

Does productive agreement morphology increase sensitivity to agreement in a second language?

Sol Lago, Goethe University Frankfurt, sollago@em.uni-frankfurt.de

Elise Oltrogge, Goethe University Frankfurt, oltrogge@em.uni-frankfurt.de

Kate Stone, University of Hull, k.stone@hull.ac.uk

Adult language learners have variable performance with subject-verb number agreement. But it is unclear whether their performance additionally depends on the availability of agreement morphology in their first language. To address this question, we conducted a self-paced reading task comparing different speaker groups: (a) first vs. second language speakers of German; (b) intermediate-to-advanced German learners whose first language had more or less productive number agreement morphology (Spanish vs. English). Two manipulations were used to diagnose number processing: agreement violations and agreement attraction. Our results showed decreased sensitivity to agreement violations in language learners, irrespective of the morphological productivity of their first language. Meanwhile, differences in attraction effects were inconclusive in all between-group comparisons. We suggest that second language variability with subject-verb agreement is unlikely to result from increased retrieval interference – the effect underlying attraction. Instead, variable performance more likely arises because learners have difficulties in the real-time mapping of inflectional morphemes to syntactic features.



1. Introduction

Speakers who acquired a language in early childhood – often called *native* or *L1* speakers – compute subject-verb number agreement quickly and accurately during sentence processing (De Vincenzi et al., 2003; Molinaro et al., 2011; Pearlmutter et al., 1999). By contrast, speakers who acquired a language later in life – called *second language* or *L2* speakers – show more variable performance with agreement, even at high proficiency levels (Blom et al., 2006; Johnson & Newport, 1989; Lardiere, 1998; Prévost & White, 2000). At first glance, this contrast could point to a difference in the mechanisms underlying L1–L2 agreement processing. But processing might also depend on speakers' first language. This includes not only whether the first language has agreement, but also the extent to which it is morphologically realized: Some languages have impoverished number agreement morphology (e.g., English), while others have productive and functionally relevant agreement morphology (e.g., Spanish and German). Such differences might shape L2 processing mechanisms, such that L2 speakers might rely on morphological cues differentially, depending on their productivity in their native language (Jiang et al., 2011; MacWhinney, 1992; Sagarra & Ellis, 2013; Shibuya & Wakabayashi, 2008).

For example, according to the *Morphological Congruency Hypothesis*, morphological processing can only become native-like if a morpheme exists and maps to the same meaning in speakers' L1 and L2 (Jiang et al., 2011, 2017). The Morphological Congruency Hypothesis was originally supported by findings from longitudinal studies, which tracked the development of different morphemes – for example, nominal plural and past tense morphology – in immersed learners of English whose L1 lacked the morphemes under study (e.g., Chinese and Japanese). These learners often produced morpheme errors and showed little progress over the years (Jia & Aaronson, 2003; Lardiere, 1998; Long, 2003).

More recently, the Morphological Congruency Hypothesis has been tested with real-time processing evidence (Choi & Ionin, 2021; Guillelmon & Grosjean, 2001; Ionin et al., 2021; Jiang et al., 2011; Scherag et al., 2004). For example, in a self-paced reading study in English, proficient learners with L1 Russian – a language with nominal plural morphology – showed morphological sensitivity through reading disruptions after a non-inflected noun (*several of the board members/*member*), but learners with L1 Japanese – a language without obligatory plural marking – did not (Jiang et al., 2011).

A possible extension of the Morphological Congruency Hypothesis concerns whether the productivity of an L1 morpheme – rather than its sole existence – also affects learners' real-time L2 processing. Some results support this idea in the area of grammatical tense. Sagarra and Ellis (2013) conducted an eye-tracking study with sentences containing adverb-verb violations in Spanish (e.g., *Ayer el chico cocinó/*cocina*, 'Yesterday the boy cook._{PAST}/*cook._{PRESENT}'). Spanish learners with L1 Romanian (productive tense morphology) looked longer at the infelicitous verbs than learners with L1 English (poor tense morphology). This suggests that learners processed L2

tense morphology to a greater extent if their L1 had morphologically richer tense. The current study investigates whether this kind of morphological influence extends to the processing of subject-verb number agreement by comparing English and Spanish intermediate-to-advanced learners of German.

1.1 Morphological productivity and agreement processing

Questions about the role of morphological productivity in agreement processing have received little attention. However, some information can be gleaned from studies on a more widely researched question: Whether speakers whose L1 doesn't have a specific grammatical feature – like number or gender – have difficulties using it for agreement in their L2. Answering this question is important for processing theories: While some accounts propose that the absence of a morphological feature in an L1 hinders its L2 processing (Hopp, 2022; MacWhinney, 2005) or prevents its target-like representation as a functional feature in the L2 grammar (Franceschina, 2005; Hawkins & Franceschina, 2004), other accounts don't assume L1 influence and focus more on L2 speakers' reliance on other types of knowledge (e.g., lexical and semantic) than on morphosyntactic knowledge during processing (Clahsen & Felser, 2018).

Processing evidence provides a nuanced answer to the question of whether L1–L2 similarity affects agreement processing in proficient L2 speakers. This evidence comes mostly from two sources: agreement violation paradigms used in event-related potential (ERP) studies, and predictive paradigms used in visual world eye-tracking studies. These previous findings often conflict. Some studies have reported that L1–L2 similarity enhances L2 speakers' sensitivity to agreement violations (Foucart & Frenck-Mestre, 2011; Gillon Dowens et al., 2011; Sabourin & Stowe, 2008) and the predictive use of agreement (Hopp & Lemmerth, 2018). However, other studies have failed to find an effect of L1–L2 similarity on either agreement violations (Alemán Bañón et al., 2018; White et al., 2012) or agreement-based predictions (Dussias et al., 2013; Lago et al., 2023). In fact, a meta-analysis of 41 ERP studies on syntactic processing – including number and gender agreement – found that L1–L2 similarity did not significantly predict the presence or absence of any ERP component, in contrast with factors like L2 proficiency and length of immersion (Caffarra et al., 2015).

However, most previous studies used paradigms that involved explicit knowledge, such as error-detection tasks. Thus, it is unclear whether participants' responses reflected processing mechanisms vs. task-specific strategies (Jiang, 2007). For instance, in violation-based ERP paradigms, participants were often asked to read the sentences with the goal of making grammaticality judgments (Alemán Bañón et al., 2018; Gillon Dowens et al., 2010; Sabourin & Stowe, 2008). Thus, the recorded brain responses might reflect not only participants' sensitivity to agreement, but also specific heuristics aimed at solving the judgment task. In this regard, evidence provided by studies on morphosyntactic predictions might be more reflective of implicit

processing, as participants were not asked to judge sentences, but rather to click on images on a computer screen (Dussias et al., 2013; Hopp & Lemmerth, 2018; Lago et al., 2023). However, there is growing evidence that – likely due to the simple and repetitive nature of the task – participants in visual world studies can become aware that agreement is the object of study, and that they use it strategically to perform the task (Curcic et al., 2019; Koch et al., 2021).

To complement previous studies, it is important to use more implicit tasks to diagnose the processing mechanisms underlying L2 subject-verb number agreement. The next section describes a psycholinguistic effect that fulfills this criterion: agreement attraction. Our study measured agreement attraction using a reading-for-comprehension paradigm without any judgment component. Our goal was to assess whether L1 morphological differences modulated subject-verb agreement processing in a second language.

1.2 Attraction in first language comprehension

Agreement attraction is a useful diagnostic of L2 agreement processing, because its underlying cognitive mechanisms are well-studied in L1 speakers. Attraction research on subject-verb number agreement in comprehension typically uses a 2×2 design manipulating the grammaticality of a sentence, i.e., whether the critical verb disagrees in number with the noun heading the subject – *the key to the cabinet(s)* – and whether the sentence contains a noun that mismatches the number of the subject head, typically called an *attractor* – bolded in example (1) below (Bock & Miller, 1991; Lago et al., 2021).

(1) *Grammatical agreement (+/- MATCH)*

- a. The key to the cabinet was rusty from many years of disuse.
- b. The key to the **cabinets** was rusty from many years of disuse.

Ungrammatical agreement (+/- MATCH)

- c. *The key to the cabinet were rusty from many years of disuse.
- d. *The key to the **cabinets** were rusty from many years of disuse.

Previous studies have found that L1 speakers show processing disruptions after encountering a verb with an agreement violation in (1c–d) vs. (1a–b), which results in longer reading times and increased brain responses. Critically, ungrammatical verbs elicit smaller disruptions when the attractor noun matches the verb in number. This yields faster reading times and/or reduced brain responses – e.g., in (1d) vs. (1c) – resulting in a so-called *agreement attraction effect* (Jäger et al., 2020; Lago et al., 2021; Nicol et al., 1997; Pearlmutter et al., 1999; Shen et al., 2013; Tanner et al., 2014; Wagers et al., 2009).

A widely adopted interpretation of attraction effects is that they reflect facilitatory interference due to the partial match of the attractor with the lexical features of the verb, which are used as retrieval cues by the parser (Jäger et al., 2017). When a verb is encountered, speakers use

its syntactic, semantic, and morphological features as cues to retrieve an appropriate number controller from memory. Memory chunks corresponding to preceding items are queried in parallel, and the chunk that best matches the retrieval cues is the most likely to be selected. Sentences with an attractor like (1d) elicit interference during retrieval due to partial match: While the subject head matches the syntactic cues of the verb, the plural attractor matches the number cue. Under activation-based retrieval models, this partial match increases the probability of misretrieving the attractor (Dillon et al., 2013; Engelmann et al., 2019; Nicenboim & Vasishth, 2018). Since the attractor in (1d) matches more retrieval cues than in (1c), it is also retrieved faster, which leads to a faster average retrieval time. This retrieval speedup is assumed to generate the faster processing times in the attraction condition (1d), as compared to (1c).

Attraction effects are robustly observed in ungrammatical sentences, while their existence in grammatical sentences – (1b) vs. (1a) – is unclear. In a survey of previous studies, Hammerly and colleagues (2019) reported that statistically significant number attraction effects were detected more often in ungrammatical than grammatical sentences: 91% vs. 38% of the surveyed studies. Further, work using Bayesian meta-analyses demonstrated a consistent attraction effect in ungrammatical sentences – on average, –22 ms, with a [–36, –9] ms 95% credible interval – but a negligible and task-dependent effect in grammatical sentences – on average, –7 ms, with a [–16, 4] ms 95% credible interval (Jäger et al., 2020). Since more research addressing the sources of this variation is needed, our study diagnosed attraction effects only in ungrammatical sentences, in which the existence of attraction effects is uncontroversial.

