Agreement attraction in native and nonnative speakers of German

SOL LAGO and CLAUDIA FELSER
Potsdam Research Institute for Multilingualism

ADDRESS FOR CORRESPONDENCE
Sol Lago, Potsdam Research Institute for Multilingualism, University of Potsdam, Campus Golm,
Haus 2, Karl-Liebknecht-Strasse 24-25, 14476 Potsdam, Germany.
E-mail: marlago@uni-potsdam.de

ABSTRACT
Second language speakers often struggle to apply grammatical constraints such as subject–verb agreement. One hypothesis for this difficulty is that it results from problems suppressing syntactically unlicensed constituents in working memory. We investigated which properties of these constituents make them more likely to elicit errors: their grammatical distance to the subject head or their linear distance to the verb. We used double modifier constructions (e.g., *the smell of the stables of the farmers*), where the errors of native speakers are modulated by the linguistic relationships between the nouns in the subject phrase: second plural nouns, which are syntactically and semantically closer to the subject head, elicit more errors than third plural nouns, which are linearly closer to the verb (2nd-3rd-noun asymmetry). In order to dissociate between grammatical and linear distance, we compared embedded and coordinated modifiers, which were linearly identical but differed in grammatical distance. Using an attraction paradigm, we showed that German native speakers and proficient Russian speakers of German exhibited similar attraction rates and that their errors displayed a 2nd-3rd-noun asymmetry, which was more pronounced in embedded than in coordinated constructions. We suggest that both native and second language learners prioritize linguistic structure over linear distance in their agreement computations.

Keywords: agreement attraction; German; linear distance; Russian

Research on sentence processing has revealed a mixed profile of successes and failures in how second language (L2) speakers apply grammatical constraints in real time. Whereas some constraints can be applied quickly and reliably, others pose problems, even for highly proficient speakers (for review, see Dallas & Kaan, 2008; Roberts, 2013). One contributing factor to these problems is difficulties in the integration of information across multiple levels of encoding (e.g., Clahsen & Felser, 2006a, 2006b; Hopp, 2016; Sorace, 2004). Here we focus on two factors, one grammatical (linguistic structure) and one nongrammatical (linear distance), and we examine their interaction in the application of a specific constraint, subject–verb agreement. We inspect a phenomenon, called agreement attraction, which has provided useful insight into the nature of agreement computations in a first language (Bock & Miller, 1991). Our study uses attraction to address how L2 speakers weight linguistic structure and linear distance in a nonnative language.

Difficulties with agreement are well attested in L2 research. These difficulties affect comprehension and production, and they appear even when speakers demonstrate accurate untimed knowledge of agreement constraints (Chen, Shu, Liu, Zhao, & Li, 2007; Grüter, Lew-Williams, & Fernald, 2012; Jiang, 2004; Lim & Christianson, 2015; Sato & Felser, 2010; Shibuya & Wakabayashi, 2008; Van-Patten, Keating, & Leeser, 2012; Wakabayashi, 1997). However, fewer studies have addressed the variables that might cause agreement errors, such as the linear distance between the (dis)agreeing elements. The studies that have examined this issue are Keating (2009, 2010) and Foote (2011).

Keating (2009) used an eye-tracking paradigm to examine whether gender agreement violations between nouns and adjectives were detected differently depending on whether they were adjacent and within the same phrase (e.g., *una casa pequeña*, "a.FEM house.FEM small.FEM") or across phrases and separated by intervening material (e.g., *la casa es bastante pequeña*, "the.FEM house.FEM is pretty small.FEM"). Participants included Spanish native speakers and English learners of Spanish with different proficiency levels. The results showed that native speakers always detected gender violations, whereas beginning and intermediate Spanish learners did not, and advanced learners were only able to detect them in the adjacent conditions.

However, note that Keating's study confounded linear and grammatical distance, as the nonadjacent conditions involved an additional phrase boundary. However, similar results were obtained by Foote (2011), who tested number and gender violations. In the number agreement conditions, the subject head and verb were always in different phrases, either adjacent (*I see that your father is/ *are from Texas*) or separated by intervening material (*The watch of the man is/ *are from Switzerland*). In two self-paced reading tasks, Spanish native speakers and English learners of Spanish were able to detect number violations, but the disruptions in reading times were significantly larger in adjacent configurations, suggesting that speakers' sensitivity to violations was diminished when the disagreeing elements were linearly more distant.

These previous studies suggest that the linear distance between the (dis)agreeing elements affects L2 speakers' ability to detect agreement violations. The stronger influence of linear distance in nonnative speakers was interpreted by Keating (2009) as support for the shallow structure hypothesis (SSH), which proposes that parsing is shallower in a second than in a native language, with greater reliance on semantic, associative, and surface information than on syntactic cues (Clahsen & Felser, 2006a, 2006b). Although the SSH itself does not equate surface information with linear distance, Keating interpreted his findings as supporting the SSH and proposed that the observed L2 agreement difficulties in nonadjacent configurations reflected a processing deficit, which was due to a reduced capacity to hold gender information in working memory across intervening material.

A stronger influence of linear distance in L2 than native language (L1) processing can be derived from a different group of accounts, which posit that L2 grammatical difficulties stem from working memory limitations (Coughlin & Tremblay, 2013; McDonald, 2006; Sagarra & Herschensohn, 2013). A recent instantiation of these accounts is offered by Cunnings (2016), whose proposal is relevant for this study because it concerns the phenomenon of interest, agreement attraction.

Attraction occurs when a verb fails to agree with its grammatical controller and agrees instead with a nearby but grammatically illicit modifier, called an *attractor*. For instance, in preambles like *the key to the cabinets*, the presence of the plural attractor *cabinets* can mislead speakers to wrongly produce a plural verb or to accept it as grammatical in comprehension. A dominant view is that attraction reflects errors that occur during cue-based memory retrieval (Badecker & Kuminiak, 2007; Lewis & Vasishth, 2005; Lorimor, Jackson, & Foote, 2015; Wagers, Lau, & Phillips, 2009). Under this account, when speakers read or produce a verb, they use its morphosyntactic features as cues to retrieve an appropriate controller from working memory. Since memory access mechanisms are noisy and susceptible to retrieval interference, the plural attractor *cabinets* is sometimes selected instead of *key*, misleading speakers to use a plural verb.

Cunnings's account proposes that some L2 difficulties result from an increased susceptibility to retrieval interference. Although this account does not deal with linear distance per se, if linear distance is augmented by introducing intervening material between subjects and verbs, it should affect retrieval interference, both by increasing the decay of the subject phrase in memory and by creating additional competing elements for retrieval. Therefore, if L2 speakers have more difficulty with memory retrieval than native speakers, they might show larger attraction effects and possibly a stronger role of the linear distance between the attractor nouns and the verb. For the example above, an increased susceptibility to interference could render L2 speakers more likely than native speakers to wrongly retrieve the plural attractor *cabinets*, which, due to its linear proximity to the verb, will have been processed more recently and should be more activated in memory than the subject head. The prediction that L2 speakers should make more attraction errors than native speakers has not been examined to date: although previous work has reliably found attraction effects in L2, the error rates of L2 speakers have not been systematically compared with those of native speakers (Foote, 2010; Hoshino, Dussias, & Kroll, 2010; Jacob et al., 2016; Jegerski, 2016; Lim & Christianson, 2015; Nicol & Greth, 2003; Nicol, Teller, & Greth, 2001; Tanner, 2011).

In short, previous work has suggested that the processing of agreement in a L2 might be error prone due to working memory limitations. The nature of these limitations predicts a stronger role for linear distance in L2 than L1 agreement processing: as the distance between the subject head and the verb increases, agreement computations should be more error prone, either because the subject head decays more in memory (Keating, 2009, 2010) or because linearly closer, intervening nouns are more prominent and thus more likely to be wrongly retrieved as agreement licensors (Cunnings, 2016).

However, these findings do not necessarily mean that L2 speakers prioritize linear over grammatical information in their agreement computations, as these two factors were confounded in previous work. In Keating's study, the diminished L2 sensitivity to gender violations could reflect a purely linear phenomenon (i.e., the number of intervening words) or a structural one (i.e., the depth of embedding of the modifying adjective relative to its controller noun). As discussed below, structural distance plays an important role in L1 attraction errors, raising the possibility that it might also affect L2 agreement computations. Moreover, previous work in this area is limited to comprehension, so it is unknown whether linear distance also

affects L2 production. To address these issues, our study directly compared linear and grammatical distance in L2 subject–verb agreement. In contrast with previous studies, we used a task more similar to production, and we manipulated the linear distance between the verb and the attractor noun, rather than the subject head. Our goal was to address whether L2 speakers were able to dissociate between linear and structural distance, and if so, which of these factors more strongly influenced their agreement computations. We compared attraction errors in native and L2 speakers, in order to assess whether attraction effects were stronger in a nonnative language. To elicit attraction, we used double modifier constructions, as described below.

THE 2ND-3RD-NOUN ASYMMETRY IN AGREEMENT ATTRACTION

Previous work initially assumed that attraction errors mostly occurred when a plural attractor linearly intervened between the subject head and the verb (Francis, 1986; Quirk, Greenbaum, Leech, & Svartvik, 1985). As explained before, under a memory retrieval account, the fact that the attractor is linearly closer to the verb than the subject head should render it more active in memory and thus more likely to be misretrieved.

However, subsequent work found that native speakers were more affected by the linguistic relationship between the attractor and the subject head than by its linear distance to the verb (Bock & Cutting, 1992; Franck, Vigliocco, & Nicol, 2002; Gillespie & Pearlmutter, 2011; Solomon & Pearlmutter, 2004; Vigliocco & Nicol, 1998). This work emphasized the role of linguistic over linear distance by showing that attractors within relative clauses (e.g., *The editor who rejected the books*) did not induce attraction as much as attractors within prepositional phrase modifiers (despite identical linear distance between the attractors and the verb) or that attraction occurred even in cases where the attractor and verb were not linearly contiguous, as in questions such as *Is/are the helicopter for the flights safe?* For example, a production study (Franck et al., 2002) manipulated the plurality of double modifier constructions by making either the first or the second modifier plural ([1a] vs. [1b]). The authors hypothesized that if linear distance to the verb was the main cause of agreement attraction, then more errors should be obtained when the attractor was linearly closer to the verb (1b). However, their results showed the opposite pattern: native speakers of English and French made more errors when the second (linearly farther), but not the third noun (linearly closer), was plural.

