna Fuchs. 2018. In support of representational economy: agreement in heritage Spanish. *Glossa: A Journal of General Linguistics* 3(1). 1. DOI: *https://doi.org/10.5334/gjgl.164*

Shen, Erica Y., Adrian Staub & Lisa D. Sanders. 2013. Event-related brain potential evidence that local nouns affect subject–verb agreement processing. *Language and Cognitive Processes* 28(4). 498–524. DOI: *https://doi.org/10.1080/01690965.2011.650900*

Slioussar, Natalia & Anton Malko. 2016. Gender agreement attraction in Russian: production and comprehension evidence. *Frontiers in Psychology* 7. DOI: *https://doi.org/10.3389/fpsyg.2016.01651*

Tanner, Darren, Janet Nicol & Laurel Brehm. 2014. The time-course of feature interference in agreement comprehension: multiple mechanisms and asymmetrical attraction. *Journal of Memory and Language* 76. 195–215. DOI: *https://doi.org/10.1016/j.jml.2014.07.003*

Tucker, Matthew A., Ali Idrissi & Diogo Almeida. 2015. Representing number in the real-time processing of agreement: self-paced reading evidence from Arabic. *Frontiers in Psychology* 6. DOI: *https://doi.org/10.3389/fpsyg.2015.00347*

Van Dyke, Julie A. 2007. Interference effects from grammatically unavailable constituents during sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 33(2). 407–430. DOI: *https://doi.org/10.1037/0278-7393.33.2.407*

Van Dyke, Julie A. & Brian McElree. 2011. Cue-dependent interference in comprehension. *Journal of Memory and Language* 65(3). 247–263. DOI: *https://doi.org/10.1016/j.jml.2011.05.002*

Vasishth, Shravan & Bruno Nicenboim. 2016. Statistical methods for linguistic research: foundational ideas – Part I: statistical methods for linguistics. *Language and Linguistics Compass* 10(8). 349–369. DOI: *https://doi.org/10.1111/lnc3.12201*

Vigliocco, G., B. Butterworth & C. Semenza. 1995. Constructing subject-verb agreement in speech: the role of semantic and morphological factors. *Journal of Memory and Language* 34(2). 186–215. DOI: *https://doi.org/10.1006/jmla.1995.1009*

Vigliocco, Gabriella, Brian Butterworth & Merrill F. Garrett. 1996. Subject-verb agreement in Spanish and English: differences in the role of conceptual constraints. *Cognition* 61(3). 261–298. DOI: *https://doi.org/10.1016/S0010-0277(96)00713-5*

Vigliocco, Gabriella & Julie Franck. 1999. When sex and syntax go hand in hand: gender agreement in language production. *Journal of Memory and Language* 40(4). 455–478. DOI: *https://doi.org/10.1006/jmla.1998.2624*

Vigliocco, Gabriella & Tiziana Zilli. 1999. Syntactic accuracy in sentence production: the case of gender disagreement in Italian language-impaired and unimpaired speakers. *Journal of Psycholinguistic Research* 28(6). 623–648. DOI: *https://doi.org/10.1023/A:1023268911312*

Villata, Sandra & Julie Franck. 2020. Similarity-based interference in agreement comprehension and production: evidence from object agreement. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 46(1). 170–188. DOI: *https://doi.org/10.1037/xlm0000718*

Villata, Sandra, Whitney Tabor & Julie Franck. 2018. Encoding and retrieval interference in sentence comprehension: evidence from agreement. *Frontiers in Psychology* 9. 2. DOI: *https://doi.org/10.3389/fpsyg.2018.00002*

Wagers, Matthew W., Ellen F. Lau & Colin Phillips. 2009. Agreement attraction in comprehension: representations and processes. *Journal of Memory and Language* 61(2). 206–237. DOI: *https://doi.org/10.1016/j.jml.2009.04.002*

Xiang, Ming, Brian Dillon & Colin Phillips. 2009. Illusory licensing effects across dependency types: ERP evidence. *Brain and Language* 108(1). 40–55. DOI: *https://doi.org/10.1016/j.bandl.2008.10.002*

Xiang, Ming, Julian Grove & Anastasia Giannakidou. 2013. Dependency-dependent interference: NPI interference, agreement attraction, and global pragmatic inferences. *Frontiers in Psychology* 4. DOI: *https://doi.org/10.3389/fpsyg.2013.00708*