1.3 Attraction in second language comprehension

Proficient L2 speakers typically show number attraction effects in comprehension, which suggests that they also use a retrieval mechanism for subject-verb agreement (Bian et al., 2021; Cheng et al., 2021; Hoshino et al., 2010; Jegerski, 2016; Lee & Phillips, 2023; Lim & Christianson, 2015; Tanner et al., 2012). But it is less clear whether attraction effects are larger in L2 speakers than in L1 speakers, with mixed results in previous studies (Jacob et al., 2017). The comparison of attraction effects in L1 vs. L2 speakers is theoretically relevant, because larger L2 attraction effects are predicted by accounts like the *Interference Hypothesis* (Cunnings, 2017a, 2017b). The Interference Hypothesis proposes that both L1 and L2 speakers rely on cue-based memory retrieval to process linguistic dependencies like subject-verb agreement. However, L2 speakers are more susceptible to retrieval interference during processing, which increases the likelihood of misretrievals and should result in larger attraction effects than those exhibited by L1 speakers. Our study evaluates this prediction by comparing L1 and L2 speakers of German.

Of additional interest to our study is whether morphological richness plays a role. Specifically, speakers with productive L1 agreement might be more susceptible to attraction if richer morphology affects retrieval through differential cue weighting. L1 Spanish speakers might

weigh morphological cues like number more heavily at retrieval – as compared to other cues, for example, syntactic cues, which should be weighted similarly by Spanish and English speakers. A higher weighting of number cues should increase interference from number-matching elements in the attraction conditions, leading to more misretrievals and, thus, larger attraction effects in Spanish speakers than in English speakers.

To our knowledge, no previous study has evaluated this possibility by comparing attraction effects between two learner groups who differ in the richness of their L1 morphology. Therefore, inferences about these variables need to be based on comparisons across studies. These reveal an unclear pattern of results. With speakers whose L1 has subject-verb agreement, studies tend to find qualitatively similar L1–L2 attraction effects, for example, in proficient Spanish speakers of English (Tanner et al., 2012), in proficient Russian speakers of German (Lago & Felser, 2018), and in English near-native speakers of Spanish (Jegerski, 2016).

With speakers whose L1 lacks subject-verb agreement, results are more variable. Studies on proficient Korean speakers of English have shown similar L1–L2 attraction profiles, but the effects depended on the grammatical environment of the attractor noun, i.e., whether it appeared in a relative clause (e.g., *The artist who made **the sculpture(s)** was/were...*) or in a prepositional phrase (Lee & Phillips, 2023; Lim & Christianson, 2015). For example, in the study by Lee and Phillips (2023), Korean learners of L2 English showed attraction effects in relative clause constructions, but there was no evidence of attraction in prepositional phrase constructions.

Meanwhile, studies on Chinese speakers of English have reported similar L1–L2 attraction profiles (Cheng et al., 2021), no L2 attraction (Jiang, 2004) or similar L1–L2 attraction profiles but with a smaller size in L2 (Bian et al., 2021) or a different distribution of brain responses (Chen et al., 2007). However, several studies on Chinese learners measured attraction in grammatical sentences (Cheng et al., 2021; Jiang, 2004), where even L1 findings are inconclusive. If only studies including ungrammatical sentences are considered, then the results suggest behaviorally similar L1–L2 attraction profiles (Chen et al., 2007; Hoshino et al., 2010; Jegerski, 2016; Tanner et al., 2012).

Crucially, previous studies focused on comparing L1–L2 speakers, rather than comparing L2 learners with different L1 backgrounds. Direct comparisons between learner groups are important, because to appropriately demonstrate L1 influence, it's not enough to show that a particular L2 behavior is *consistent* with speakers' L1. Instead, one has to show that the behavior is *specific* to speakers' L1, for example, by directly comparing L2 speakers whose L1s differ in the properties of the target construction (Felix, 1976; Jarvis, 2000; Pienemann et al., 2005). Otherwise, one might misattribute to L1 influence behaviors that are due to general-purpose processing mechanisms, which may be used universally by L2 learners. The current study addressed this gap by directly comparing German speakers with English or Spanish as a native language.

1.4 The present study

English and Spanish provide useful comparison points for questions about morphological richness. Morphological richness can be quantified in terms of paradigmatic richness, which describes the number of formally distinct inflected word-forms per lemma (Dressler, 2003; Xanthos et al., 2011). Based on this, English has impoverished number agreement morphology, such that only third person singular verbs mark number agreement, e.g., *I/you/they/we eat* vs. *she/he/it eats*. By contrast, Spanish verbs mark agreement in all syntactic persons, with clearly differentiable singular and plural verb forms (e.g., *comió* ‘ate.3SG’ vs. *comieron* ‘ate.3PL’). Further, English has a relatively fixed word order, but word order in Spanish is freer and grammatical subjects can be omitted – Spanish is a null-subject language – which makes sentence position less useful for subject identification. As a consequence, verb morphology is both productive and reliable as a cue for subject identification in Spanish, while English speakers might rely more on word order and syntactic information (Kail, 1989; MacWhinney, 2001; MacWhinney et al., 1984).

Our study compared L1 Spanish and English speakers in their processing of subject-verb number agreement in German – a language with productive agreement morphology, like Spanish. We tested intermediate-to-advanced L2 speakers, to ensure that any potential between-group contrasts were attributable to differences between processing mechanisms, rather than an incomplete knowledge of German. We also included speakers who had acquired German as a first language, in order to test the predictions of the Interference Hypothesis (Cunnings, 2017a, 2017b). Thus, two analyses were performed: the first compared the L1 vs. L2 groups – thus collapsing L2 speakers into one group. The second analysis focused on the L2 learners and compared English vs. Spanish native speakers, to assess whether differences in their native number systems affected their processing of subject-verb agreement in German.

We defined two experimental effects of interest. The first concerned *agreement violations*. This effect was diagnosed as the processing disruption elicited by a verb disagreeing with the subject in number. An example is provided by sentences (1a) vs. (1c) above: in contrast with (1a), the verb in (1c) has an agreement violation. Note that the subject and attractor noun always match in number, such that the ungrammatical verb cannot be licensed in number.

The second experimental effect concerned *agreement attraction*. This effect was diagnosed as the processing facilitation elicited by an ungrammatical verb in the presence of a plural attractor, as compared to an equally ungrammatical sentence without an attractor noun. An example is the comparison between (1c) and (1d).

Our study adopted an error-driven account of retrieval, which proposes that participants use memory retrieval after encountering a verb whose number does not match the number predicted on the basis of the number of the subject (Lago et al., 2015; Schlueter et al., 2019; Wagers et al., 2009). According to this account, readers first detect an agreement violation, and only

afterward experience attraction due to misretrieval. Thus, agreement violations and attraction effects convey complementary information. Participants' reactions to agreement violations provide a measure of their real-time ability to detect whether the number of the verb matches the number expectation generated by the subject. Meanwhile, reactions to attraction measure participants' susceptibility to retrieval interference, which allows one to establish whether a retrieval mechanism is used to compute agreement.

Our predictions were as follows. L1–L2 comparisons allowed us to address the question of whether intermediate-to-advanced L2 speakers could become native-like in the processing of subject-verb number agreement in German. If so, L1 and L2 speakers should show comparable reading disruptions after encountering a verb with an agreement violation. With regard to agreement attraction, if L2 speakers are more affected by interference, as posited by the Interference Hypothesis (Cunnings, 2017a, 2017b), then attraction should be larger in L2 speakers than L1 speakers. Alternatively, L1 and L2 speakers might be equally prone to interference and show similar attraction effects, consistent with native-like retrieval mechanisms.

Meanwhile, comparisons between English and Spanish speakers allowed us to address whether differences in L1 morphological productivity affect the processing of subject-verb number agreement in German. Specifically, due to the richness and reliability of number inflection in Spanish, Spanish speakers might be more fine-tuned to agreement morphology in German, resulting in stronger reading disruptions when an agreement violation is encountered. This outcome would support an extended version of the Morphological Congruency Hypothesis (Jiang et al., 2011). With regard to attraction, a higher weighting of morphological cues should give rise to stronger attraction in Spanish speakers than English speakers.

Alternatively, the morphological richness of speakers' L1 might not affect their processing of subject-verb agreement. Specifically, intermediate-to-advanced L2 German speakers with different morphological systems might have developed similar processing mechanisms for subject-verb agreement, regardless of their first language. If so, comparable responses to agreement violations and agreement attraction should arise in the L2 groups.

2. Self-paced reading experiment

Our experiment examined the processing of number agreement, using a self-paced reading paradigm that required participants to read for comprehension – without any judgment component. The experimental materials tested the agreement relationship between a subject and a verb within a relative clause (RC), with a non-intervening attractor noun located outside the RC. Non-intervening attractors have been used extensively in previous research (Avetisyan et al., 2020; Dillon et al., 2013; Lago et al., 2021; Staub, 2009; Wagers et al., 2009), and they offer several advantages. First, they allowed us to keep constant the region immediately preceding the verb, which is one strategy for avoiding confounding effects from the plurality of the attractor

noun (Wagers et al., 2009). Second, recent work has shown that L2 speakers may sometimes show attraction with relative clauses, but not with prepositional phrases (Lee & Phillips, 2023; Lim & Christianson, 2015). Because we wanted to maximize our probability of eliciting attraction in L2 speakers, we used relative clauses.

A sample item set is illustrated in (2). Each experimental item consisted of four conditions created by crossing two factors. The first factor was the grammaticality of the sentence. It was manipulated by varying the number of the RC subject: it was plural in the grammatical conditions (e.g., *die Diplomaten*._{PL} in (2a–b)), such that it agreed with the RC plural verb (e.g., *unterzeichneten*). By contrast, the RC subject was singular in the ungrammatical conditions, such that it disagreed with the RC verb (e.g., *der Diplomat*._{SG}” in (2c–d)). This design allowed us to hold the RC verb constant across conditions, such that reading times at the verb could not be affected by differences in the verb’s length or frequency across conditions. Following Lago et al. (2015), the attractor noun was always inanimate, and the agreement controller noun was animate. This configuration has been shown to facilitate the comprehension of relative clause constructions (Mak et al., 2002), which helped ensure that L2 participants could process the target sentences.

The second factor was whether the attractor and the RC subject matched or mismatched in number ((2a,c) vs. (2b,d)). In the match conditions, both noun phrases had the same number, while in the mismatch conditions, they had different numbers. Because agreement attraction effects are typically found in ungrammatical sentences in mismatch configurations, the comparison between (2c) and (2d) was our key comparison for diagnosing attraction.

The critical region was the word immediately after the verb, the *verb + 1* region (underlined below), which is where the previous self-paced reading studies have typically found attraction effects (Avetisyan et al., 2020; Lago et al., 2021; Patson & Husband, 2016; Wagers et al., 2009). A supplementary analysis of the verb region – which was the earliest region in which an effect could arise – can be found in Supplemental Materials S5.

