CrossMark

Studies in Second Language Acquisition, 2016, **38**, 5–33.
doi:10.1017/S027226311400059X

NUMBER ATTRACTION EFFECTS IN NEAR-NATIVE SPANISH SENTENCE COMPREHENSION

Jill Jegerski
University of Illinois at Urbana-Champaign

Grammatical agreement phenomena such as verbal number have long been of fundamental interest in the study of second language (L2) acquisition. Previous research from the perspective of sentence processing has documented nativelike behavior among nonnative participants but has also relied almost exclusively on grammar violation paradigms. The present investigation examined the online comprehension of subject-verb number agreement by native and nonnative speakers of Mexican Spanish using the more subtle agreement attraction paradigm, with stimuli grammatical in both conditions, in addition to a nonlocal grammaticality paradigm. The results of a self-paced reading experiment confirmed that nonnatives can acquire nativelike online sensitivity to verbal number agreement, even without grammar errors, and that such sensitivity develops only at very high levels of L2 proficiency.

The rapidly growing field of second language (L2) psycholinguistics has as one of its primary goals the description of online sentence processing (i.e., real-time comprehension) in nonnative languages acquired in adulthood, with particular emphasis on the identification of specific traits that might distinguish nonnative processing from native processing.

The author would like to thank Marilyn Buck for help with participant recruitment and access to facilities in Mexico City; the faculty, staff, and students at the CELE of the Universidad Nacional Autónoma de México in Mexico City for help with access to facilities; and Clyde "Chico" James and Danilo Tunac for help with participant recruitment. This research was supported by NSF grant BCS-0844035.

Correspondence concerning this article should be addressed to Jill Jegerski, Department of Spanish and Portuguese, University of Illinois, MC-176, 707 South Mathews Avenue Urbana, IL 61801. E-mail: jegerski@illinois.edu

© Cambridge University Press 2014

https://doi.org/10.1017/S027226311400059X Published online by Cambridge University Press

Although there is fairly widespread consensus that first language (L1)-L2 differences probably do exist on some level, and there are claims that differences in processing may be the culprits behind what appears to be an overall lack of success in adult L2 acquisition as compared to childhood acquisition, the specific nature of such differences remains open and widely debated. More specifically, current research in this area seeks to test three main types of accounts of nonnative processing: (a) that L2 sentence processing deficiency is grammatical in nature and is therefore best described in terms of modules of linguistic knowledge such as syntax and semantics (Clahsen & Felser, 2006a, 2006b; Clahsen, Felser, Neubauer, Sato, & Silva, 2010; Sorace, 2005); (b) that L2 processing deficiency is primarily a performance issue that can be best described in terms of extralinguistic constructs such as working memory and automaticity (Coughlin & Tremblay, 2013; Hopp, 2006, 2010; Jiang, 2007; Segalowitz, 2003) and that may even be overstated (Witzel, Witzel, & Nicol, 2012); and (c) that transfer or interference from the native language can be a critically limiting factor in L2 processing (Frenck-Mestre & Pynte, 1997; Jackson, 2010; Juffs, 1998). The present study examined the validity of the predictions of a theory of the first type, that is, a grammar-based account.

The shallow structure hypothesis (Clahsen & Felser, 2006a, 2006b; Clahsen et al., 2010) is a theory of nonnative comprehension that proposes that L2 syntactic, morphosyntactic, and morphological processing can be distinguished primarily by deficiencies at the level of grammatical representation and in the application of such grammatical knowledge during language comprehension. Specifically, the theory proposes that L2 processing is less detailed than native processing, or is *shallow*, in that it fails to integrate abstract and hierarchical elements of phrase structure, relying instead on extrasyntactic sources of information such as lexical specifications, semantics, and pragmatics as well as on nonlinguistic information such as general world knowledge. By way of compensating for deficiencies in syntax, the available sources of linguistic and nonlinguistic information receive a disproportionately greater emphasis or level of activation during comprehension. Whereas other L2 processing theories view divergent processing as a stage in the developmental trajectory (Pienemann, 1998; VanPatten, 1996, 2007), the shallow structure hypothesis proposes that even the most proficient nonnatives, who are often ultimately able to arrive at the same level of global comprehension as native speakers after reading text or listening to an utterance, will always be critically different from natives when it comes to the moment-by-moment linguistic processes involved in sentence comprehension.

One key feature of online sentence comprehension that is purportedly lacking in shallow L2 processing is certain morphosyntactic phenomena, which have long been of interest in L2 research. According to the

https://doi.org/10.1017/S027226311400059X Published online by Cambridge University Press

shallow structure hypothesis, L2 sensitivity to morphosyntax during online processing is diminished, even at high levels of proficiency, and online sensitivity to syntactic dependencies will only be observed "between adjacent or locally related words" (Clahsen & Felser, 2006b, p. 565). To illustrate the significance of distance in computing syntactic dependencies, Clahsen et al. (2010) propose that the L2 processing of accusative case is relatively less problematic because it is carried out within the local domain, the verb phrase, whereas "SV agreement dependencies span the entire clause" (Clahsen et al., 2010, p. 37). Thus, sentence comprehension paradigms that involve the processing of subject-verb agreement, particularly over distance, are appropriate for testing the empirical predictions of the shallow structure theory and, more broadly, for determining whether its focus on grammar as the source of L2 deficiency is plausible as an account of nonnative sentence processing behavior.