(1) a. The statue in the **gardens** by the mansion
    b. The statue in the garden by the **mansions**

Currently, there are two accounts of this 2nd-3rd-noun asymmetry. Both accounts agree that the grammatical relationship between the attractor and the subject head matters more than its linear distance to the verb, but they differ in whether they posit a syntactic (Franck et al., 2002) or semantic origin (Gillespie & Pearlmutter, 2011, 2013). Under a syntactic account, the asymmetry arises because the second noun is structurally closer, or less embedded, than the third noun with

respect to the subject head, which makes it more likely to determine the grammatical number of the subject phrase.[1] Alternatively, Gillespie and Pearlmutter have proposed a scope of planning account, under which the 2nd-3rd-noun asymmetry occurs because the second noun is semantically closer to the subject head than the third noun. Under this account, elements that are more semantically integrated are more likely to be planned simultaneously during the formulation of an utterance. Using a norming task adapted from Solomon and Pearlmutter (2004), Gillespie and Pearlmutter showed that in the materials by Franck et al., participants rated the head and the second noun as more tightly linked than the head and the third noun. They suggested that the combination of being linearly closer to the head and semantically more integrated rendered the second noun more active in memory when the number of the subject phrase was planned, and thus, more likely to influence its number encoding.

Our study was not designed to arbitrate between semantic and syntactic accounts of the 2nd-3rd-noun asymmetry. Instead, we focused on the larger question of whether L2 speakers were more sensitive to grammatical than linear distance. In what follows, we use the term "grammatical distance" to refer to the syntactic and/or semantic modification relationships between the constituents of the subject phrase, independent of their linear configuration. We hypothesized that if L2 speakers were more sensitive to linear distance than native speakers, they should have increased problems inhibiting the most recently encountered second modifier. In this case, they should show either no 2nd-3rd-noun asymmetry (in contrast with native speakers) or a reversal, with more errors for third than second plural nouns, because third plural nouns were linearly closer to the verb. In contrast, if L2 speakers are sensitive to grammatical relationships to a similar extent as native speakers, their attraction errors should also show a 2nd-3rd-noun asymmetry. This pattern would show a larger role of grammatical structure over linear distance, suggesting proficient use of linguistic cues in the processing of agreement in a L2.

THE PRESENT STUDY

We examined whether Russian speakers of German were more likely to make attraction errors than German native speakers, and whether their errors were modulated by the 2nd-3rd-noun asymmetry. We chose Russian speakers because their agreement and case systems largely pattern with German, thus minimizing the possibility of their performance being affected by negative L1 influence. Our materials were constructed in German using double modifier constructions similar to those in the study of Franck et al. (2002). One advantage of using German as a target language is that its case marking system made the syntactic structure of double modifier constructions unambiguous. This differed from previous studies in French and English, where prepositional modifiers were syntactically ambiguous, such as *by the mansion(s)* in (1), which could modify either the first or the second noun phrase. In contrast, our materials contained no modification ambiguities.

We also addressed another potential issue in the materials of Franck et al. Although these authors attributed the 2nd-3rd-noun asymmetry to the syntactic closeness between the second noun and the subject head, their findings cannot completely rule out a linear distance account because the second noun was both

structurally and linearly closer to the head noun. To address this concern, we added two conditions. In addition to having sentences similar to those by Franck et al. (*embedded conditions*: e.g., *the smell of the stable of the farmer*), we included a set of conditions that kept the linear distance between the modifiers constant but varied their syntactic and semantic relationship by coordinating them (*coordinated conditions*: e.g., *the smell of the stable and the farmer*).

Most formal analyses of coordination posit that conjunction phrases are hierarchically structured (Munn, 1993; Ross, 1967; Zoerner, 1995). They differ, however, in whether they consider that both elements of the conjunction are heads of the coordinated phrase, resulting in a multiheaded structure (Gazdar, Pullum, Sag, & Wasow, 1982; Ingria, 1995) or whether the coordinated phrase is headed by the conjunction *and* (Camacho, 1997; Munn, 1987). In the latter case, the conjuncts are treated as either specifiers, complements (Johannessen, 1998; Munn, 1987), or adjuncts (Munn, 1993).

Under most of these analyses, coordinated nouns have a closer semantic and syntactic relationship than embedded nouns (for review, see Progovac, 1998a, 1998b). Therefore, in contrast with the embedded conditions, where the second noun was syntactically and semantically closer to the head than the third noun, in the coordinated conditions, the asymmetry between the two modifiers was reduced. In the embedded conditions, the third noun modified the second one both syntactically and semantically, whereas in the coordinated conditions, the second and third nouns jointly modified the first noun. In the coordinated conditions, this should reduce the grammatical distance between the third and the head noun, effectively decreasing the 2nd-3rd-noun asymmetry. Thus, if agreement attraction is modulated by grammatical distance, we expected a 2nd-3rd-noun asymmetry in the embedded conditions but a smaller or no asymmetry in the coordinated conditions. This interaction would further support the role of grammatical structure in attraction, since any difference between second and third nouns in the embedded and coordinated conditions cannot be due to linear distance, which was held constant across the two constructions.

Agreement attraction was elicited using a novel experimental paradigm introduced by Staub (2009, 2010). In this paradigm, subject phrases are shown in rapid serial visual presentation and participants are asked to choose whether the singular or plural form of the verb *to be* provides a suitable continuation. This task is similar to spoken production tasks in that participants are given a sentence preamble and are asked to select an appropriate verb. The main difference is that in spoken tasks participants articulate the verb, whereas in the forced-choice task, they select it from a set of two options.

We adopted the forced-choice task because several of its properties make it advantageous for studying L2 processing. First, spoken production studies have to discard some percentage of the trials due to participants' producing false starts, uninflected verbs, or verbs with ambiguous number morphology. In contrast, the forced-choice task involves key presses, thus reducing the number of invalid trials and increasing experimental power. Second, there are known differences in how L2 speakers access and represent inflection: for instance, lexical retrieval takes longer and is more error prone for L2 than for native speakers (Haznedar & Schwartz, 1997; Prévost & White, 2000; Tokowicz, 2015). As our study focused on

morphosyntactic processing, we wanted to increase the likelihood that any differences between our speaker groups would result from their grammatical computations, as opposed to variation in the lexical retrieval of the verbs. Since the forced-choice task minimizes lexical variability by presenting the highly frequent and semantically empty verb *to be*, it allowed us to focus on the processing of the verb's plural feature itself, rather than on the way in which the feature was realized lexically. Third, the forced-choice task measures not only verb choices but also response latencies, and thus can offer insight into the time course of agreement computations, even in cases where no error is actually made.

EXPERIMENT 1

The first experiment examined whether German and Russian–German speakers showed a 2nd-3rd-noun asymmetry in agreement attraction. Based on previous studies, we expected this asymmetry to be present for native speakers. Our goal was to examine whether the asymmetry would be stronger in embedded than in coordinated conditions, as would be expected if attraction is more strongly influenced by syntactic or semantic structure than by linear distance.

For Russian–German speakers, our research questions were whether they were more prone to attraction than native speakers and whether they showed a 2nd-3rd-noun asymmetry. Based on previous findings, we hypothesized that linear distance might affect L2 speakers more strongly than native speakers. In this case, their agreement computations should be more influenced by the third noun, which was adjacent to the verb, resulting in either no 2nd-3rd-noun asymmetry or a reversal, with more errors for third than for second plural nouns. Alternatively, if proficient L2 speakers are more sensitive to grammatical than linear distance (similarly to native speakers), then they should display a 2nd-3rd-noun asymmetry. Furthermore, the asymmetry should be stronger in embedded than coordinated preambles, due to stronger differences in grammatical distance (either syntactic or semantic) for the former type of constructions (Franck et al., 2002; Gillespie & Pearlmutter, 2011).

Method

Participants.   Forty native speakers of German (mean age: 22 years, range: 18–42, 31 females, 39 right-handed) were recruited from the University of Potsdam community. Forty native speakers of Russian were recruited from the same community. All participants stated to have learned German after the age of 6 and to have been living in Germany for at least a year. One participant was excluded due to low accuracy in the filler trials. The remaining 39 participants were entered into the analysis (mean age: 27 years, range: 20–43, 31 females, 37 right-handed). To gauge their level of proficiency in German, participants completed the web-based Goethe Institute Placement Test (Goethe Institute, 2010). Their scores corresponded to an advanced C1–C2 level (mean = 88%; range: 73%–100%). In this and all following experiments, participants provided informed consent and received either course credit or financial compensation for their participation. All procedures were in accordance with the Declaration of Helsinki.

Stimuli.   Stimuli consisted of 50 preamble sets arranged in five conditions. Each preamble contained a singular noun (the head of the subject phrase) and a double modifier. In half of the preambles, the double modifier was a double genitive construction as shown in (2). As it was not possible to use double genitive constructions for all materials, the other half of preambles consisted of a prepositional phrase modified by a genitive phrase, as shown in (3). The nouns within the modifiers are referred to as the second and third noun, because the head noun was the first noun in the subject phrase.

All preambles were syntactically unambiguous, such that the first modifier could only modify the subject head and the second modifier could only modify the first modifier. This contrasts with English, where (3d,e) would be ambiguous between these two parses: [*the pen* [*next to* [*the letters and the diary*]]] and [*the pen* [*next to the letters*]] *and* [*the diary*]. In German, the use of oblique case marking made only the first parse available. Items were divided into phrases for their on-screen display, as indicated in (2–3). The segments containing plural attractors are indicated in boldface.