(2) a. GRAMMATICAL, MATCH

Die Verträge, die die Diplomaten feierlich unterzeichneten
 The treaties_{.NOM.PL} that_{.ACC.PL} the diplomats_{.NOM.PL} solemnly signed_{.3PL}
bei der Konferenz, hatten viele Seiten.
 at the conference_{.NOM} had_{.3PL} many pages_{.ACC}
 ‘The treaties that the diplomats solemnly signed at the conference had many pages.’

b. GRAMMATICAL, MISMATCH

Der Vertrag, den die Diplomaten feierlich unterzeichneten
 The treaty_{.NOM.SG} that_{.ACC.SG} the diplomats_{.NOM.PL} solemnly signed_{.3PL}
bei der Konferenz, hatte viele Seiten.
 at the conference_{.NOM} had_{.3SG} many pages_{.ACC}
 ‘The treaty that the diplomats solemnly signed at the conference had many pages.’

c. UNGRAMMATICAL, MATCH

*Der Vertrag, den der Diplomat feierlich unterzeichneten
 The treaty_{.NOM.SG} that_{.ACC.SG} the diplomat_{.NOM.SG} solemnly signed_{.3PL}
 bei der Konferenz, hatte viele Seiten.
 at the conference_{.NOM} had_{.3SG} many pages_{.ACC}
 ‘The treaty that the diplomat solemnly signed at the conference had many pages.’

d. UNGRAMMATICAL, MISMATCH

*Die Verträge, die der Diplomat feierlich unterzeichneten
 The treaties_{.NOM.PL} that_{.ACC.PL} the diplomat_{.NOM.SG} solemnly signed_{.3PL}
 bei der Konferenz, hatten viele Seiten.
 at the conference_{.NOM} had_{.3PL} many pages_{.ACC}
 ‘The treaties that the diplomat solemnly signed at the conference had many pages.’

QUESTION: Wurde etwas unterzeichnet während der Konferenz? (‘Was anything signed during the conference?’).

TARGET ANSWER: Ja (‘yes’)

2.1 Method

2.1.1 Participants

Participants consisted of 127 speakers who had been exposed to German from birth (L1 group) and 132 speakers who had started learning German after childhood, outside their home environment (L2 group). Of the L2 speakers, 69 had English as their first language, and 63 had Spanish as their first language. They were all residing in Germany at the time of testing and reported having learned German in a combination of formal and informal contexts (e.g., immersion combined with language classes). We excluded data from participants who reported language impairments ($n = 1$), exposure to languages other than their first language at home during childhood ($n = 3$), and participants outside the 18–60 age recruitment criteria ($n = 2$).¹ We also excluded L2 participants who did not rate themselves as having at least an intermediate level of German ($n = 3$) or who did not perform with at least 50% accuracy in the untimed test on agreement knowledge (none). The two L2 groups were matched in their age of acquisition and self-rated proficiency in German (Table 1). After exclusions, 259 participants were entered in the analysis: 124 L1 German speakers (mean age = 29 years, 52 female, 113 right-handed), 67 L1 English speakers (mean age = 30 years; 39 female; 60 right-handed) and 58 L1 Spanish speakers (mean age = 34 years; 47 female; 57 right-handed).

¹ One L1 participant was additionally excluded due to scoring less than 75% in a grammatical test used to assess German knowledge in self-reported native speakers of German.

Table 1: Demographic and linguistic profile of the L2 group.

Variables	L1 English	L1 Spanish
Age of German acquisition (years) ^a	19 (10–32)	23 (5–44)
Self-rated German proficiency (%) ^b	73 (10)	75 (11)
Untimed agreement test accuracy(%) ^b	98 (5)	96 (6)

^a Mean age (age range). ^b Mean score (standard deviation).

Self-ratings were collected for each of the four skills (speaking, listening, reading and writing), using a scale from 1–10. The ratings were averaged to get an overall measure of language proficiency per participant. We used self-ratings, because previous research has shown that they offer a good correspondence with formal language tests (Blanche & Merino, 1989; Marian et al., 2007; Ross, 1998), also in L2 populations similar to ours (Lago et al., 2023). Due to the length of the experiment, it was not feasible to use a formal test to provide an additional measure of participants' proficiency. This is a limitation of our study.

2.1.2 Materials

The materials consisted of 24 experimental items and 56 filler items. The experimental sentences consisted of a main clause and an embedded object relative clause (RC). The subject of the main clause was the attractor noun phrase (e.g., 'the treaties'), which was followed by an RC including the relative pronoun 'that' (e.g., *den/die* in (2)), an animate RC subject (e.g., 'the diplomats'), an adverbial, a third person plural verb (e.g., 'signed') and a two-to-three-word prepositional phrase (e.g., 'at the conference'). This prepositional phrase acted as a spillover region, and its first word was the critical region (verb + 1 region). After the RC, the main clause was completed with a simple sentence continuation (e.g., 'had many pages'). The RC subject (e.g., 'the diplomats') always had masculine grammatical gender, to ensure unambiguous plural morphology in the preceding determiner – in German, nominative determiner forms differ from singular to plural for masculine nouns, but not for feminine nouns.

The remaining 56 items consisted of 24 sentences from another experiment (not reported here) and 32 filler sentences. The fillers consisted of simple subject-verb-object-adverbial sentences like 'The couple took their dog for a long walk in the park'. Some contained an embedded relative clause, so that they resembled the structure of the experimental items, but only subject relative clauses were used, which differed from the experimental sentences. All the filler sentences were grammatical. This resulted in 15% of trials being ungrammatical across the experiment. The rationale for having a majority of grammatical sentences was to prevent participants from getting habituated to ungrammatical stimuli, in order to maximize the likelihood of processing

disruptions to ungrammatical sentences. This followed previous self-paced and eye-tracking research on attraction, which typically keeps the percentage of ungrammatical sentences below 20–25% (Avetisyan et al., 2020; Lago et al., 2015, 2021; Wagers et al., 2009).

In addition to the self-paced reading task, L2 participants completed two untimed tests. The first test probed for their knowledge of subject-verb agreement in German (Supplemental Materials S1). It comprised ten sentences that contained a gap where the verb was located, with two response options to fill the gap (e.g., ‘Sometimes the kids ___ in the garden’; response options: ‘play’ and ‘plays’). The second test was a vocabulary test in which participants were asked to write down the plural form of the attractor nouns used in the self-paced reading task (Supplemental Materials S2).

2.1.3 Procedure

L1 participants were recruited over the platform Prolific (<http://www.prolific.com/>) and tested online using Ibex Farm (Drummond, 2013). L2 participants were also tested using Ibex Farm, but in a lab setting. Materials were presented in a word-by-word self-paced reading paradigm. At the beginning of each sentence, an asterisk was visible which indicated the location of the first word. Participants pressed the spacebar to reveal each word. When a word was revealed, the preceding one was covered. After each sentence, a comprehension question appeared that could be answered with “yes” by pressing the F key or with “no” by pressing the J key. Participants were asked to answer as fast and as accurately as possible. Half of the questions had “yes” as a target answer.

Experimental items and fillers were presented in the Latin Square design, such that they were distributed across four lists. Each participant saw six practice trials. Presentation order was pseudo-randomized, such that each experimental sentence was followed by two fillers. After the self-paced reading task, L1 participants performed a test to verify their knowledge of German. L2 participants performed the untimed agreement test and the vocabulary test.

2.1.4 Analysis

Data processing. Trials containing nouns whose plural form was incorrectly given in the vocabulary test were excluded from analysis (4.66% of trials in L1 English and 8.33% in L1 Spanish). Reading times lower than 200 ms or higher than 4000 ms were excluded (Chiuchiù & Benati, 2020; Shoghi et al., 2022). In the verb region, this resulted in the exclusion of 3.02% of reading times in the German group, 0.20% in the English group, and 0.00% in the Spanish group. In the critical verb + 1 region, this resulted in the exclusion of 2.02% of the reading times in the German group, 0.20% in the English group, and 0.00% in the Spanish group.

Reading times. All statistical models were Bayesian linear mixed-effects models fit to the self-paced reading data in the software R (R Development Core Team, 2024). The dependent variable was the reading time in the critical verb + 1 region. Reading times were reciprocally transformed, following the Box-Cox procedure (Box & Cox, 1964). Maximal random-effects structures were used to jointly capture by-item and by-participant variability.

The statistical analysis consisted of two stages. In the first stage, each speaker group was evaluated individually, in order to diagnose grammaticality and attraction effects in the three tested samples. In the second stage, between-group comparisons were performed, to address our research questions. Specifically, we evaluated the interaction of Grammaticality and Attraction with either Group (L1 vs. L2) or L1 status (L1 English vs. L1 Spanish). We report these analyses in two separate sections, one for the L1 vs. L2 group comparisons and the other for the L1 English vs. L1 Spanish comparisons.

The rationale for performing separate L1–L2 and Spanish–English comparisons is twofold. First, L1–L2 comparisons addressed Cunnings’ Interference Hypothesis (Cunnings 2017a,b), which is specifically about whether L2 speakers differ from L1 speakers in their susceptibility to similarity-based interference. The Interference Hypothesis does not make predictions about morphological richness, but rather about L1- vs. L2-status. Second, the Spanish–English speaker comparisons allowed us to use self-rated proficiency in German as an additional predictor – this factor was absent for L1 German speakers – and to use model priors that were informed by previous research, especially in the case of L1–L2 comparisons. We are not aware of previous reading time research that would have allowed us to build informative priors for other comparisons, for example, L1 German vs. L1 English (or L1 German vs. L1 Spanish).

The model that assessed the effect of grammaticality was applied in the conditions in which the subject and attractor noun matched in number, to avoid any potential attraction effects. This model used as a predictor Grammaticality (sum coded, -0.5 grammatical/ 0.5 ungrammatical). In the model comparing the L1–L2 groups, Grammaticality was used as a fixed effect, together with Group (sum coded, -0.5 L1 German/ 0.5 L2 German) and their interaction. In the model comparing English and Spanish speakers, Grammaticality was used together with L1 (sum coded, -0.5 L1 English/ 0.5 L1 Spanish) and their interaction. The key comparisons concerned whether grammaticality effects differed either between L1–L2 groups or between the two groups of German learners. Therefore, the key effects of interest were the Grammaticality \times Group and the Grammaticality \times L1 interactions. Participants’ self-rated proficiency in German was centered and used as an additional predictor.

The model that assessed the effect of attraction was run in the ungrammatical conditions, which is where attraction effects have been consistently found in previous research. The attraction model used as a predictor Match (sum coded, -0.5 match/ 0.5 mismatch). An effect of attraction was

predicted to emerge as a processing facilitation in the mismatch condition. In the model comparing the L1–L2 groups, Match was used as a fixed effect, together with Group and their interaction. In the model comparing English and Spanish speakers, Match was used with L1 and their interaction, together with participants' self-rated German proficiency. The key comparisons concerned whether attraction effects differed either between L1–L2 groups or between the two groups of German learners. Thus, the key effects of interest were the Match \times Group and the Match \times L1 interactions. For ease of reference, the results below use Match and Attraction interchangeably.

The models for the analysis of the individual groups used nested contrasts to estimate the effects of Grammaticality (sum coded, -1 grammatical-match/ 1 ungrammatical-match/ 0 mismatch), Attraction within ungrammatical sentences (sum coded, -1 ungrammatical-match/ 1 ungrammatical-mismatch/ 0 grammatical), and Attraction within grammatical sentences (sum coded, -1 grammatical-match/ 1 grammatical-mismatch/ 0 ungrammatical). This coding reflects the hypothesis matrix, which was then inverted to obtain a contrast matrix for the model (see code). The use of $-1/1$ for the nested contrasts in the hypothesis matrix allowed us to retain the $-0.5/0.5$ coding for the non-nested main effect of grammaticality.

We set the model priors based on our prior experience and on a meta-analysis of previous research on agreement attraction (Jäger et al., 2017). The priors for the model intercept encoded our assumption that reading times would be with 95% probability between 2–4 words per second, with a mean of three, $N(3, 0.5)$. In the L1–L2 group analyses, the grammaticality effect was assumed to most likely increase reading time by up to 0.15 of a word per second, with a mean of -0.15 , $N(-0.15, 0.075)$.² On the millisecond scale, this corresponds to an inhibitory effect of 50 ± 50 ms. The prior for the effect of attraction was based on the Bayesian meta-analytic estimate provided by Jäger and colleagues: -22 ms [$-36, -9$] ms. Our prior $N(0.065, 0.03)$ corresponded to a facilitatory effect of -22 ± 20 ms. The effect of group was expected to most likely increase reading time by up to 0.3 of a word per second, with a mean of -0.3 , $N(-0.3, 0.15)$. On the millisecond scale, this corresponds to an inhibitory effect of 100 ± 100 ms. The interactions of grammaticality and attraction with group were expected to most likely influence reading time by up to 0.4 of a word per second, with a mean of zero, $N(0, 0.2)$ – this encoded our uncertainty about the direction of any potential between-group differences.

In the models for the individual groups, all priors were identical to the L1–L2 group analyses. The nested effect of grammaticality within the match condition used the group analysis grammaticality prior of $N(-0.15, 0.075)$. The nested effect of attraction within ungrammatical

² The seemingly counterintuitive signs in the prior specifications are due to the reciprocal transformation of the reading time data. Due to the reciprocal transformation ($1000/\text{reading time}$), the effect of an experimental effect should be interpreted as a change in speed, rather than raw reading time. Thus, an *increase* in reading time (e.g., in ungrammatical vs. grammatical sentences) corresponds to a *decrease* in speed – thus, the negative sign on the prior of the grammaticality effect.

sentences used the prior of $N(0.065, 0.03)$, based on Jäger et al. (2017). An additional effect examined in this analysis was the effect of attraction within grammatical sentences. We based this prior on the meta-analytic estimate of the same effect in Jäger et al. (2017), which was a facilitatory effect of -7 ms $[-16, 4]$ ms. Therefore, we specified the mean of this prior as a speed-up of 0.021 words per second, $+/- 20$ ms: $N(0.021, .03)$. In the L2 analyses, all priors were identical to the L1–L2 group analysis, except for the fixed effects, which had less informative $N(0, 0.2)$ priors, due to the lack of meta-analytic estimates on L2 attraction.

The prior for the random effects and the residual variance used a half-normal distribution $N+(0, 0.2)$ and $N+(0, 1)$, respectively, because random effects and residual variance cannot be negative. Within the variance-covariance matrices of the by-participant and by-item random effects, priors were defined for the correlation matrices using a so-called Lewandowski-Kurowicka-Joe (LKJ) prior (Lewandowski et al., 2009). This prior has a parameter η , which, when set to 2, has the regularizing effect of disfavoring extreme correlations.

Hypothesis tests with Bayes factors were conducted for the critical predictors in the between-group analyses, i.e., the interaction effects. Bayes factors quantified the ratio of evidence for the alternative hypothesis (H_1) versus the null hypothesis (H_0) by comparing a model with the interaction of interest versus a model without it. A BF_{10} of 1 suggests equivalent evidence for either hypothesis. In line with Lee and Wagenmakers's scale (2013) – adapted from Jeffreys (1961) – a BF_{10} above 1 indicates evidence for H_1 , supporting an interaction effect: Values between 1 and 3 are considered inconclusive, values between 3 and 10 provide moderate evidence, and values over 10 provide strong evidence. Conversely, a BF_{10} below 1 indicates evidence for H_0 , against an interaction effect: Values between 1 and 0.3 are considered inconclusive, between 0.3 and 0.1 provide moderate evidence, and below 0.1 provide strong evidence. Since Bayes factors are sensitive to the choice of prior, we conducted sensitivity analyses by computing Bayes factors at a range of prior standard deviations representing different ranges of a priori assumed plausible effect sizes (Schad et al., 2022).

Untimed subject-verb agreement test. A Bayesian generalized linear mixed-effects model was fit to the binomial accuracy data with a maximal random-effects structure and (centered) self-rated proficiency and L1 as predictors (L1 was sum contrast-coded: -0.5 L1 English/ 0.5 L1 Spanish). The prior for the L1 effect was $N(0, 1.5)$, and the prior for the model intercept encoded our assumption that the average probability of a correct response on the log-odds scale followed a normal distribution of $N(1, 1.5)$, with a lower bound of 0 – which corresponds to 50% on the percentage scale. This reflected the fact that participants with less than 50% accuracy in the untimed test were excluded prior to analysis, and so these scores could not occur in the data. To quantify evidence for the hypothesis that the two L1 groups differed in their untimed knowledge, we computed a Bayes factor comparing the ratio of evidence for a model with L1 as a predictor vs. one without.

2.2 Results

Comprehension question accuracy in the filler trials was above 95% across all groups (L1 German mean = 98%, SD = 2%; L1 English mean = 97%, SD = 4%; L1 Spanish mean = 96%, SD = 4%). Comprehension question accuracy in the experimental trials was also high (L1 German mean = 97%, SD = 4%; L1 English mean = 94%, SD = 6%; L1 Spanish mean = 90%, SD = 8%). All L2 speakers knew German subject-verb agreement constraints, as indicated by the untimed test (mean accuracy L1 English = 98%, SD = 5%; mean accuracy L1 Spanish = 96%, SD = 6%). The statistical model did not suggest an accuracy difference between the L2 groups, and the Bayes factor showed equivalent evidence for either hypothesis, $\hat{\beta} = -0.60$, 95% CrI [-1.45, 0.31], $BF_{10} = 0.9$.

Empirical reading times in the spillover region, together with the preceding and following regions, are shown in **Figure 1**. The results of the comparisons within and between groups are described below. For each comparison, we report the mean of the posterior probability distribution of each effect, together with a 95% credible interval (CrI), which represents the interval in which it is 95% certain that the true effect lies, given the data and the model. For easier interpretability, reading measures are backtransformed to milliseconds in the text, but statistical analyses and inferences are based on the untransformed model coefficients.

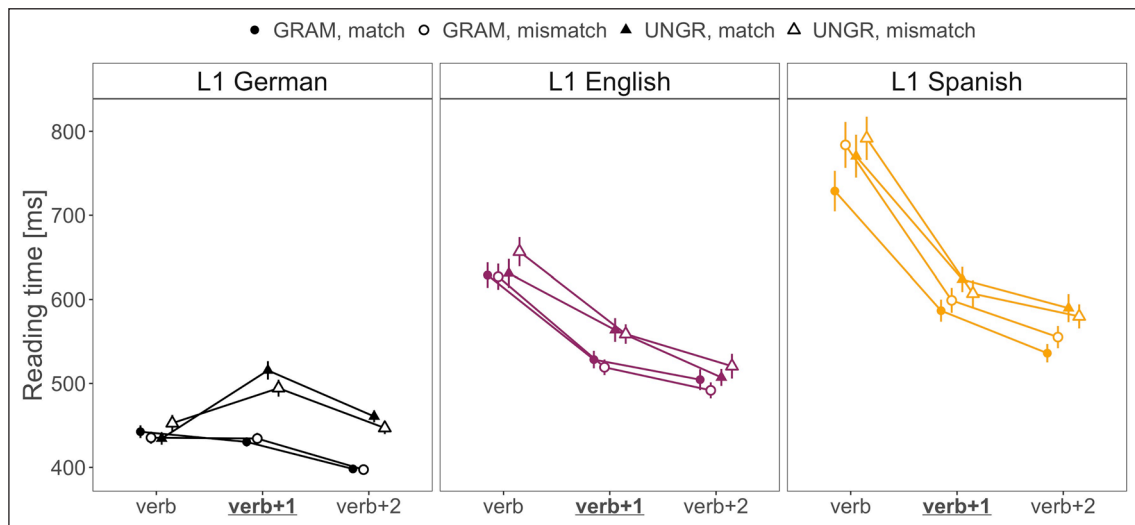


Figure 1: Empirical reading times in the critical *verb + 1* region, together with the preceding and following regions. Points show condition means and error bars ± 1 standard errors calculated for each condition across participants and items.

2.2.1 Individual groups

In the L1 German group, the model in the critical *verb + 1* region supported a slowdown of 31 ms in ungrammatical vs. grammatical sentences, with a 95% CrI spanning 21 to 41 ms, $\hat{\beta} = 31$

ms, 95% CrI [21,41] ms. The model was also consistent with a facilitatory attraction effect in ungrammatical sentences, $\hat{\beta} = -9$ ms, 95% CrI [-17,-2] ms, but not in grammatical sentences, $\hat{\beta} = -2$ ms, 95% CrI [-9,5] ms.

In the L1 Spanish group, the model was consistent with a slowdown in ungrammatical sentences, $\hat{\beta} = 19$ ms, 95% CrI [-3,42] ms. Since the 95% credible interval of the effect contained zero and negative effects, but it mainly appeared to favor positive values, we additionally calculated the proportion of the posterior samples that were consistent with a positive effect: 96% of the samples were compatible with a slowdown, $P(\beta > 0) = 0.96$. The model was also consistent with an attraction effect in ungrammatical sentences, $\hat{\beta} = -18$ ms, 95% CrI [-41,4] ms, $P(\beta < 0) = 0.95$, but not in grammatical sentences, $\hat{\beta} = -1$ ms, 95% CrI [-25,23] ms.

In the L1 English group, the model was consistent with a slowdown in ungrammatical sentences: $\hat{\beta} = 17$ ms, 95% CrI [0,34] ms, $P(\beta > 0) = 0.97$, but not with attraction in ungrammatical sentences, $\hat{\beta} = 9$ ms, 95% CrI [-7,25] ms, or grammatical sentences, $\hat{\beta} = 4$ ms, 95% CrI [-13,22] ms.

2.2.2 Group comparisons: L1 vs. L2

In addition to posterior means and credible intervals, this (and following) sections report Bayes factors, which allowed us to quantify evidence for the effects relevant to our research questions, i.e., differences in the size of grammaticality and attraction effects between groups. With regard to the effect of grammaticality, the model supported a difference between the L1 and L2 groups: compared to the L1 group, the L2 group showed less disruption in the spillover region of ungrammatical sentences, consistent with a reduced sensitivity to agreement violations, Grammaticality \times Group $\hat{\beta} = -27$ ms, 95% CrI [-44,-10] ms (**Figure 2**). A Bayes factor (BF_{10}) of 23 suggested strong evidence in favor of the interaction, and the sensitivity analysis showed that a different prior would not have changed this conclusion. The model was also consistent with a main effect of grammaticality: ungrammatical sentences were read more slowly across both groups, $\hat{\beta} = 27$ ms, 95% CrI [19,36] ms. There was also a main effect of group, with L2 speakers reading more slowly than L1 speakers across sentence types, $\hat{\beta} = 98$ ms, 95% CrI [71,124] ms.