(2) a. Der Geruch / des Stalls$_{.gen.sg}$ / des Landwirts$_{.gen.sg}$
        The smell of the stable of the farmer
     b. Der Geruch / **der Ställe$_{.gen.pl}$** / des Landwirts$_{.gen.sg}$
        *The smell of **the stables** of the farmer*
     c. Der Geruch / des Stalls$_{.gen.sg}$ / **der Landwirte$_{.gen.pl}$**
        *The smell of the stable of **the farmers***
     d. Der Geruch / **der Ställe$_{.gen.pl}$** / und des Landwirts$_{.gen.sg}$
        *The smell of **the stables** <u>and</u> the farmer*
     e. Der Geruch / des Stalls$_{.gen.sg}$ / und **der Landwirte$_{.gen.pl}$**
        *The smell of the stable <u>and</u> **the farmers***

(3) a. Der Stift / neben dem Brief$_{.dat.sg}$ / des Pastors$_{.gen.sg}$
        The pen next to the letter of the priest
     b. Der Stift / **neben den Briefen$_{.dat.pl}$** / des Pastors$_{.gen.sg}$
        *The pen next to **the letters** of the priest*
     c. Der Stift / neben dem Brief$_{.dat.sg}$ / **der Pastoren$_{.gen.pl}$**
        *The pen next to the letter of **the priests***
     d. Der Stift / **neben den Briefen$_{.dat.pl}$** / und dem Tagebuch$_{.dat.sg}$
        *The pen next to **the letters** <u>and</u> the diary*
     e. Der Stift / neben dem Brief$_{.dat.sg}$ / und **den Tagebüchern$_{.dat.pl}$**
        *The pen next to the letter <u>and</u> **the diaries***

We varied the plurality of the second and third nouns and whether they appeared in embedded or coordinated constructions. In some cases, keeping the third noun constant across constructions rendered the coordinated conditions unnatural, as judged by a native speaker. In these cases, the third noun in the coordinated constructions was changed ($n = 40$ out of 50). When possible, we selected masculine modifier nouns to ensure that they had distinct singular and plural forms in the nominative and genitive cases (97 out of 100). The plural forms were regular and salient (e.g., *–n* and *–s* morphemes).

Table 1. *Mean ratings in the plausibility and semantic integration norming tasks*

	Plausibility Rating			Semantic Integration Rating	
Condition	Embedded	Coordinated	*Noun pair*	Embedded	Coordinated
Baseline	4.58 (0.78)	–	N1-N2	3.58 (1.32)	3.59 (1.22)
2nd pl.	4.24 (1.04)	3.87 (1.23)	N1-N3	2.57 (1.23)	3.07 (1.29)
3rd pl.	4.08 (1.08)	3.79 (1.33)	N2-N3	3.90 (1.22)	2.59 (1.19)

Note: The plausibility task used a 1 ('very implausible') to 5 ('very plausible') rating scale. The semantic integration task used a 1 ('loosely linked') to 5 ('tightly linked') rating scale. The semantic integration ratings between the second and third noun (N2-N3) are displayed for completeness but were not included in the analyses due to their lack of relevance for the scope of planning account, which focuses on the relationship between each modifier and the head noun, and not on the relationship between the modifiers themselves (Gillespie & Pearlmutter, 2011; 2013).

The head noun in the preambles was singular, such that the correct answer was always a singular verb. To ensure a 1:1 ratio of plural-to-singular target responses, we constructed 50 filler preambles where the correct response was a plural verb. Of these, 24 fillers consisted of double modifier constructions with a structure similar to the experimental preambles (e.g., *Die Geschenke der Eltern mit viel Geld*, "the gifts of the parents with lots of money"), 13 consisted of plural head nouns with a single modifier (e.g., *Die Flaggen in der Stadt*, "the flags in the city"), and 13 consisted of coordinated noun phrases (e.g., *Die Mutter und die Tochter*, "the mother and the daughter"). In addition, to encourage participants to read for comprehension, we created 16 short sentences followed by *yes/no* comprehension questions, which were interspersed with the experimental preambles (e.g., *The grandmother plays in the afternoon with the young child*. Question: *Is the child young?*).

Experimental stimuli were normed for plausibility and semantic integration. In the plausibility norming task, participants ($n = 15$ German native speakers, mean age = 22 years, age range = 19–27, 13 females, 14 right-handed) were asked to rate the preambles using a scale from 1 (*very implausible*) to 5 (*very plausible*). In the semantic integration task, participants (28 German native speakers, mean age: 24 years, range: 19–39, 25 females, 27 right-handed) rated the semantic integration of the two capitalized nouns in the singular conditions of each preamble, using a scale from 1 (*loosely linked*) to 5 (*tightly linked*). The instructions from Solomon and Pearlmutter (2004) were adapted to German. They included example phrases such as *the KETCHUP or the MUSTARD* and *the BRACELET made of GOLD*. They stated that although *ketchup* and *mustard* were similar in meaning, they did not share a specific relationship in the example, in contrast to *bracelet* and *gold*, which were closely related as the phrasing made it clear that the bracelet contained gold.

Table 1 shows the mean ratings from the norming tasks. In the plausibility task, preambles with plural modifiers were rated as less plausible than the singular condition ($\hat{\beta} = -0.592$, $SE = 0.120$, $t = -4.962$, $p = .000$). After verifying this overall difference in plausibility between singular and plural conditions, we removed the

baseline condition and analyzed the remaining four conditions using a 2×2 model with plural noun position (2nd vs. 3rd), construction type (embedded vs. coordinated), and their interaction as fixed effects. The coordinated constructions were rated as less plausible than the embedded constructions overall ($\hat{\beta} = -0.328$, $SE = 0.149$, $t = -2.896$, $p = .007$), but there were no differences in plausibility between second and third plural noun preambles within each type of construction (*ns* Plural Noun Position $\times$ Construction Type: $\hat{\beta} = -0.075$, $SE = 0.149$, $t = -0.501$, $p = .617$).

The semantic integration task was analyzed using a model with Pair (N1–N2 vs. N1–N3), construction type (embedded vs. coordinated), and their interaction as fixed effects. N1–N2 pairs were rated as more closely integrated than N1–N3 pairs overall ($\hat{\beta} = -0.773$, $SE = 0.130$, $t = -5.943$, $p = .000$). However, there was also a significant Pair $\times$ Construction Type interaction, because the difference between N1–N2 and N1–N3 ratings was larger in embedded than in coordinated constructions ($\hat{\beta} = 0.473$, $SE = 0.135$, $t = 3.513$, $p = .000$).

In sum, the norming tasks revealed that: coordinated preambles were considered less plausible than embedded preambles; and the second noun was perceived as more tightly integrated with the head noun than the third noun. The latter pattern was stronger in embedded than in coordinated conditions. This shows that in the embedded conditions, second nouns were both syntactically closer and semantically more integrated with the head noun than third nouns, as compared with the coordinated conditions.

Procedure. We used a speeded forced-choice task based on Staub (2009, 2010). Sentence preambles were presented phrase-by-phrase in the center of the screen for 700 ms. Following the last phrase, a question mark appeared in the center of the screen with two response options at the bottom in uppercase letters: *IST* ("is") and *SIND* ("are"). These two options were used for both experimental and filler preambles. Singular verbs were always shown on the right of the screen and plural verbs on the left. Participants were instructed to select the verb that provided a grammatical continuation to the sentence by pressing either the F or J key on the keyboard. In comprehension trials, the F and J keys were used for "yes" and "no," respectively. After each response, participants pressed the spacebar to begin the next trial. No feedback was provided.

Sentence preambles were distributed across five lists in a Latin-square design. Experimental and filler preambles were randomized on a by-participant basis. Each experimental session began with three practice trials. The task instructions emphasized the importance of both accuracy and speed. The experiment was run on a laptop PC using the Linger software (Rohde, MIT). For the native German speakers, each session lasted approximately 45 min.

For the Russian native speakers, the procedure was similar with a few exceptions. First, participants performed a vocabulary test before the speeded task. They were shown a booklet with the nouns that would later appear in the experimental materials. Each noun was presented with a definite article, a picture, and a Russian translation (e.g., *der Geruch*, *запах*). Participants identified unfamiliar nouns, and items containing these nouns were later excluded from analysis on a by-participant basis.

Second, participants performed an untimed choice task after the speeded task. The goal of the task was to ensure that they understood German agreement and case constraints, so that any effects in the speeded task could not be attributed to incomplete grammatical knowledge. Participants read 20 sentences at their own pace and were asked to circle the option that was grammatical in the context of the sentence: 10 sentences probed for knowledge of genitive and dative case (e.g., *Alle Besucher wollen {dem Baby / des Babys} einen Kuss geben*, "All visitors want to give the baby$_{DAT/*GEN}$ a kiss") and 10 sentences probed for knowledge of subject–verb agreement (e.g., *Manchmal {spielen / spielt} die Kinder im Garten*, "Sometimes the children play/*plays in the garden"). Each session lasted approximately 1 hr.

Analysis. We analyzed accuracy and response times for correctly answered trials. To ensure that the analysis only included participants who were able to perform proficiently in the task, participants with accuracy lower than 60% in the filler trials were excluded. In addition, responses longer than 4000 ms or shorter than 200 ms were removed (Staub, 2010). To estimate the appropriate transformation for the response time data we used the Box–Cox procedure (Box & Cox, 1964). This procedure yielded similar results across all experiments, with the optimal value of the λ coefficient clustering around 0 (range: –0.30–0.02). Therefore, response times were log transformed. Accuracy was analyzed with mixed-effects logistic regression (Jaeger, 2008) using the bobyqa optimizer (Powell, 2009). Response times were analyzed with mixed linear models (Baayen, Davidson, & Bates, 2008). Analyses were performed with R, an open-source programming language and environment for statistical computing (R Development Core Team, 2017).

The statistical analysis was performed in two stages. In the first stage, we verified that our manipulation had successfully elicited attraction errors by comparing the conditions with plural nouns against the baseline condition (i.e., main effect of attractor: [1a] vs. [1b–e]). In the second stage, we removed the baseline condition and analyzed the remaining four conditions with a 2 × 2 model with plural noun position (2nd vs. 3rd), construction type (embedded vs. coordinated), and their interaction as fixed effects. Fixed effects were coded using orthogonal contrasts: for the plural noun position factor, the mean of the second plural noun conditions was compared with the mean of the third plural noun conditions. For the construction type factor, the mean of the coordinated conditions was compared with the mean of the embedded conditions. In addition, the Goethe test scores of Russian participants were centered and added as a predictor in the analyses to examine whether their German proficiency modulated their error rates or response times. These scores did not turn out to be significant predictors in most of the experiments reported below, but they are reported in the tables for completeness.

For the random effects structure of the model, we followed current guidelines in psycholinguistics, and we initially constructed a maximal model that included random intercepts and slopes for all fixed effects and their interactions (Barr, Levy, Scheepers, & Tily, 2013). When this maximal model failed to converge, we gradually simplified the random effects structure following the suggestions by Barr et al. We report the results from the model with the maximal random effects structure that converged and that did not contain correlations between the random effects

with an absolute value of 1 (Baayen et al., 2008). Unless noted below, all models included uncorrelated by-subject and by-item random intercepts and slopes for all fixed effects. We report effect sizes using the model estimates ($\hat{\beta}$), standard errors (*SE*) and *t* and *z* statistics. *P* values were computed using Satterthwaite's approximation for denominator degrees of freedom (Kuznetsova, Bruun Brockhoff, & Haubo Bojesen Christensen, 2014). Data for this and following experiments, as well as experimental materials, can be found at the Center for Open Science Framework website (https://osf.io/).

Results of Experiment 1

Data exclusion. In the online task, native and L2 speakers were highly accurate in the filler trials (92% and 90%, respectively), and no participants were excluded due to low accuracy in the fillers. Participants also performed accurately in the comprehension questions (93% and 90%). The exclusion of excessively long or short responses removed 1.6% of experimental trials for native speakers and 3% of experimental trials for L2 speakers. For L2 speakers, the exclusion of items with unknown words removed a further 3.6% of trials.[2] In the untimed test, L2 speakers performed near ceiling (mean = 99.4%, range: 90%–100%), suggesting that they knew German agreement constraints and could use the genitive and dative cases.

Group analyses. In order to compare native and L2 speakers, we examined whether there were main effects or interactions between group (German vs. Russian–German) and the predictors entered for the first- and second-stage analyses. In the first-stage analysis, which verified the existence of attraction, accuracy measures showed a main effect of attraction ($\hat{\beta} = -2.923$, $SE = 1.018$, $z = -2.871$, $p = .004$) but no interaction with group. Response time measures showed a main effect of attraction ($\hat{\beta} = -0.115$, $SE = 0.02$, $t = -5.354$, $p = .000$) and a main effect of group ($\hat{\beta} = 0.398$, $SE = 0.066$, $t = 6.008$, $p = .000$), with longer response times for L2 than native speakers. These factors did not interact, suggesting similar attraction rates in both groups.

In the second-stage analyses, which addressed whether the attraction effect was modulated by the position of the plural noun or the type of construction, accuracy measures showed a three-way interaction between group and these factors ($\hat{\beta} = -1.449$, $SE = 0.707$, $z = -2.050$, $p = .040$), suggesting that German and Russian–German speakers differed in their response to the experimental manipulations. There were also main effects of plural noun position ($\hat{\beta} = 0.660$, $SE = 0.220$, $z = 5.893$, $p = .004$) and construction type ($\hat{\beta} = 1.167$, $SE = 0.281$, $z = 4.150$, $p = .000$). Response time measures showed an effect of group ($\hat{\beta} = 0.398$, $SE = 0.072$, $t = 5.495$, $p = .000$), with longer response times for L2 than native speakers, and an effect of construction type ($\hat{\beta} = 0.108$, $SE = 0.020$, $t = 5.193$, $p = .006$).

The results of each group are presented below. Table 2 shows the mean percentage of agreement errors and the averaged correct response times by condition for each speaker group. Figure 1 displays the mean percentage of agreement errors by condition. Table 3 shows the results of the statistical analyses. Pairwise comparisons are reported in the text. The figures and text display the experimental

Table 2. *Mean agreement error rates and response times of German native speakers and Russian-German speakers*

	German Native Speakers		Russian-German Speakers	
	Error Rate [CI] (%)	RT [SE] (ms)	Error Rate [CI] (%)	RT [SE] (ms)
Experiment 1				
Baseline	2 [1,4]	704 [21]	2 [1,4]	1056 [31]
Embedded, 2 pl.	4 [3,7]	792 [28]	7 [5,10]	1103 [33]
Embedded, 3 pl.	2 [1,5]	730 [22]	2 [1,5]	1150 [37]
Coordinated, 2 pl.	12 [9,16]	868 [33]	14 [11,19]	1202 [45]
Coordinated, 3 pl.	6 [4,9]	850 [28]	14 [11,18]	1277 [44]
Experiment 2				
Baseline	3 [2,4]	696 [13]	3 [2,5]	713 [13]
Embedded, 2 pl.	9 [7,11]	760 [16]	5 [3,7]	753 [14]
Embedded, 3 pl.	2 [1,4]	779 [16]	3 [1,4]	754 [14]
Coordinated, 2 pl.	16 [13,19]	756 [17]	9 [7,12]	783 [16]
Coordinated, 3 pl.	10 [8, 13]	809 [18]	11 [9,15]	799 [18]

Note: Mean percentages are provided with 95% binomial confidence intervals, and mean response time with standard errors.
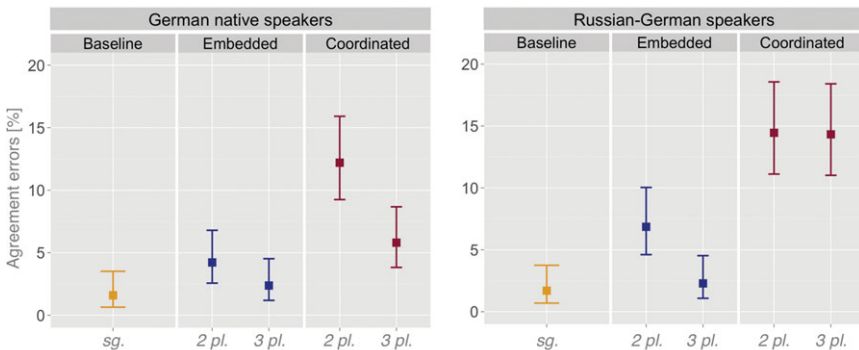

Figure 1. (Color online) Attraction errors in Experiment 1. Squares represent the mean percentage of agreement errors, and bars correspond to 95% binomial confidence intervals.

effects in percentages for easier interpretability, but the accuracy analyses were always performed on log odds, such that differences between conditions may look different in percentages than in log odds.

German native speakers. Accuracy measures showed a marginal attraction effect, with preambles with a plural noun eliciting more agreement errors than the baseline condition ($\hat{\beta} = 2.491$, $SE = 1.381$, $z = 1.804$, $p = .071$). There were also

Table 3. *Model results of Experiment 1*

Experiment 1	Accuracy				Response Times			
	$\hat{\beta}$	SE	z	p	$\hat{\beta}$	SE	t	p
German speakers								
Pl Noun Position	**0.858**	**0.425**	**2.017**	**.044**	0.029	0.025	1.169	.243
Construction type	**1.188**	**0.500**	**2.375**	**.018**	**0.111**	**0.025**	**4.435**	**.000**
Pl Noun × Const. type	0.237	0.529	0.448	.655	−0.059	0.050	−1.182	.237
Russian-German speakers								
German proficiency	1.543	2.065	0.747	.455	−0.890	0.609	−1.461	.152
Pl Noun Position	**0.622**	**0.304**	**2.049**	**.041**	−0.040	0.032	−1.282	.205
Construction type	**1.567**	**0.342**	**4.576**	**.000**	**0.089**	**0.036**	**2.455**	**.019**
Pl Noun × Const. type	**−1.211**	**0.488**	**−2.484**	**.013**	−0.029	0.057	−0.512	.608

Note: Model estimates ($\hat{\beta}$) are expressed in log odds for accuracy and log milliseconds for response times. For the PLURAL NOUN POSITION factor, a positive estimate means that 2nd plural nouns elicited more agreement errors (or longer response times) than 3rd plural nouns. For the CONSTRUCTION TYPE factor, a positive estimate means that the coordinated conditions elicited more agreement errors (or longer response times) than the embedded conditions. Significant effects at the $\alpha = .05$ level are bolded.

main effects of plural noun position and construction type: participants made more errors with second than third plural nouns and with coordinated than embedded constructions. However, there was no interaction between noun position and construction type, suggesting that the 2nd-3rd-noun asymmetry affected embedded and coordinated preambles similarly.

The response times of correctly answered trials showed an attraction effect, with plural noun preambles eliciting longer latencies than the baseline condition ($\hat{\beta} = 0.103$, $SE = 0.030$, $t = 3.349$, $p = .001$). Further, coordinated constructions elicited longer latencies than embedded constructions.[3]

Russian–German speakers. Accuracy measures showed an attraction effect, with preambles with a plural noun eliciting more agreement errors than the baseline condition ($\hat{\beta} = 1.742$, $SE = 0.557$, $t = 3.12$, $p = .002$). Participants also made more agreement errors with second than third plural nouns and with coordinated than embedded constructions. There was an interaction between noun position and construction type, showing that the 2nd-3rd-noun asymmetry affected the embedded and coordinated constructions differently. Pairwise comparisons showed that participants made more errors for second than third plural nouns in embedded constructions ($\hat{\beta} = 1.161$, $SE = 0.422$, $z = 2.754$, $p = .006$), but there was no difference in coordinated constructions ($\hat{\beta} = -0.017$, $SE = 0.233$, $z = -0.073$, $p = .942$).

The response times of correctly answered trials showed an attraction effect, with plural noun preambles eliciting longer latencies than the baseline condition ($\hat{\beta} = 0.095$, $SE = 0.031$, $t = 3.032$, $p = .003$). In addition, responses in coordinated constructions showed longer latencies than in embedded constructions.

Discussion of Experiment 1

Experiment 1 tested whether German and Russian–German speakers made attraction errors and whether they displayed a 2nd-3rd-noun asymmetry with double modifier constructions, as was previously found in English and French (Franck et al., 2002; Gillespie & Pearlmutter, 2011). Both groups showed attraction, making more errors when the sentence preambles contained a plural attractor as compared with the baseline condition. In addition, the response times of correctly answered trials were longer when the preambles contained a plural attractor, which suggests that verb choices in these conditions were more difficult. Attraction rates did not significantly differ between German and Russian–German speakers, showing that native and nonnative speakers were similarly prone to attraction errors.