With regard to the effect of attraction, we focused on ungrammatical sentences, which is where we predicted attraction to arise. There was inconclusive evidence against an Attraction \times Group interaction, suggesting similar effects in the L1 and L2 groups, $\hat{\beta} = 8$ ms, 95% CrI [-15,30] ms. The sensitivity analysis indicated that priors assuming a wider range of plausible effect sizes would have provided more conclusive evidence against an interaction, but that the evidence was inconclusive for more constraining priors. The model was consistent with a main effect of attraction, $\hat{\beta} = -10$ ms, 95% CrI [-17,-2] ms. The model was also consistent with slower reading times in L2 speakers than L1 speakers, $\hat{\beta} = 93$ ms, 95% CrI [67,121] ms.

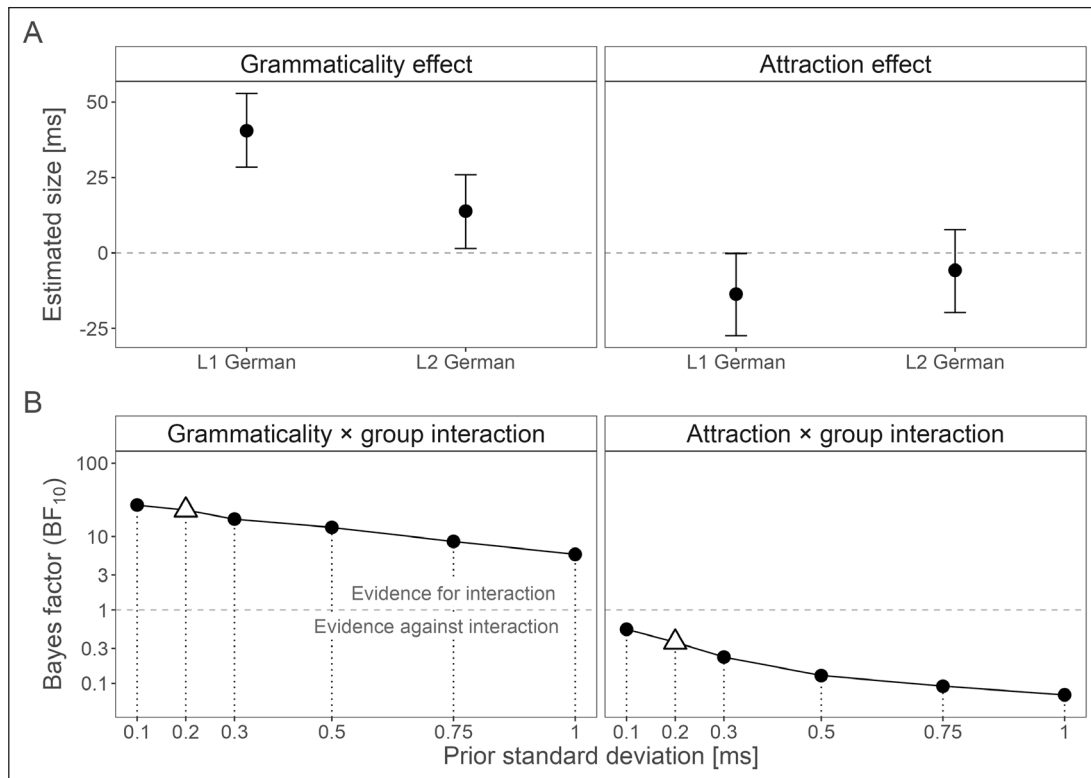


Figure 2: **A.** Model-estimated means and 95% credible intervals of grammaticality and attraction effects in L1 and L2 at the critical *verb + 1* region. **B.** Bayes factor sensitivity analyses for the interaction between grammaticality and group, and attraction and group. Bayes factors were computed at a range of prior standard deviations representing different ranges of a priori assumed plausible effect sizes. Triangles indicate the a priori prior standard deviations used in the main analysis and the Bayes factors reported in the text.

2.2.3 Learner comparisons: L1 English vs. L1 Spanish

With regard to grammaticality, there was moderate evidence against a Grammaticality × L1 interaction, suggesting that Spanish and English speakers did not differ in their sensitivity to agreement violations, $\hat{\beta} = 4$ ms, 95% CrI [-24,31] ms, $BF_{10} = 0.27$ (**Figure 3**). The sensitivity analysis indicated that a different choice of prior would not have changed our conclusions. By contrast, the model was consistent with a main effect of grammaticality, with ungrammatical sentences being read more slowly than grammatical sentences across L2 speakers, $\hat{\beta} = 17$ ms, 95% CrI [2,32] ms. There was also a main effect of L1, with Spanish speakers reading more slowly than English speakers, $\hat{\beta} = 47$ ms, 95% CrI [14,80] ms. The model was not consistent with an effect of proficiency, $\hat{\beta} = -7$ ms, 95% CrI [-95,78] ms.

With regard to attraction in ungrammatical sentences, the results were inconclusive about an Attraction × L1 interaction: while Spanish speakers had a numerically larger attraction effect than English speakers, the Bayes factor was inconclusive, $\hat{\beta} = -27$ ms, 95% CrI [-56,1] ms,

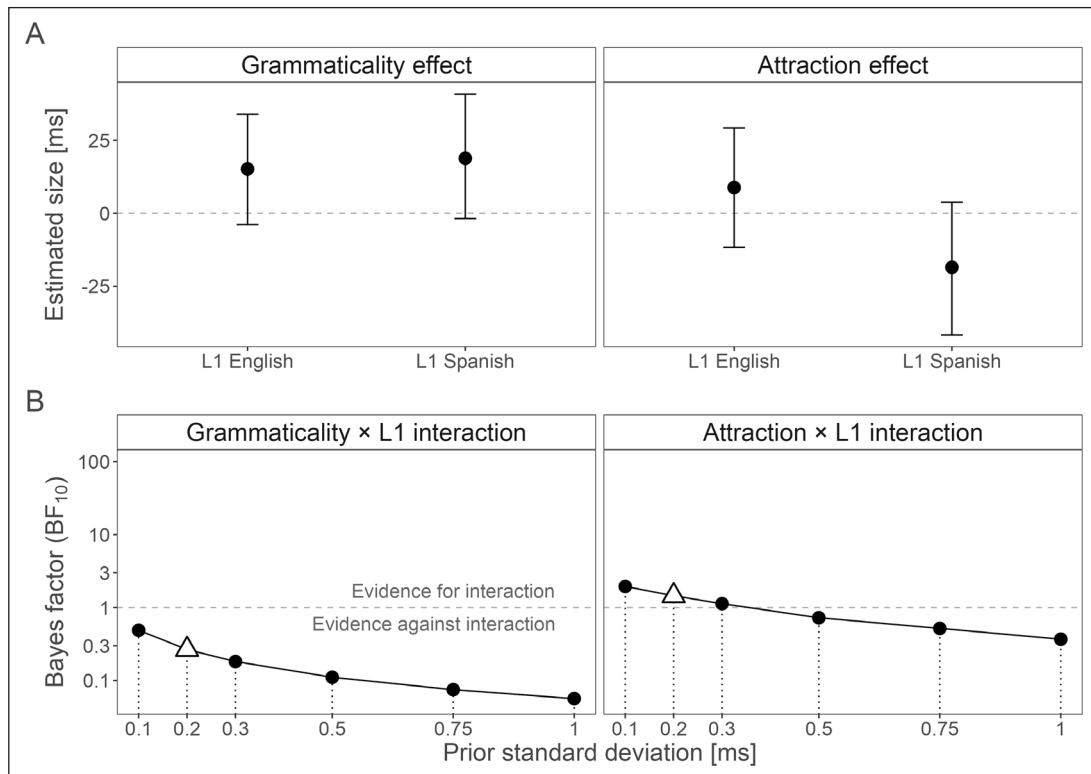


Figure 3: **A.** Model-estimated means and 95% credible intervals of grammaticality and attraction effects for L2 speakers at the critical *verb + 1* region. **B.** Bayes factor sensitivity analyses for the interaction between grammaticality and L1, and attraction and L1. Bayes factors were computed at a range of prior standard deviations representing different ranges of a priori assumed plausible effect sizes. Triangles indicate the a priori prior standard deviations used in the main analysis and the Bayes factors reported in the text.

$BF_{10} = 1.56$. The sensitivity analysis indicated that a different choice of prior would not have yielded a more conclusive Bayes factor. Additionally, the model was not consistent with a main effect of attraction, $\hat{\beta} = -5$ ms, 95% CrI [-22,11] ms, but it was suggestive of a facilitatory effect in the Spanish group, $\hat{\beta} = -19$ ms, 95% CrI [-42,4] ms. The model was consistent with slower reading times in Spanish speakers than English speakers, $\hat{\beta} = 37$ ms, 95% CrI [5,71] ms. The model was not consistent with an effect of proficiency, $\hat{\beta} = 0$ ms, 95% CrI [-90,88] ms.

3. General discussion

This study examined sensitivity to subject-verb number agreement through two between-group comparisons: (a) L1 vs. L2 German speakers; (b) intermediate-to-advanced L2 speakers of German whose L1 had either productive (Spanish) or less productive (English) number agreement morphology. To assess participants' implicit processing, we used a self-paced reading paradigm that did not involve grammaticality judgments. Two behavioral markers were examined.

First, processing disruptions to agreement violations were measured, to diagnose participants' ability to apply German agreement constraints in real time. Second, participants' susceptibility to agreement attraction was measured, to diagnose whether they used a retrieval mechanism to compute agreement and to address whether the productivity of their L1 agreement morphology affected retrieval processes.

Our results can be summarized as follows. With regard to sensitivity to agreement violations, L2 speakers showed smaller reading disruptions than L1 speakers after encountering a verb with an agreement violation. However, sensitivity to violations was similar in the two learner groups, regardless of L1 morphological productivity. With regard to agreement attraction, the results were inconclusive, i.e., they did not provide clear evidence for or against differential attraction between groups. Although attraction effects were numerically larger in L1 speakers than L2 speakers, and also in Spanish learners of German than in English learners of German, the Bayes factors associated with the relevant between-group interactions were inconclusive. We discuss the implications of these results below.

3.1 Agreement violations: Sensitivity decreases in a second language, but it is not modulated by L1 morphological richness

Compared to L1 speakers, L2 speakers were less sensitive to agreement violations. This may be surprising, given that we tested constructions in which the subject and verb were linearly close – separated only by an adverb – and that we focused on intermediate-to-advanced German learners: Most of our participants had spent years living in Germany, they rated themselves as proficient, and they were fully aware of German subject-verb agreement constraints, as attested by their accurate untimed test performance – 98% and 96% mean accuracy in L1 English and L1 Spanish, respectively. We also made sure that agreement processing was not confounded with lexical problems: participants' knowledge of the plural form of the critical nouns was separately assessed, and our analyses only included trials with appropriately identified plural forms.

Due to this, our results suggest that L2 speakers are less able than L1 speakers to apply knowledge about L2 inflectional morphology when reading sentences in real time (Jiang, 2007). Specifically, while knowledge about L2 morphology may be represented in the L2 grammar, its application during sentence processing might be effortful and require more attentional resources, as opposed to other types of knowledge that can be applied automatically and/or without conscious awareness. In this sense, our results echo findings in single-word studies, which suggest variable effects in the priming of inflectional morphemes (Ciaccio & Veríssimo, 2022; Feldman et al., 2010; Veríssimo et al., 2018).

Further, English and Spanish speakers showed comparable sensitivity to agreement violations, with moderate evidence against a between-group difference. Thus, having productive L1 agreement morphology does not automatically confer an advantage in real-time L2 use, at

least at high proficiency levels. This speaks against an extended version of the Morphological Congruency Hypothesis, according to which morphology can only be fully processed if a morpheme is congruent between a speaker's L1 and L2 (Jiang et al., 2011). We entertained an extended version of the Morphological Congruency Hypothesis based on a reading eye-tracking study in the domain of tense processing (Sagarra & Ellis, 2013). This study found that learners whose L1 had productive tense morphology – Romanian – were more sensitive to L2 verb tense violations than learners whose L1 had poorer morphology – English.

We can think of two explanations for the different findings across studies. The first is methodological: the eye-tracking method used by Sagarra and Ellis (2013) has a better temporal resolution than self-paced reading. For example, if our study had used eye-tracking, evidence of between-group differences could have emerged in measures related to the re-reading of the verb + 1 region and/or regressions to preceding sentence regions. Such measures are not available in a self-paced reading study.

Alternatively, the contrast between studies might be due to inherent differences between tense and number. Both are grammatical properties, but tense has stronger semantic consequences: a tense violation like *Yesterday the boy cooks* leaves comprehenders uncertain about the temporal dimension of the event, as it is unclear how to resolve the conflict between inflectional and lexical information. By contrast, number agreement, especially in languages with overt (non pro-drop) subjects, like English and German, is a process with reduced semantic import – the number of the subject is always available to comprehenders, because it is marked independently from the verb. As a result, verb agreement might not be salient – regardless of L1 status – which might further explain why reactions to agreement violations were reduced in L2 learners.

Finally, while our results do not support an extended version of the Morphological Congruency Hypothesis, they are consistent with the original version of this hypothesis, which concerns the congruency of a morpheme between a speaker's L1 and L2 – a morpheme is congruent if it conveys the same meaning across languages. Under this criterion, German plural agreement morphology is congruent for both Spanish and English speakers, as both languages have corresponding morphemes to express a number distinction in the third person. Therefore, the original Morphological Congruency Hypothesis would not predict differences between English and Spanish learners of German, which is consistent with our results.

3.2 Attraction effects do not increase in a second language, and the role of L1 morphological productivity is unclear

All between-group comparisons regarding agreement attraction yielded inconclusive results. Thus, we keep their discussion short, in order to avoid overinterpretation. Nonetheless, the observed patterns are informative for current discussions in the L2 literature. The first finding concerns the comparison of attraction effects between L1 and L2 speakers: Attraction effects

were numerically smaller in L2 speakers than in L1 speakers. Although the Bayes factor for this effect was inconclusive, the effect's direction was the opposite than predicted by the Interference Hypothesis (Cunnings, 2017a, 2017b). The Interference Hypothesis proposes that cue-based memory retrieval is used to compute syntactic dependencies, like subject-verb agreement. According to this hypothesis, L2 speakers are more prone to interference than L1 speakers; thus, attraction effects – which are due to interference – should be larger in L2 than L1. But our results suggest that agreement attraction, if anything, is smaller in L2 German than in L1 German. While the inconclusiveness of the analysis prevents us from claiming that L2 effects were conclusively smaller, they were clearly not larger.

We suggest that our results demonstrate reduced sensitivity to inflectional morphology in L2, i.e., difficulties with morphological processing, rather than increased interference during syntactic licensing. According to an error-driven view of attraction, comprehenders only initiate a retrieval process when they detect that the number of the verb does not match the number predicted by the parser on the basis of the number of the subject (Lago et al., 2015; Schlueter et al., 2019; Wagers et al., 2009). But in order to notice an error, L2 speakers need to have successfully implemented at least two processes: (i) they have to generate a number prediction after reading the subject (e.g., a projected memory chunk for the verb with 3rd person plural features); (ii) they have to map the morphological information on the verb to syntactic features, in order to compare them directly against the original number prediction (i.e., map *-eten* in *unterzeichnen* to '3rd person plural'). There is evidence that both these processes are compromised in L2 speakers (for an overview of L2 predictive skills, see Schlenter, 2023; for evidence of processing difficulties with inflection, see Veríssimo et al., 2018). Under this explanation, the numerically smaller attraction effect in L2 was a downstream consequence of L2 speakers' reduced ability to perform processes (i) and/or (ii).

Before discussing between-group differences in L2, one limitation of the current study should be acknowledged: the observed attraction effects were always small, which could have created a floor effect and prevented the detection of between-group modulations: even the L1 German group showed a modest processing facilitation of -9 [-17,-2] ms. A smaller attraction effect might be expected, since we used comprehension questions (as opposed to acceptability judgment questions), which focused participants' attention on the meaning of the sentence, as opposed to its grammatical status. A second potential reason for the small attraction effects might be our choice of relative clause materials. In relative clauses, the attractor is not linearly adjacent to the verb, as is the case in prepositional constructions (e.g., *the key to the cabinets*). However, it has been shown that the linear distance between the attractor noun and the verb is not critical. Rather, it is the linear and syntactic distance between the target and attractor nouns that matters. For example, in double prepositional constructions (e.g., *the smell of the stable of the farmer*), attraction rates increase when the second – rather than the third – noun is plural (Franck

et al., 2002; Lago & Felser, 2018). Since the second noun is farther from the verb, but closer to the head noun, this suggests that linear and/or syntactic closeness to the agreement controller (rather than the verb) modulates attraction effects.

The increased distance between the target and attractor noun could explain the reduced attraction in our relative clause constructions, since the attractor noun was syntactically more distant from the subject noun. But this would be surprising, given previous findings with Korean speakers of English, who were more prone to attraction with relative clauses than prepositional phrases (Lee & Phillips, 2023; Lim & Christianson, 2015). At present, the role of the grammatical environment of the attractor and its potentially different role in L1 and L2 processing is unclear: more research will be necessary to resolve this issue. A final reason for our small attraction effects could be the animacy difference between the (inanimate) attractor and the (animate) target noun. Since the target verbs always required animate licensors (e.g., the verb *sign*), the use of animacy as a retrieval cue could have reduced interference and boosted the retrieval of the target noun.

Our second finding was that attraction effects were numerically larger in Spanish speakers than English speakers. The Bayes factor was inconclusive, but its direction supports an extended version of the Morphological Congruency Hypothesis, according to which the productivity of a L1 morpheme affects its L2 processing. With regard to attraction, an extended version of the Morphological Congruency Hypothesis might predict that Spanish speakers would weigh number features more highly at retrieval (due to their productive L1 agreement morphology). This should result in increased interference – and, thus, more attraction – in Spanish speakers than English speakers. Interestingly, since there was evidence against an extended Morphological Congruency Hypothesis with regard to agreement violations (see 3.1), proponents of this hypothesis might need to restrict it by assigning it a more selective role in L2 processing. For example, it could be proposed that L1 morphological productivity affects the weighting of shared L1–L2 features during syntactic retrieval (thus, yielding different attraction effects across L2 speakers), but not the initial detection of an agreement error (thus, yielding similar violation effects across L2 speakers).

As English-Spanish differences in attraction were inconclusive, more research is necessary to assess the hypothesis above. And it is also possible that the apparent difference arose due to a confound. Recall that Spanish speakers consistently showed longer reading times than English speakers, as evidenced by the main effect of L1 in all statistical comparisons. It is possible that the longer reading times in the verb + 1 region allowed the more successful detection of attraction effects in the Spanish group. By contrast, the attraction effect in English speakers might have been distributed across regions, thus preventing its reliable measurement.

Alternatively, differences in attraction effects between English and Spanish speakers might be real, but we may have lacked the statistical power to detect them. Recent work suggests that samples of more than 300 participants are needed to detect a 2×2 medium-size interaction in between-subject designs (Brysbaert, 2019). Crucially, attraction effects are small: on average

–22 ms [–36, –9] ms in ungrammatical sentences (Jäger et al., 2017). Because it is logistically challenging to recruit hundreds of L2 speakers – matched in demographic properties – we think that multi-laboratory collaborations will be crucial to properly assessing whether attraction is modulated by L1 morphological productivity.

4. Conclusion

This study examined whether sensitivity to subject-verb agreement differed between L1 and L2 speakers of German, as well as between intermediate-to-advanced L2 German speakers whose L1 differed in the productivity of agreement morphology – Spanish vs. English. Two experimental effects were measured: agreement violations and agreement attraction. We found that sensitivity to agreement violations was reduced in a second language, and that it was not modulated by the productivity of L1 agreement morphology. Meanwhile, attraction results did not support increased interference in a second language, and they were inconclusive about the role of L1 morphological productivity.

Abbreviations

NOM = nominative, ACC = accusative, SG = singular, PL = plural, 3PL = third person plural, 3SG = third person singular

Data accessibility statement

Experimental materials, data and analysis code, as well as all Supplemental Materials for this article are publicly available at <https://doi.org/10.17605/OSF.IO/VW6UF>.

Ethics and consent

The procedures of the study were reviewed and approved by an ethics committee at the University of Potsdam (reference number 16/2016).

Acknowledgments

We thank Lisa Becker and Eleonora Pinotti for their assistance with materials creation, participant recruitment, and data collection. We thank João Veríssimo and the members of the Knoeferle and Vasishth laboratories for useful discussion. We thank three anonymous reviewers and the associate editor, Dr. Matt Husband, for their thorough reviews of the article. Sol Lago, Elise Oltrogge and Kate Stone were supported by the German Research Foundation, project number 317308350.

Competing interests

The authors have no competing interests to declare.

Author contributions

Sol Lago: conceptualization, methodology, formal analysis (supporting), data curation, funding acquisition, writing – original draft, writing – review & editing. **Kate Stone:** formal analysis (lead), methodology, writing – review & editing. **Elise Oltrogge:** investigation, software, writing – review & editing.