A second similarity between the groups is that they made more agreement errors and showed longer response times in coordinated than in embedded preambles. This pattern was unpredicted, and it could be due to several reasons. First, coordinated preambles may have elicited more plural verb choices because they contained more cues to plurality, due to the lexically plural conjunction *and*, and to the fact that the two modifiers together formed a conceptually plural entity. A second possibility is that the pattern was due to lexical differences between embedded and coordinated constructions. As mentioned before, in order to ensure that all preambles were pragmatically felicitous, the third noun was often changed across constructions. This change introduced lexical differences that decreased the plausibility and semantic integration of coordinated preambles, as shown in the norming tasks. Therefore, differences in either plausibility or semantic integration may have made coordinated preambles harder to process, giving rise to more agreement errors.

The main difference between German and Russian–German speakers is that bilingual speakers made more agreement errors with second than third plural nouns in embedded but not in coordinated conditions, whereas German speakers showed an overall 2nd-3rd-noun asymmetry, which did not differ across constructions. These results show that the linear distance between the modifiers and the verb cannot solely modulate L2 agreement errors. First, because linear distance was identical between embedded and coordinated constructions. Second, because the third noun never elicited more agreement errors, despite being linearly closer to the verb. Thus, these patterns suggest that when computing agreement, L2 speakers weighed the grammatical distance between then modifiers more strongly than their linear distance to the verb.

In contrast to the bilinguals, native German speakers did not show a stronger 2nd-3rd-noun asymmetry in embedded than coordinated constructions. This was unexpected, especially given previous findings that native processing is more affected by grammatical than linear distance (Franck et al., 2002; Gillespie & Pearlmutter, 2011; Solomon & Pearlmutter, 2004; Vigliocco & Nicol, 1998), and is

difficult to account for both under a syntactic (Franck et al., 2002) and a scope of planning account (Gillespie & Pearlmutter, 2011). Under a syntactic account, the coordination of the second and third nouns should have made them syntactically more symmetric, thus reducing the 2nd-3rd-noun asymmetry in the coordinated conditions. Under a scope of planning account, the integration asymmetry was larger for embedded than for coordinated preambles, as verified in the integration norming task. However, the reduced syntactic/semantic distance between the modifiers in the coordinated constructions did not lead to any measurable reduction of the 2nd-3rd-noun asymmetry.

However, one concern about the results is that the error rates of native speakers were extremely low, as suggested by the marginal effect of agreement attraction in accuracy measures. This was specially the case in embedded conditions, which elicited only 2%–4% error rates. These rates contrast with previous studies that used a forced-choice paradigm, where attraction errors ranged between 15%–27% (Staub, 2009) and 23% (Staub, 2010). Thus, it is possible that the lack of a stronger 2nd-3rd-noun asymmetry in the embedded conditions was due to a floor effect, such that errors were not frequent enough to show any modulations.

A possible reason for the low errors rates is that the task may have been too easy for native speakers, since it used a phrase-by-phrase display with a generous presentation rate (700 ms). By contrast, previous forced-choice studies used a word-by-word presentation with a shorter rate (400 ms; Staub, 2009, 2010). Our phrase-by-phrase presentation may have made the task too easy for native speakers, giving them more time to process the preambles and to overcome attraction effects. This is less of a concern for bilinguals, for whom processing a nonnative language might already be more difficult than for native speakers. To address the possibility of floor effects, we conducted a follow-up study where we attempted to increase the rate of agreement errors by making the forced-choice task more difficult.

EXPERIMENT 2

Experiment 2 aimed to foster agreement errors by making the forced-choice task more difficult. The procedure was modified as follows. Sentence preambles were presented word-by-word. We reasoned that this presentation mode would require participants to structurally decompose the preambles on their own, in contrast with Experiment 1, where the preambles appeared already split into phrases, making it easier to assign them syntactic and prosodic structure. In addition, we increased the presentation speed to 400 ms per word. Participants were also given a response deadline that required them to respond within 2 s, or otherwise the message *Zu langsam!* ("Too slow!") appeared in red letters. Finally, trials with comprehension questions were removed, as a pilot study showed that they were difficult to answer within the new response deadline.

Method

Participants.   Sixty-two German native speakers were recruited remotely via the web by advertising the experiment on social media and student groups affiliated with German universities. Three participants were excluded from analysis: 1 due to

dyslexia and 2 due to low accuracy in the filler trials. The remaining 59 participants were entered in the analysis (mean age: 28 years, range: 19–52, 38 females, 53 right-handed).

Sixty-three Russian native speakers were recruited remotely via the web by advertising the experiment on social media and groups for Russian communities in Germany. All participants stated to have learned German after the age of 6 and to have been living in Germany for at least a year. One participant was excluded from analysis due to low accuracy in the filler trials. The remaining 62 participants were entered into the analysis (mean age: 27 years, range: 18–42, 42 females, 60 right-handed). All participants completed the Goethe test with a range of scores that corresponds to an intermediate to advanced B2–C2 level (mean = 87%; range = 57%–100%).

Stimuli, procedure, and analysis. The stimuli were the same as in Experiment 1, but the procedure was modified as described above. The word-by-word presentation consisted of each word being displayed for 3000 ms with a 100 ms interstimulus interval, for a total stimulus onset asynchrony of 400 ms. The experiment was run remotely on a web-based platform using the Ibex Farm software (Drummond; http://spellout.net/ibexfarm). Web-based testing was used because it allowed us to expand the number of participants, and because this method has been found to yield reliable results in previous psycholinguistics studies (Chemla, Cummins, & Singh, 2016; Dillon, Clifton, & Frazier, 2014; Enochson, & Culbertson, 2014; Gibson, Piantadosi, & Fedorenko, 2011; Sprouse, 2011; Wagers & Phillips, 2014). For German native speakers, each session lasted approximately 30 min. For Russian native speakers, the only difference with Experiment 1 was that the nouns in the vocabulary test appeared without the pictures and Russian glosses. Each session lasted approximately 45 min.

The analysis was identical to Experiment 1, except that only responses shorter than 200 ms were removed, as the 2-s response deadline already ensured the exclusion of excessively long responses.

Results of Experiment 2

Data exclusion. In the online task, native and L2 speakers were highly accurate in the filler trials (88% and 90%, respectively). Two native speakers and one L2 speaker were removed due to low accuracy in the filler trials. Time-outs affected 3.54% of experimental trials of native speakers and 2.95% of trials of L2 speakers. The exclusion of excessively short responses removed 0.68% and 2.6% of trials, respectively. For L2 speakers, the exclusion of items with unknown words removed a further 9.6% of trials. In the untimed test, L2 speakers performed near ceiling (mean = 99.5%, range = 90%–100%), showing that they knew German agreement and case constraints.

Group analyses. These analyses examined whether there were main effects or interactions between group and the predictors entered for the first- and second-stage analyses. In the first-stage analysis, accuracy measures showed a main effect

Table 4. *Model results of Experiment 2*

	Accuracy				Response Times			
Experiment 2	$\hat{\beta}$	SE	z	p	$\hat{\beta}$	SE	t	p
German speakers								
Pl Noun Position	**1.294**	**0.386**	**3.356**	**.001**	−0.034	0.022	−1.542	.128
Construction type	**1.351**	**0.364**	**3.708**	**.000**	0.030	0.020	1.461	.150
Pl Noun × Const. type	**−1.411**	**0.487**	**−2.898**	**.004**	−0.042	0.033	−1.240	.215
Russian-German speakers								
German proficiency	−0.725	0.848	−0.855	.393	0.117	0.136	0.860	.393
Pl Noun Position	0.469	0.352	1.334	.182	−0.008	0.021	−0.360	.720
Construction type	**0.946**	**0.251**	**3.770**	**.000**	0.044	0.022	1.966	.056
Pl Noun × Const. type	**−1.050**	**0.417**	**−2.520**	**.012**	−0.025	0.033	−0.759	.448

Note: Model estimates ($\hat{\beta}$) are expressed in log odds for accuracy and log milliseconds for response times. For the PLURAL NOUN POSITION factor, a positive estimate means that 2nd plural nouns elicited more agreement errors (or longer response times) than 3rd plural nouns. For the CONSTRUCTION TYPE factor, a positive estimate means that the coordinated conditions elicited more agreement errors (or longer response times) than the embedded conditions. Significant effects at the $\alpha = .05$ level are bolded.

of attraction ($\hat{\beta} = -1.706$, $SE = 0.490$, $z = -3.482$, $p = .000$) but no interaction with group. Response time measures showed a main effect of attraction ($\hat{\beta} = -0.086$, $SE = 0.013$, $t = -6.359$, $p = .000$) and no interaction. These results supported similar attraction effects across groups.

In the second-stage analyses, accuracy measures showed significant effects of plural noun position ($\hat{\beta} = 0.930$, $SE = 0.240$, $z = 3.883$, $p = .000$), construction type ($\hat{\beta} = 1.044$, $SE = 0.237$, $z = 4.404$, $p = .000$), and a two-way interaction between them ($\hat{\beta} = -0.937$, $SE = 0.347$, $z = -2.704$, $p = .007$). In contrast with Experiment 1, the interaction between group and the other two factors was not significant, suggesting that German and Russian–German speakers were similarly affected by the experimental manipulations. However, the two-way interaction showed that across constructions, the effect of plural noun position was significant for German speakers ($\hat{\beta} = 0.923$, $SE = 0.318$, $z = 2.904$, $p = .004$), but not for Russian speakers ($\hat{\beta} = 0.322$, $SE = 0.318$, $z = 1.015$, $p = .310$). Response time measures showed an effect of construction type ($\hat{\beta} = 0.040$, $SE = 0.014$, $t = 2.848$, $p = .006$). Table 4 shows the results of the statistical analyses. Figure 2 displays the mean percentage of agreement errors by condition.

German native speakers. Accuracy measures showed an attraction effect, with preambles with a plural noun eliciting more agreement errors than the baseline condition ($\hat{\beta} = 2.109$, $SE = 0.675$, $z = 2.992$, $p = .003$).[4] As in Experiment 1,
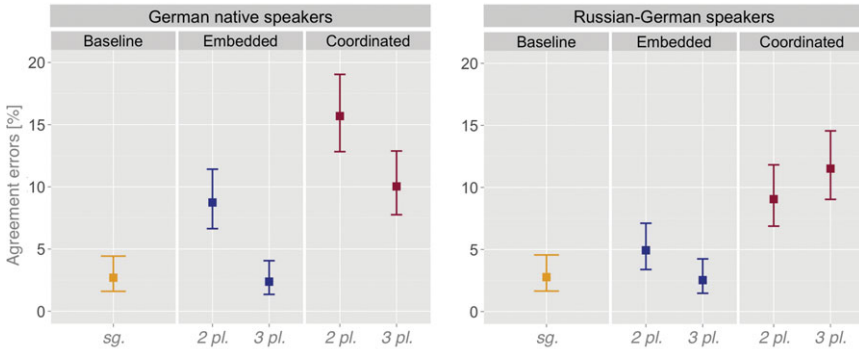
Figure 2. (Color online) Attraction errors in Experiment 2. Squares represent the mean percentage of agreement errors, and bars correspond to 95% binomial confidence intervals.

participants made more errors with second than with third plural nouns, and they made more errors with coordinated than with embedded constructions. However, and in contrast with Experiment 1, there was a significant interaction between noun position and construction type, which suggests that the 2nd-3rd-noun asymmetry affected the embedded and coordinated constructions differently. Pairwise comparisons showed that the 2nd-3rd-noun asymmetry was almost twice as strong in the embedded ($\hat{\beta} = 1.515$, $SE = 0.337$, $z = 4.500$, $p = .000$) than in the coordinated ($\hat{\beta} = 0.606$, $SE = 0.214$, $z = 2.836$, $p = .005$) constructions.

The response times of correctly answered trials showed an attraction effect, with plural noun preambles eliciting longer latencies than the baseline condition ($\hat{\beta} = 0.101$, $SE = 0.020$, $t = 4.989$, $p = .000$). In addition, second plural noun preambles elicited shorter latencies than third plural noun preambles.

Russian–German speakers. Accuracy measures showed an attraction effect, with preambles with a plural noun eliciting more agreement errors than the baseline condition ($\hat{\beta} = 0.943$, $SE = 0.332$, $z = 2.840$, $p = .005$). There were also more errors in coordinated than in embedded constructions. There was an interaction between noun position and construction type, which was due to the fact that the 2nd-3rd-noun asymmetry marginally affected the embedded constructions ($\hat{\beta} = 1.892$, $SE = 1.0105$, $z = 1.873$, $p = .061$), but there was no difference in the coordinated conditions ($\hat{\beta} = 0.140$, $SE = 0.4337$, $z = 0.323$, $p = .747$).

The response times of correctly answered trials showed an attraction effect, with plural noun preambles eliciting longer latencies than the baseline condition ($\hat{\beta} = 0.066$, $SE = 0.020$, $t = 3.391$, $p = .001$).

Discussion of Experiment 2

In order to foster agreement errors, Experiment 2 increased the difficulty of the forced-choice task by using a fast word-by-word presentation rate and a

response deadline designed to put participants under time pressure. The error rates of native speakers in the embedded conditions doubled as compared to Experiment 1. Further, we observed a significant interaction between noun position and construction type, which did not interact with the group factor, showing that both native and bilingual speakers displayed a stronger 2nd-3rd-noun asymmetry in embedded than in coordinated preambles. As in Experiment 1, bilinguals were not more prone to agreement attraction than native speakers.

The presence of a stronger 2nd-3rd-noun asymmetry in embedded than in coordinated conditions does not support the claim that the agreement errors of L2 speakers are more influenced by linear distance than native speakers, at least at high levels of proficiency. Otherwise, if linear distance to the verb was the main factor modulating attraction, there should have been more errors for third than second plural nouns, since third nouns were linearly closer to the verb. In addition, the number of attraction errors should have been similar for embedded and coordinated conditions, which had identical linear order. Instead, second plural nouns elicited more agreement errors and processing difficulty than third plural nouns in embedded conditions, with no difference in coordinated constructions. Therefore, our findings suggest that both native and bilingual speakers prioritize the grammatical relationship between the attractors and the head noun when computing agreement, as proposed by syntactic and semantic accounts (Franck et al., 2002; Gillespie & Pearlmutter, 2011).

Our results cannot distinguish between these accounts because our materials contained both syntactic and semantic asymmetries. However, as suggested by a reviewer, there is one property of our items that may help arbitrate between syntactic and semantic accounts. Recall that our preambles were of two types: in half, the double modifier was a double genitive construction, such as *the smell of the stable(s) of the farmer(s)* (henceforth "of-of items"). The other half consisted of a prepositional phrase modified by a genitive phrase, such as *the pen next to the letter(s) of the priest(s)* (henceforth "PP-of items"). Syntactically, the distance between the first and second modifiers was identical across both item types, as the first modifier was unambiguously attached to the subject head, whereas the second modifier (always a genitive) was unambiguously attached to the first modifier. However, the two item types differed in semantic integration, as shown by the norming study: in of-of items, the first modifier was rated as more semantically integrated to the subject head than in PP-of items, rendering the semantic 2nd-3rd-noun asymmetry higher for of-of than PP-of items (1.04 vs. 0.5).

To examine whether differences in semantic integration affected attraction errors above and beyond syntactic structure, we reanalyzed the data of Experiments 1 and 2 jointly using preposition type (of-of vs. PP-of) as a further predictor of participants' responses, in addition to noun position (2nd vs. 3rd) and construction type (embedded vs. coordinated). For both German and Russian–German speakers, attraction rates were higher in of-of than in PP-of items, as shown by a significant main effect of preposition type (German speakers: $\hat{\beta} = 0.846$, $SE = 0.276$, $z = 3.068$, $p = .002$; Russian–German speakers: $\hat{\beta} = 0.472$, $SE = 0.198$, $z = 2.382$, $p = .017$). The effect of preposition type did not interact with the other two experimental factors (German speakers: $\hat{\beta} = -0.907$, $SE = 0.664$, $z = -1.364$; $p = .176$; Russian–German speakers: $\hat{\beta} = 0.097$, $SE = 0.629$, $z = 0.155$, $p = .877$).

Overall, these analyses suggest that native and bilingual speakers made more attraction errors with of-of items, where the second noun was more integrated with the head than PP-of items. Although these analyses should be taken with care because they were done post hoc, they suggest that semantic distance may modulate attraction errors above and beyond syntactic distance.

Finally, both speaker groups made more errors in coordinated than in embedded constructions, similarly to Experiment 1. As explained before, coordinated constructions contained an additional cue to plurality (the conjunction *and*) and due to the use of different nouns, they further differed in plausibility and semantic integration from the embedded constructions. Any of these variables may have resulted in the overall differences between constructions. A way to differentiate between these variables would be to include an additional baseline coordinated condition (e.g., *the smell of the stable <u>and</u> the farmer*): the comparison between embedded and coordinated baselines could address whether the use of the conjunction itself contributed to the production of errors. We did not include this additional baseline because we worried that, given participants' low error rates, adding one extra condition to a five-condition design would reduce experimental power. We acknowledge this is a limitation of the present design, and we believe that further research will be needed to address this issue.

GENERAL DISCUSSION

Our study examined how native and L2 speakers of German produced agreement attraction errors with double modifier constructions. We addressed two questions. First, we asked whether L2 speakers were more prone to attraction than native speakers, consistent with accounts where L2 grammatical difficulties are attributed to working memory limitations (McDonald, 2006; Sagarra & Herschensohn, 2013) and recently, to increased susceptibility to interference (Cunnings, 2016). Second, given previous findings of L2 difficulties with agreement in nonadjacent configurations (Foote, 2011; Keating, 2009, 2010), we asked whether L2 attraction errors were differentially affected by grammatical and linear distance, as previously found in native speakers.

Our study focused on proficient L2 speakers to ensure that attraction errors were not due to insufficient knowledge of German agreement constraints. In untimed measures, L2 speakers demonstrated nativelike knowledge of agreement. However, they made attraction errors in the forced-choice task, consistent with previous production work (Foote, 2010; Hoshino et al., 2010; Nicol & Greth, 2003). Similarly to some of these studies, our L2 population involved advanced L2 speakers (Foote, 2010; Hoshino et al., 2010) who were in an immersion situation at the time of testing (Jegerski, 2016).

When compared directly, native and proficient L2 speakers erred at similar rates: attraction rates in both groups always ranged between 0 and 20% and did not vary significantly between groups. Under a view where attraction results from interference during memory retrieval, these findings suggest that L2 difficulties with agreement are unlikely to result from increased susceptibility to interference, as recently proposed by Cunnings (2016). However, our results can only speak to the processing of subject–verb agreement, so a higher predisposition to interference

might still account for L1–L2 differences in the processing of other grammatical dependencies, such as pronouns, reflexives, and negative polarity items. To address this issue, further research comparing the amount of interference in native and nonnative speakers is needed.

Our second question was whether L2 speakers weighted linguistic and nonlinguistic information differently in the computation of agreement. We addressed this question by manipulating the distance between the attractor nouns and the verb, and whether this distance was linear or also linguistic (syntactic or semantic) in nature. Jointly, our results showed that both native and L2 speakers prioritized linguistic over linear distance. In embedded constructions such as *the smell of the stable(s) of the farmer(s)*, both participant groups displayed a 2nd-3rd-noun asymmetry and made more errors for second than third plural nouns. However, the 2nd-3rd-noun asymmetry was reduced in coordinated conditions, which had identical linear order to the embedded conditions but a smaller syntactic and semantic distance between the two modifiers. In coordinated constructions, the 2nd-3rd-noun asymmetry was reduced in native speakers (Experiment 2) and absent in L2 speakers (Experiments 1 and 2). For native speakers, these results replicate previous work on English and French (Franck et al., 2002; Gillespie & Pearlmutter, 2011). For L2 speakers, these results suggest that, similarly to native speakers, their agreement computations are more influenced by modifiers that are grammatically close to the subject head than modifiers that are linearly close to the verb.

With regard to previous work on linear distance effects (Foote, 2011; Keating, 2009, 2010), Keating argued that L2 speakers were less sensitive to agreement violations in nonadjacent configurations because they were more likely to engage in shallow processing than native speakers. Our results do not invalidate Keating's finding, but they provide an important qualification to his claim: they show that if shallow parsing is understood as the prioritization of linear information over linguistic structure, this is unlikely to occur in L2 agreement processing. Instead, linear distance effects could be explained by proposing that, as more material intervenes between the subject head and the verb, L2 speakers are less likely to try to license the number of the verb through memory retrieval, thus avoiding engaging in agreement computations altogether. This would result in increased failure to detect agreement violations, without yielding stronger attraction effects, if the latter arise due to misretrieval.