ORCID IDs

Sol Lago: <https://orcid.org/0000-0002-4966-1913>

Elise Oltrogge: <https://orcid.org/0009-0004-4953-9186>

Kate Stone: <https://orcid.org/0000-0002-2180-9736>

References

- Alemán Bañón, J., Fiorentino, R., & Gabriele, A. (2018). Using event-related potentials to track morphosyntactic development in second language learners: The processing of number and gender agreement in Spanish. *PLOS ONE*, 13(7), e0200791. <https://doi.org/10.1371/journal.pone.0200791>
- Avetisyan, S., Lago, S., & Vasishth, S. (2020). Does case marking affect agreement attraction in comprehension? *Journal of Memory and Language*, 112, 104087. <https://doi.org/10.1016/j.jml.2020.104087>
- Bian, J., Zhang, H., & Sun, C. (2021). An ERP study on attraction effects in advanced L2 learners. *Frontiers in Psychology*, 12. <https://www.frontiersin.org/articles/10.3389/fpsyg.2021.616804>.
- Blanche, P., & Merino, B. J. (1989). Self-assessment of foreign-language skills: Implications for teachers and researchers. *Language Learning*, 39(3), 313–338. <https://doi.org/10.1111/j.1467-1770.1989.tb00595.x>
- Blom, E., Poliššenská, D., & Weerman, F. (2006). Effects of age on the acquisition of agreement inflection. *Morphology*, 16(2), 313–336. <https://doi.org/10.1007/s11525-007-9110-1>
- Bock, K., & Miller, C. A. (1991). Broken agreement. *Cognitive Psychology*, 23(1), 45–93. [https://doi.org/10.1016/0010-0285\(91\)90003-7](https://doi.org/10.1016/0010-0285(91)90003-7)
- Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, 26(2), 211–252. <http://www.jstor.org/stable/2984418>.
- Brysbaert, M. (2019). How many participants do we have to include in properly powered experiments? A tutorial of power analysis with reference tables. *Journal of Cognition*, 2(1), 16. <https://doi.org/10.5334/joc.72>
- Caffarra, S., Molinaro, N., Davidson, D., & Carreiras, M. (2015). Second language syntactic processing revealed through event-related potentials: An empirical review. *Neuroscience & Biobehavioral Reviews*, 51, 31–47. <https://doi.org/10.1016/j.neubiorev.2015.01.010>
- Chen, L., Shu, H., Liu, Y., Zhao, J., & Li, P. (2007). ERP signatures of subject–verb agreement in L2 learning. *Bilingualism: Language and Cognition*, 10(2), 161–174. <https://doi.org/10.1017/S136672890700291X>
- Cheng, Y., Cunnings, I., Miller, D., & Rothman, J. (2021). Double-number marking matters for both L1 and L2 processing of nonlocal agreement similarly: An ERP investigation. *Studies in Second Language Acquisition*, 1–21. <https://doi.org/10.1017/S0272263121000772>
- Chiuchiù, G., & Benati, A. (2020). The effects of structured input and textual enhancement on the acquisition of Italian subjunctive: A self-paced reading study. *Instructed Second Language Acquisition*, 4(2), 235–257. <https://doi.org/10.1558/isla.40659>
- Choi, S. H., & Ionin, T. (2021). Plural marking in the second language: Atomicity, definiteness, and transfer. *Applied Psycholinguistics*, 42(3), 549–578. <https://doi.org/10.1017/S0142716420000569>
- Ciaccio, L. A., & Veríssimo, J. (2022). Investigating variability in morphological processing with Bayesian distributional models. *Psychonomic Bulletin & Review*, 29(6), 2264–2274. <https://doi.org/10.3758/s13423-022-02109-w>

- Clahsen, H., & Felser, C. (2018). Some notes on the Shallow Structure Hypothesis. *Studies in Second Language Acquisition*, 40(3), 693–706. <https://doi.org/10.1017/S0272263117000250>
- Cunnings, I. (2017a). Interference in native and non-native sentence processing. *Bilingualism: Language and Cognition*, 20(4), 712–721. <https://doi.org/10.1017/S1366728916001243>
- Cunnings, I. (2017b). Parsing and working memory in bilingual sentence Processing. *Bilingualism: Language and Cognition*, 20(4), 659–678. <https://doi.org/10.1017/S1366728916000675>
- Curcic, M., Andringa, S., & Kuiken, F. (2019). The role of awareness and cognitive aptitudes in L2 predictive language processing. *Language Learning*, 69(S1), 42–71. <https://doi.org/10.1111/lang.12321>
- De Vincenzi, M., Job, R., Di Matteo, R., Angrilli, A., Penolazzi, B., Ciccarelli, L., & Vespignani, F. (2003). Differences in the perception and time course of syntactic and semantic violations. *Brain and Language*, 85(2), 280–296. [https://doi.org/10.1016/S0093-934X\(03\)00055-5](https://doi.org/10.1016/S0093-934X(03)00055-5)
- Dillon, B., Mishler, A., Sloggett, S., & Phillips, C. (2013). Contrasting intrusion profiles for agreement and anaphora: Experimental and modeling evidence. *Journal of Memory and Language*, 69(2), 85–103. <https://doi.org/10.1016/j.jml.2013.04.003>
- Dressler, W. U. (2003). Degrees of grammatical productivity in inflectional morphology. *Rivista Di Linguistica*, 15, 31–62.
- Drummond, A. (2013). *Ibex Farm* [Computer software]. <http://spellout.net/ibexfarm>.
- Dussias, P. E., Valdés Kroff, J. R., Guzzardo Tamargo, R. E., & Gerfen, C. (2013). When gender and looking go hand in hand: Grammatical gender processing in L2 Spanish. *Studies in Second Language Acquisition*, 35(2), 353–387. <https://doi.org/10.1017/S0272263112000915>
- Engelmann, F., Jäger, L. A., & Vasishth, S. (2019). The effect of prominence and cue association on retrieval processes: A computational account. *Cognitive Science*, 43(12), e12800. <https://doi.org/10.1111/cogs.12800>
- Feldman, L. B., Kostić, A., Basnight-Brown, D. M., Đurđević, D. F., & Pastizzo, M. J. (2010). Morphological facilitation for regular and irregular verb formations in native and non-native speakers: Little evidence for two distinct mechanisms. *Bilingualism: Language and Cognition*, 13(2), 119–135. <https://doi.org/10.1017/S1366728909990459>
- Felix, S. W. (1976). Interference, interlanguage, and related issues. In S. W. Felix (Ed.), *Second language development. Trends and issues* (pp. 93–107). Narr.
- Foucart, A., & Frenck-Mestre, C. (2011). Grammatical gender processing in L2: Electrophysiological evidence of the effect of L1–L2 syntactic similarity. *Bilingualism: Language and Cognition*, 14(3), 379–399. <https://doi.org/10.1017/S136672891000012X>
- Franceschina, F. (2005). *Fossilized second language grammars: The acquisition of grammatical gender*. John Benjamins. <https://doi.org/10.1075/lald.38>
- Franck, J., Vigliocco, G., & Nicol, J. (2002). Subject-verb agreement errors in French and English: The role of syntactic hierarchy. *Language and Cognitive Processes*, 17(4), 371–404. <https://doi.org/10.1080/01690960143000254>

- Gillon Dowens, M. G., Guo, T., Guo, J., Barber, H., & Carreiras, M. (2011). Gender and number processing in Chinese learners of Spanish – Evidence from Event Related Potentials. *Neuropsychologia*, 49(7), 1651–1659. <https://doi.org/10.1016/j.neuropsychologia.2011.02.034>
- Gillon Dowens, M. G., Vergara, M., Barber, H. A., & Carreiras, M. (2010). Morphosyntactic processing in late second-language learners. *Journal of Cognitive Neuroscience*, 22(8), 1870–1887. <https://doi.org/10.1162/jocn.2009.21304>
- Guillelmon, D., & Grosjean, F. (2001). The gender marking effect in spoken word recognition: The case of bilinguals. *Memory & Cognition*, 29(3), 503–511. <https://doi.org/10.3758/BF03196401>
- Hammerly, C., Staub, A., & Dillon, B. (2019). The grammaticality asymmetry in agreement attraction reflects response bias: Experimental and modeling evidence. *Cognitive Psychology*, 110, 70–104. <https://doi.org/10.1016/j.cogpsych.2019.01.001>
- Hawkins, R., & Franceschina, F. (2004). Explaining the acquisition and non-acquisition of determiner-noun gender concord in French and Spanish. In P. Prévost & J. Paradis (Eds.), *The acquisition of French in different contexts: Focus on functional categories* (pp. 175–205). John Benjamins Publishing Company. <https://doi.org/10.1075/lald.32.10haw>
- Hopp, H. (2022). Second language sentence processing. *Annual Review of Linguistics*, 8(1), 235–256. <https://doi.org/10.1146/annurev-linguistics-030821-054113>
- Hopp, H., & Lemmerth, N. (2018). Lexical and syntactic congruency in L2 predictive gender processing. *Studies in Second Language Acquisition*, 40(1), 171–199. <https://doi.org/10.1017/S0272263116000437>
- Hoshino, N., Dussias, P. E., & Kroll, J. F. (2010). Processing subject–verb agreement in a second language depends on proficiency. *Bilingualism: Language and Cognition*, 13(2), 87–98. <https://doi.org/10.1017/S1366728909990034>
- Ionin, T., Choi, S. H., & Liu, Q. (2021). Knowledge of indefinite articles in L2-English: Online vs. offline performance. *Second Language Research*, 37(1), 121–160. <https://doi.org/10.1177/0267658319857466>
- Jacob, G., Lago, S., & Patterson, C. (2017). L2 processing and memory retrieval: Some empirical and conceptual challenges. *Bilingualism: Language and Cognition*, 20(4), 691–693. <https://doi.org/10.1017/S1366728916000948>
- Jäger, L. A., Engelmann, F., & Vasishth, S. (2017). Similarity-based interference in sentence comprehension: Literature review and Bayesian meta-analysis. *Journal of Memory and Language*, 94, 316–339. <https://doi.org/10.1016/j.jml.2017.01.004>
- Jäger, L. A., Mertzen, D., Van Dyke, J. A., & Vasishth, S. (2020). Interference patterns in subject-verb agreement and reflexives revisited: A large-sample study. *Journal of Memory and Language*, 111, 104063. <https://doi.org/10.1016/j.jml.2019.104063>
- Jarvis, S. (2000). Methodological rigor in the study of transfer: Identifying L1 influence in the interlanguage lexicon. *Language Learning*, 50(2), 245–309. <https://doi.org/10.1111/0023-8333.00118>
- Jeffreys, H. (1961). *The theory of probability*. Clarendon.