An interesting question concerns how the present results relate to existing models of agreement production. There are very few production studies on cue-based retrieval (Badecker & Kuminiak, 2007; Lorimor et al., 2015), but we think that our account is compatible with this work. Like Badecker and Kuminiak (2007), we assume that the assignment of an appropriate number inflection to a verb will depend on the morphosyntactic features of the subject phrase, which was assembled prior to the verb. In order to license the verb, the subject phrase will need to be retrieved from working memory so that the appropriate features can be checked. During this process, nouns other than the subject phrase, such as the two attractors in the present study, might be misretrieved due to feature overlap with the subject head, thus eliciting attraction.

Note that this account does not assume that attraction errors in production occur only due to misretrieval. Errors might also originate during the encoding of the

subject phrase itself, as proposed by other accounts (Eberhard, Cutting, & Bock, 2005; Foote & Bock, 2012; Vigliocco & Nicol, 1998). Under these accounts, the planning of the verb is launched during the encoding of subject phrase, such that encoding errors result in the assignment of the wrong verb number. The semantic and morphological properties of the attractor nouns will affect encoding, either by decreasing the likelihood of attraction (when the attractors and subject head mismatch in grammatical features; e.g., Lorimor et al., 2015) or by increasing it (when the attractors and the subject head overlap in features). At present, we cannot determine whether the errors observed in our experiments were due to problems at encoding or retrieval. Instead, we think that both types of processes should be seen as relying on working memory, such that the grammatical properties of the attractors will modulate their relative levels of activation, thus making them more likely to interfere with the representation of the subject phrase (at encoding) or more likely to be wrongly recovered as agreement controllers (at retrieval).

To account for the 2nd-3rd-noun asymmetry, we suggest two alternatives. The first is to allow semantic features to differentially affect memory encoding. For instance, modifiers that are conceptually more integrated with the head noun might be more salient in memory. This increased activation should render them more likely to interfere in the planning of the verb or in the selection of the appropriate agreement controller at retrieval. Distinguishing between these possibilities will require the use of time-sensitive paradigms in order to measure whether processing disruptions associated with the 2nd-3rd-noun asymmetry arise during the reading of the subject preamble or during the selection of the verb.

Alternatively, the 2nd-3rd-noun asymmetry could arise due to the role of syntactic attachments in modulating activation levels (Vasishth & Lewis, 2006). Under current implementations of cue-based retrieval (e.g., Aho & Ullman, 1972; Lewis & Vasishth, 2005; Nicenboim & Vasishth, 2017), each phrase (e.g., *the smell*) is stored as a node (e.g., noun phrase) in memory. When a modifier is attached, it reactivates the node in memory that it attaches to and boosts its activation level. Thus, the attachment of the second modifier will reactivate the first modifier directly (by attaching to it) and the head of the subject phrase indirectly (because it was attached to the first modifier). In the conditions where the third noun was plural (e.g., *the smell of the stable of the farmers*), the reduction of attraction errors could have resulted from two opposing forces: the third plural noun was linearly closer to the verb (and thus should have been more active in memory) but its attachment to the second noun should have jointly reactivated the second and head nouns and increased their activation levels. As both nouns were singular, their retrieval should have reduced attraction errors, compared to the conditions with a second plural noun. Similarly to the first explanation, this account explains the 2nd-3rd-noun asymmetry as the result of the relative activation of constituents in memory, but it does not specify whether activation differences arise at encoding, retrieval, or both.

Finally, whereas qualitative patterns were similar between native and L2 speakers, there were quantitative differences: native speakers merely showed a reduction of the 2nd-3rd-noun asymmetry in the coordinated conditions (Experiment 2), but the asymmetry fully disappeared for L2 speakers (Experiments 1 and 2). At present, we cannot offer a conclusive explanation for why L2 speakers showed a stronger modulation of the 2nd-3rd-noun asymmetry than native

speakers. There are several possibilities that merit further research. The difference could be indicative of how nonnative speakers represent conjoined structures in their L2. In contrast with native speakers, L2 speakers may compute a "flat" analysis of coordination (Jackendoff, 1977), which would render the second and the third noun equidistant from the head noun, thereby eliminating the 2nd-3rd-noun asymmetry.

Another possibility is that L2 speakers are more influenced by lexical–semantic cues than L1 speakers. Under this explanation, L2 speakers may have used the conjunction *and* in the coordinated conditions as the most relevant cue to select a plural verb, prioritizing its lexical meaning over the information provided by the syntactic structure of the preamble. In contrast, the embedded conditions did not include any lexical cues to plurality beyond the plural morpheme on the modifiers. Thus, there was a lesser conflict between syntactic and lexical cues in the embedded conditions, and L2 speakers behaved similarly to native speakers. This explanation is in line with research on L2 ambiguity resolution, which has found that L2 speakers prioritize lexicosemantic and pragmatic cues over structural cues to resolve different types of lexical and structural attachment ambiguities (Felser, Roberts, Marinis, & Gross, 2003; Juffs, 1998; Juffs & Harrington, 1996; Pan & Felser, 2011; Pan, Schimke, & Felser, 2015; Papadopoulou & Clahsen, 2003).

Finally, a different explanation arises under an account where the 2nd-3rd-noun asymmetry is due to the semantic relationship between the modifiers and the head noun (Gillespie & Pearlmutter, 2011). Under this account, the contrast between language groups could indicate that L2 speakers were more sensitive than native speakers to semantic integration. The obvious question is why this should be the case. One possible answer is suggested by emerging work that proposes that L2 speakers have a less incremental scope of planning than native speakers, such that they prefer to process larger quantities of input before producing a response (Konopka & Forest, 2016). In the coordinated conditions, a larger scope of planning combined with the increased integration between the third noun and the head may have delayed L2 speakers' computation of the number of the subject phrase. Thus, in contrast with native speakers, L2 speakers may have still been in the process of computing the subject number when the third noun was displayed, making the third noun more likely to interfere with the number encoding of the subject phrase. It should be noted, however, that given the few studies on the extent of the scope of planning in native and L2 speakers, more research is necessary to assess whether a differential sensitivity to semantic information could be captured by the scope of planning differences.

While the cause of the stronger modulation of the 2nd-3rd-noun asymmetry in L2 than in native speakers merits further studies, our results yield clear conclusions regarding the processing of agreement in a nonnative language. First, they demonstrate that proficient L2 speakers can err at similar rates as native speakers when the impact of variables related to lexical processing is reduced. Second, we show that native and L2 speakers prioritize linguistic structure over linear distance in their agreement computations. Overall, our findings illustrate a less-explored but useful fact: that native and L2 speakers can err in very similar ways, and that these errors can reveal how they deploy different types of linguistic information during processing.

ACKNOWLEDGMENTS

NOTES
1. Structural distance can be quantified as the number of intervening syntactic nodes in the phrase structure representation of the subject phrase. The count of the number of nodes depends on the choice of a syntactic theory, which is beyond the scope of this paper. However, under most syntactic accounts, fewer nodes intervene between the second noun and the head noun than between the third noun and the head.
2. During the analysis, we discovered that two experimental items had been coded incorrectly, with a plural instead of singular subject head. These two items were excluded from analysis.
3. Due to nonconvergence, the model for response times in the native group was simplified by removing the random slopes for the factor plural noun position. In the L2 group, the model for accuracy was simplified by removing the random by-subject slope for the factor construction type.
4. Due to nonconvergence, the model for accuracy in the native group was simplified by removing the random by-item slope for the factor attractor. In the L2 group, the model for accuracy was simplified by removing the by-item slope for the factor construction type; the model for response times was simplified by removing the random by-item slope for the factor attractor.

REFERENCES
Aho, A. V., & Ullman, J. D. (1972). *The theory of parsing, translation and compiling: Vol. 1. Parsing.* Englewood Cliffs, NJ: Prentice Hall.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390–412.

Badecker, W., & Kuminiak, F. (2007). Morphology, agreement, and working memory retrieval in sentence production: Evidence from gender and case in Slovak. *Journal of Memory and Language, 56*, 65–85.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. (2013). Random-effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*, 255–278.

Bock, J. K., & Cutting, J. C. (1992). Regulating mental energy: Performance units in language production. *Journal of Memory and Language, 31*, 99–127.

Bock, J. K., & Miller, C. A. (1991). Broken agreement. *Cognitive Psychology, 23*, 45–93.

Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological), 26*, 211–252.

Camacho, J. A. (1997). *The syntax of NP coordination* (Unpublished doctoral dissertation, University of Southern California, Los Angeles).

Chemla, E., Cummins, C., & Singh, R. (2016). Training and timing local scalar enrichments under global pragmatic pressures. *Journal of Semantics, 34*, 107–126.

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*, 161–174.

Clahsen, H., & Felser, C. (2006a). Grammatical processing in language learners. *Applied Psycholinguistics, 27*, 3–42.

Clahsen, H., & Felser, C. (2006b). Continuity and shallow structures in language processing. *Applied Psycholinguistics, 27*, 107–126.

Coughlin, C. E., & Tremblay, A. (2013). Proficiency and working-memory-based explanations for nonnative speakers' sensitivity to agreement in sentence processing. *Applied Psycholinguistics, 34*, 615–646.

Cunnings, I. (2016). Parsing and working memory in bilingual sentence processing. *Bilingualism: Language and Cognition*. Advance online publication. doi:10.1017/S1366728916000675

Dallas, A. C., & Kaan, E., (2008). Second language processing of filler-gap dependencies by late learners. *Language and Linguistics Compass, 2, 3*, 372–388.

Dillon, B., Clifton, C., Jr., & Frazier, L. (2014). Pushed aside: Parentheticals and processing. *Language, Cognition and Neuroscience, 29*, 483–498.

Eberhard, K. M., Cutting, J. C., & Bock, K. (2005). Making syntax of sense: Number agreement in sentence production. *Psychological Review, 112*, 531–559.

Enochson, K., & Culbertson, J. (2015). Collecting psycholinguistic response time data using Amazon Mechanical Turk. *PLOS ONE, 10*. doi:10.1371/journal.pone.0116946

Felser, C. L., Roberts, T. M., Marinis, T., & Gross, R. (2003). The processing of ambiguous sentences by first and second language learners of English. *Applied Psycholinguistics, 24*, 453–489.

Foote, R. (2010). Age of acquisition and proficiency as factors in language production: Agreement in bilinguals. *Bilingualism: Language and Cognition, 13*, 99–118.

Foote, R. (2011). Integrated knowledge of agreement in early and late English–Spanish bilinguals. *Applied Psycholinguistics, 32*, 187–220.

Foote, R., & Bock, J. K. (2012). The role of morphology in subject-verb number agreement: A comparison of Mexican and Dominican Spanish. *Language and Cognitive Processes, 27*, 429–461.

Francis, W. N. (1986). Proximity concord in English. *Journal of English Linguistics, 19*, 309–317.

Franck, J., Vigliocco, G., & Nicol, J. (2002). Attraction in sentence production: The role of syntactic structure. *Language and Cognitive Processes, 17*, 371–404.

Gazdar, G., Pullum, G. K., Sag, I. A., & Wasow, T. (1982). Coordination and transformational grammar. *Linguistic Inquiry, 13*, 663–676.

Gibson, E., Piantadosi, S., & Fedorenko, K. (2011). Using Mechanical Turk to obtain and analyze English acceptability judgments. *Language and Linguistic Compass, 5*, 509–524.

Gillespie, M., & Pearlmutter, N. J. (2011). Hierarchy and scope of planning in subject–verb agreement production. *Cognition, 118*, 377–397.

Gillespie, M., & Pearlmutter, N. J. (2013). Against structural constraints in subject–verb agreement production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*, 515–528.

Grüter, T., Lew-Williams, C., & Fernald, A. (2012). Grammatical gender in L2: A production or a real-time processing problem? *Second Language Research, 28*, 191–215.

Haznedar, B., & Schwartz, B. (1997). Are there optional infinitives in child L2 acquisition? In E. Hughes, M. Hughes, & A. Greenhill (Eds.), *Proceedings of the 21st Annual Boston University Conference on Language Development (BUCLD)*, (pp. 293–306). Somerville, MA: Cascadilla Press.

Hopp, H. (2015a). Semantics and morphosyntax in L2 predictive sentence processing. *International Review of Applied Linguistics, 53*, 277–306.

Hopp, H. (2015b). Differential learning of prediction. *Linguistic Approaches to Bilingualism, 5*, 476–481.

Hopp, H. (2016). The timing of lexical and syntactic processes in L2 sentence comprehension. *Applied Psycholinguistics, 37*, 1253–1280.

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*, 87–98.

Ingria, R. (1990). The limits of unification. *Proceedings of the 28th annual meeting of the Association for Computational Linguistics* (pp. 194–204). Stroudsburg, PA: Association for Computational Linguistics.

Jackendoff, R. (1977). *X' Syntax*. Cambridge, Mass: MIT Press.

Jacob, G., Lago, S., & Patterson, C. (2016). L2 processing and memory retrieval: Some empirical and conceptual challenges. Commentary on Ian Cunnings' keynote article entitled "Parsing and working memory in bilingual sentence processing." *Bilingualism: Language and Cognition*. Advance online publication. doi:10.1017/S1366728916000948

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*, 434–446.

Jegerski, J. (2016). Number attraction effects in near-native Spanish sentence comprehension. *Studies in Second Language Acquisition, 1*, 5–33.

Jiang, N. (2004). Morphological insensitivity in second language processing. *Applied Psycholinguistics, 25*, 603–634.

Johannessen, J. B. (1998). *Coordination*. Oxford: Oxford University Press.

Juffs, A. (1998). Main verb vs. reduced relative clause ambiguity resolution in L2 sentence processing. *Language Learning, 48*, 107–147.

Juffs, A., & Harrington, M. (1996). Garden-path sentences and error data in second language processing research. *Language Learning, 46*, 286–324.

Keating, G. (2009). Sensitivity to violations of gender agreement in native and nonnative Spanish: An eye-movement investigation. *Language Learning, 59*, 503–535.

Keating, G. (2010). The effects of linear distance and working memory on the processing of gender agreement in Spanish. In B. VanPatten & J. Jegerski (Eds.), *Research in second language processing and parsing* (pp. 113–134). Amsterdam: John Benjamins.

Konopka, A., & Forest, T. A. (2016). *Linguistic experience (L1 vs. L2) shapes sentence formulation*. Talk at the 29th annual CUNY Conference on Human Sentence Processing, Gainesville, FL, March 3−5.

Kuznetsova, A., Bruun Brockhoff, P., & Haubo Bojesen Christensen, R. (2014). lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package). *R package version 2*.0–11. Retrieved from http://CRAN.R-project.org/package=lmerTest

Lewis, R., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science, 29*, 375–419.

Lim, H. L., & Christianson, K. (2015). Second language sensitivity to agreement errors: Evidence from eye movements during comprehension and translation. *Applied Psycholinguistics*, *36*, 1283–1315. doi:10.1017/S0142716414000290

Lorimor, H., Jackson, C. N., & Foote, R. (2015). How gender affects number: Cue-based retrieval in agreement production. *Language, Cognition and Neuroscience, 30*, 947–954.

McDonald, J. (2006). Beyond the critical period: Processing-based explanations for poor grammaticality judgment performance by late second language learners. *Journal of Memory and Language, 55*, 381–401.

Munn, A. (1987). Coordinate structure and X-bar theory. *McGill Working Papers in Linguistics, 4*, 121–140.

Munn, A. (1993). *Topics in the syntax and semantics of coordinate structures* (Unpublished doctoral dissertation, University of Maryland).

Nicenboim, B., & Vasishth, S. (2017). Models of retrieval in sentence comprehension: A computational evaluation using Bayesian hierarchical modeling. *Journal of Memory and Language*. Advance online publication.

Nicol, J., & Greth, D. (2003). Production of subject–verb agreement in Spanish as a second language. *Experimental Psychology, 50*, 196–203.

Nicol, J., Teller, M., & Greth, D. (2001). The production of verb agreement in monolingual, bilingual, and second language speakers. In J. L. Nicol (Ed.), *One mind, two languages: Bilingual language processing* (pp. 117–133). Malden, MA: Blackwell.

Pan, H.-Y., & Felser, C. (2011). Referential context effects in L2 ambiguity resolution: Evidence from self-paced reading. *Lingua, 121*, 221–236.

Pan, H.-Y., Schimke, S., & Felser, C. (2015). Referential context effects in non-native relative clause ambiguity resolution. *International Journal of Bilingualism, 19*, 298–313.

Papadopoulou, D., & Clahsen, H. (2003). Parsing strategies in L1 and L2 sentence processing: A study of relative clause attachment in Greek. *Studies in Second Language Acquisition, 25*, 501–528.

Powell, M. J. D. (2009). *The BOBYQA algorithm for bound constrained optimization without derivatives*. Report No. DAMTP 2009/NA06. Cambridge: University of Cambridge, Centre for Mathematical Sciences.

Prévost, P., & White, L. (2000). Missing surface inflection or impairment in second language acquisition? Evidence from tense and agreement. *Second Language Research, 16*, 103–133.

Progovac, L. (1998a). "Structure for Coordination, Part 1" State-of-the-Articles. *Glot International, 3*, 3–6.

Progovac, L. (1998b). "Structure for Coordination, Part 2" State-of-the-Articles. *Glot International, 3*, 6–9.

Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language*. New York: Longman.

R Development Core Team (2017). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org

Roberts, L. (2013). Sentence processing in bilinguals. In van Gompel, R. P. G. (ed.), *Sentence processing*. Hove: Psychology Press, p. 221–246 26 p. (Current Issues in the Psychology of Language).

Ross, J. R. (1967). *Infinite syntax*. Norwood, NJ: Ablex.

Sagarra, N., & Herschensohn, J. (2013). Processing of gender and number agreement in late Spanish bilinguals. *International Journal of Bilingualism, 17*, 607–627.

Sato, M., & Felser, C. (2010). Sensitivity to morphosyntactic violations in English as a second language. *Second Language, 9*, 101–118.

Shibuya, M., & Wakabayashi, S. (2008). Why are L2 learners not always sensitive to subject-verb agreement*? EUROSLA Yearbook, 8*, 235–258.

Solomon, E., & Pearlmutter, N. (2004). Semantic integration and syntactic planning in language production. *Cognitive Psychology, 49*, 1–46.

Sorace, A. (2004). Native language attrition and developmental instability at the syntax-discourse interface: Data, interpretation and methods. *Bilingualism: Language and Cognition*, 7, 143–146.

Sprouse, J. (2011). A validation of Amazon Mechanical Turk for the collection of acceptability judgments in linguistic theory. *Behavioral Research Methods, 43*, 155–167.

Staub, A. (2009). On the interpretation of the number attraction effect: Response time evidence. *Journal of Memory and Language, 60*, 308–327.

Staub, A. (2010). Response time distributional evidence for distinct varieties of number attraction. *Cognition, 114*, 447–454.

Tanner, D. (2011). *Agreement mechanisms in native and nonnative language processing: Electrophysiological correlates of complexity and interference* (Unpublished doctoral dissertation, University of Washington).

Tokowicz, N. (2015). *Lexical processing and second language acquisition*. New York: Routledge.

VanPatten, B., Keating, G. D., & Leeser, M. J. (2012). Missing verbal inflections as a representational problem. *Linguistic Approaches to Bilingualism, 2*, 109–140.

Vasishth, S., & Lewis, R. L. (2006). Argument-head distance and processing complexity: Explaining both locality and anti-locality effects. *Language, 82*, 767–794.

Vigliocco, G., & Nicol, J. (1998). Separating hierarchical relations and word order in language production. Is proximity concord syntactic or linear? *Cognition, 68*, 3–29.

Wagers, M., Lau, E., & Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language, 61*, 206–237.

Wagers, M. W., & Phillips, C. (2014). Going the distance: Memory and control processes in active dependency construction. *Quarterly Journal of Experimental Psychology, 67*, 1274–1304.

Wakabayashi, S. (1997). *The acquisition of functional categories by learners of English* (Unpublished doctoral dissertation, University of Cambridge).

Zoerner, E. (1995). *Coordination: The syntax of &P*. Unpublished doctoral dissertation, University of California, Irvine.