- Jegerski, J. (2016). Number attraction effects in near-native Spanish sentence comprehension. *Studies in Second Language Acquisition*, 38(1), 5–33. <https://doi.org/10.1017/S027226311400059X>
- Jia, G., & Aaronson, D. (2003). A longitudinal study of Chinese children and adolescents learning English in the United States. *Applied Psycholinguistics*, 24(1), 131–161. <https://doi.org/10.1017/S0142716403000079>
- Jiang, N. (2004). Morphological insensitivity in second language processing. *Applied Psycholinguistics*, 25(4), 603–634. <https://doi.org/10.1017/S0142716404001298>
- Jiang, N. (2007). Selective integration of linguistic knowledge in adult second language learning. *Language Learning*, 57(1), 1–33. <https://doi.org/10.1111/j.1467-9922.2007.00397.x>
- Jiang, N., Hu, G., Chrabaszcz, A., & Ye, L. (2017). The activation of grammaticalized meaning in L2 processing: Toward an explanation of the morphological congruency effect. *International Journal of Bilingualism*, 21(1), 81–98. <https://doi.org/10.1177/1367006915603823>
- Jiang, N., Novokshanova, E., Masuda, K., & Wang, X. (2011). Morphological congruency and the acquisition of L2 morphemes. *Language Learning*, 61(3), 940–967. <https://doi.org/10.1111/j.1467-9922.2010.00627.x>
- Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, 21(1), 60–99. [https://doi.org/10.1016/0010-0285\(89\)90003-0](https://doi.org/10.1016/0010-0285(89)90003-0)
- Kail, M. (1989). Cue Validity, cue Cost, and processing types in sentence comprehension in French and Spanish. In B. MacWhinney & E. Bates (Eds.), *The crosslinguistic study of sentence processing* (pp. 77–117). Cambridge University Press.
- Koch, E. M., Bulté, B., Housen, A., & Godfroid, A. (2021). Using verb morphology to predict subject number in L1 and L2 sentence processing: A visual-world eye-tracking experiment. *Journal of the European Second Language Association*, 5(1), Article 1. <https://doi.org/10.22599/jesla.79>
- Lago, S., Acuña Fariña, C., & Meseguer, E. (2021). The reading signatures of agreement attraction. *Open Mind*, 5, 132–153. https://doi.org/10.1162/opmi_a_00047
- Lago, S., & Felser, C. (2018). Agreement attraction in native and nonnative speakers of German. *Applied Psycholinguistics*, 39(3), 619–647. <https://doi.org/10.1017/S0142716417000601>
- Lago, S., Shalom, D. E., Sigman, M., Lau, E. F., & Phillips, C. (2015). Agreement attraction in Spanish comprehension. *Journal of Memory and Language*, 82, 133–149. <https://doi.org/10.1016/j.jml.2015.02.002>
- Lago, S., Stone, K., Oltrogge, E., & Veríssimo, J. (2023). Possessive processing in bilingual comprehension. *Language Learning*, 73(3), 904–941. <https://doi.org/10.1111/lang.12556>
- Lardiere, D. (1998). Dissociating syntax from morphology in a divergent L2 end-state grammar. *Second Language Research*, 14(4), 359–375. <https://doi.org/10.1191/026765898672500216>
- Lee, E.-K. R., & Phillips, C. (2023). Why non-native speakers sometimes outperform native speakers in agreement processing. *Bilingualism: Language and Cognition*, 26(1), 152–164. <https://doi.org/10.1017/S1366728922000414>

- Lee, M. D., & Wagenmakers, E.-J. (2013). *Bayesian cognitive modeling: A practical course*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139087759>
- Lewandowski, D., Kurowicka, D., & Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis*, 100(9), 1989–2001. <https://doi.org/10.1016/j.jmva.2009.04.008>
- Lim, J. H., & Christianson, K. (2015). Second language sensitivity to agreement errors: Evidence from eye movements during comprehension and translation. *Applied Psycholinguistics*, 36(6), 1283–1315. <https://doi.org/10.1017/S0142716414000290>
- Long, M. H. (2003). Stabilization and fossilization in interlanguage development. In *The handbook of second language acquisition* (pp. 487–535). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9780470756492.ch16>
- MacWhinney, B. (1992). Transfer and Competition in second language learning. In R. J. Harris (Ed.), *Advances in psychology* (Vol. 83, pp. 371–390). North-Holland. [https://doi.org/10.1016/S0166-4115\(08\)61506-X](https://doi.org/10.1016/S0166-4115(08)61506-X)
- MacWhinney, B. (2001). The competition model: The input, the context, and the brain. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 69–90). Cambridge University Press. <https://doi.org/10.1017/CBO9781139524780.005>
- MacWhinney, B. (2005). A unified model of language acquisition. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 49–67). Oxford University Press. <https://doi.org/10.1093/oso/9780195151770.003.0004>
- MacWhinney, B., Bates, E., & Kliegl, R. (1984). Cue validity and sentence interpretation in English, German, and Italian. *Journal of Verbal Learning and Verbal Behavior*, 23(2), 127–150. [https://doi.org/10.1016/S0022-5371\(84\)90093-8](https://doi.org/10.1016/S0022-5371(84)90093-8)
- Mak, W. M., Vonk, W., & Schriefers, H. (2002). The Influence of animacy on relative clause processing. *Journal of Memory and Language*, 47(1), 50–68. <https://doi.org/10.1006/jmla.2001.2837>
- Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research*, 50(4), 940–967. [https://doi.org/10.1044/1092-4388\(2007/067\)](https://doi.org/10.1044/1092-4388(2007/067))
- Molinaro, N., Barber, H. A., & Carreiras, M. (2011). Grammatical agreement processing in reading: ERP findings and future directions. *Cortex*, 47(8), 908–930. <https://doi.org/10.1016/j.cortex.2011.02.019>
- Nicenboim, B., & Vasishth, S. (2018). Models of retrieval in sentence comprehension: A computational evaluation using Bayesian hierarchical modeling. *Journal of Memory and Language*, 99, 1–34. <https://doi.org/10.1016/j.jml.2017.08.004>
- Nicol, J. L., Forster, K. I., & Veres, C. (1997). Subject–verb agreement processes in comprehension. *Journal of Memory and Language*, 36(4), 569–587. <https://doi.org/10.1006/jmla.1996.2497>
- Patson, N. D., & Husband, E. M. (2016). Misinterpretations in agreement and agreement attraction. *Quarterly Journal of Experimental Psychology*, 69(5), 950–971. <https://doi.org/10.1080/17470218.2014.992445>

- Pearlmutter, N. J., Garnsey, S. M., & Bock, K. (1999). Agreement processes in sentence comprehension. *Journal of Memory and Language*, 41(3), 427–456. <https://doi.org/10.1006/jmla.1999.2653>
- Pienemann, M., Di Biase, B., Kawaguchi, S., & Håkansson, G. (2005). Processing constraints on L1 transfer. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches*. Oxford University Press.
- Prévost, P., & White, L. (2000). Missing surface inflection or impairment in second language acquisition? Evidence from tense and agreement. *Second Language Research*, 16(2), 103–133. <https://doi.org/10.1191/026765800677556046>
- R Development Core Team. (2024). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. <https://www.R-project.org>.
- Ross, S. (1998). Self-assessment in second language testing: A meta-analysis and analysis of experiential factors. *Language Testing*, 15(1), 1–20. <https://doi.org/10.1191/026553298666994244>
- Sabourin, L., & Stowe, L. A. (2008). Second language processing: When are first and second languages processed similarly? *Second Language Research*, 24(3), 397–430. <https://doi.org/10.1177/0267658308090186>
- Sagarra, N., & Ellis, N. C. (2013). From seeing adverbs to seeing verbal morphology: Language experience and adult acquisition of L2 tense. *Studies in Second Language Acquisition*, 35(2), 261–290. <https://doi.org/10.1017/S0272263112000885>
- Schad, D. J., Nicenboim, B., Bürkner, P.-C., Betancourt, M., & Vasissth, S. (2022). Workflow techniques for the robust use of bayes factors. *Psychological Methods*, 28(6), 1404–1426. <https://doi.org/10.1037/met0000472>
- Scherag, A., Demuth, L., Rösler, F., Neville, H. J., & Röder, B. (2004). The effects of late acquisition of L2 and the consequences of immigration on L1 for semantic and morpho-syntactic language aspects. *Cognition*, 93(3), B97–B108. <https://doi.org/10.1016/j.cognition.2004.02.003>
- Schlenter, J. (2023). Prediction in bilingual sentence processing: How prediction differs in a later learned language from a first language. *Bilingualism: Language and Cognition*, 26(2), 253–267. <https://doi.org/10.1017/S1366728922000736>
- Schlueter, Z., Parker, D., & Lau, E. (2019). Error-driven retrieval in agreement attraction rarely leads to misinterpretation. *Frontiers in Psychology*, 10. <https://doi.org/10.3389/fpsyg.2019.01002>
- Shen, E. Y., Staub, A., & Sanders, L. D. (2013). Event-related brain potential evidence that local nouns affect subject–verb agreement processing. *Language and Cognitive Processes*, 28(4), 498–524. <https://doi.org/10.1080/01690965.2011.650900>
- Shibuya, M., & Wakabayashi, S. (2008). Why are L2 learners not always sensitive to subject-verb agreement? *EUROSLA Yearbook*, 8, 235–258. <https://doi.org/10.1075/eurosla.8.13shi>
- Shoghi, S., Arslan, S., Bastiaanse, R., & Popov, S. (2022). Does a walk-through video help the parser down the garden-path? A visually enhanced self-paced reading study in Dutch. *Frontiers in Psychology*, 13. <https://doi.org/10.3389/fpsyg.2022.1009265>

- Staub, A. (2009). On the interpretation of the number attraction effect: Response time evidence. *Journal of Memory and Language*, 60(2), 308–327. <https://doi.org/10.1016/j.jml.2008.11.002>
- Tanner, D., Nicol, J., & Brehm, L. (2014). The time-course of feature interference in agreement comprehension: Multiple mechanisms and asymmetrical attraction. *Journal of Memory and Language*, 76, 195–215. <https://doi.org/10.1016/j.jml.2014.07.003>
- Tanner, D., Nicol, J., Herschensohn, J., & Osterhout, L. (2012). Electrophysiological markers of interference and structural facilitation in native and nonnative agreement processing. *Proceedings of the 36th Annual Boston University Conference on Language Development*, 594–606.
- Veríssimo, J., Heyer, V., Jacob, G., & Clahsen, H. (2018). Selective effects of age of acquisition on morphological priming: Evidence for a sensitive period. *Language Acquisition*, 25(3), 315–326. <https://doi.org/10.1080/10489223.2017.1346104>
- Wagers, M. W., Lau, E. F., & Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language*, 61(2), 206–237. <https://doi.org/10.1016/j.jml.2009.04.002>
- White, E. J., Genesee, F., & Steinhauer, K. (2012). Brain responses before and after intensive second language learning: Proficiency based changes and first language background effects in adult Learners. *PLOS ONE*, 7(12), e52318. <https://doi.org/10.1371/journal.pone.0052318>
- Xanthos, A., Laaha, S., Gillis, S., Stephany, U., Aksu-Koç, A., Christofidou, A., Gagarina, N., Hrzica, G., Ketrez, F. N., Kilani-Schoch, M., Korecky-Kröll, K., Kovačević, M., Laalo, K., Palmović, M., Pfeiler, B., Voeikova, M. D., & Dressler, W. U. (2011). On the role of morphological richness in the early development of noun and verb inflection. *First Language*, 31(4), 461–479. <https://doi.org/10.1177/0142723711409976>