AGREEMENT ATTRACTION

One such sentence processing paradigm, which has been referred to as broken agreement, agreement attraction, or number attraction, examines the processes involved in computing verbal number agreement with complex subjects. Research in native language psycholinguistics (Bock & Miller, 1991) has observed that the presence of an embedded noun phrase (NP) of conflicting number leads to confusion of verbal number in speech production, so participants are more likely to make concordance mistakes when asked to create complete sentences by continuing a prompt such as (1b), as compared to a prompt such as (1a). In other words, the verb sometimes agrees with the embedded or attractor noun rather than the head noun in spoken language. Interestingly, the rate of errors is much lower when the embedded noun is singular rather than plural, as in (1d) versus (1c).

(1)	a. *The author of the speech . . .*	Singular-Singular (SS)
	b. *The author of the speeches . . .*	Singular-Plural (SP)
	c. *The authors of the speeches . . .*	Plural-Plural (PP)
	d. *The authors of the speech . . .*	Plural-Singular (PS)

A similar pattern has been observed in sentence comprehension research (Nicol, Forster, & Veres, 1997; Pearlmutter, Garnsey, & Bock, 1999), in which stimuli are complete sentence versions of the speech production prompts (e.g., *The author of the speech/speeches is here now* SSS/SPS). In the given stimulus conditions, the number feature of the verb agrees with that of the head noun, so the sentences are all grammatical, although the tendency of an attractor noun with a contrasting number feature to

https://doi.org/10.1017/S027226311400059X Published online by Cambridge University Press

mitigate the effects of ungrammatical agreement has commonly been observed as well. But even with entirely grammatical sentences, the computation of number agreement between the complex subject and the verb appears to lead to processing difficulty when the number feature of the embedded attractor noun conflicts with that of the head noun, which is typically evident in longer reading times on the verb in the SPS condition versus the SSS condition, though not for the PSP condition versus the PPP condition. Much of the research on agreement attraction has targeted English, but the effect is also known to occur in native Spanish, both in speech production (Bock, Carreiras, & Meseguer, 2012; Vigliocco, Butterworth, & Garrett, 1996) and in comprehension (Acuña-Fariña, Meseguer, & Carreiras, 2014).

Agreement attraction effects have been accounted for by at least three different theoretical proposals. One suggests that both production and comprehension effects are due to an inaccurate representation of the complex subject, in which the number feature of the embedded attractor noun erroneously migrates upward and is used to mark the entire subject NP (Nicol et al., 1997). This only happens when the embedded noun is plural, because single is the default or unmarked status, whereas plural number is marked with a feature (Bock & Eberhard, 1993). A second type of theory has proposed that agreement attraction effects arise not in the actual representation of number agreement but rather during the subsequent retrieval of the relevant features on the subject for the purposes of checking agreement on the verb, which only occurs when verbal morphology is inconsistent with the number that had been predicted on processing the subject. One interesting prediction that has been derived from the retrieval approach is that true attraction effects in comprehension can only occur with ungrammatical sentences because there is otherwise no need for retrieval. Indeed, the observed reading time effects on grammatical stimuli from two online comprehension studies (Pearlmutter et al., 1999; Wagers, Lau, & Phillips, 2009) have been argued by Wagers et al. to reflect spillover of a basic NP number effect rather than true attraction effects, though there remain at least two studies that have provided more solid evidence of agreement attraction effects in the comprehension of grammatical sentences (Acuña-Fariña et al., 2014; Nicol et al., 1997). Finally, a third account views agreement attraction as reflecting an adaptation of the sentence parser to typically noisy input, which leads the parser to make rational inferences regarding potential noise during sentence processing (Bergen & Gibson, 2012). A basic principle of operation under this account would arise from the application of the Bayesian size principle and the knowledge that there are more permutations for a randomly inserted morpheme than for a deleted morpheme, so a deleted morpheme is more likely to occur. As an example, the parser calculates that *speeches* could erroneously be articulated or perceived as *speech*, but the opposite

https://doi.org/10.1017/S027226311400059X Published online by Cambridge University Press

scenario would be a less frequent source of noise, so the effects of a conflicting attractor noun that is plural would be greater than those of an embedded noun that is singular, and the additional number cues provided by richer morphology would be predicted to result in fewer agreement errors. Thus, both the representational (Bock & Eberhard, 1993) and the noisy input (Bergen & Gibson, 2012) accounts predict that attraction would occur during comprehension, the target of the present study, whereas the retrieval account (Wagers et al., 2009) predicts that it would not. Additionally, under any of these three theoretical accounts, attraction effects minimally require the processor to distinguish between different verbal number morphemes and to either maintain an active representation or retrieve a stored representation of those features over linear distance. From the perspective of L2 acquisition, even though attraction effects reflect something going wrong during the processing of agreement, they require a level of detail and computation over distance that would be predicted not to occur in shallow L2 processing. If, on the other hand, nonnative processing is not necessarily shallow with regard to grammatical detail but rather less efficient than native processing and generally more prone to error, agreement attraction effects might be notably larger or persistent, perhaps even affecting overall sentence comprehension accuracy. Agreement attraction can therefore be an important point of comparison between native and nonnative sentence comprehension.

NONNATIVE COMPREHENSION OF NUMBER AGREEMENT

Several previous psycholinguistic and neurolinguistic studies have investigated the processing of different types of number agreement in nonnative languages, primarily with error recognition paradigms. One early neurolinguistic study found persistent deficiencies in the nonnative comprehension of verbal number agreement. Ojima, Nakata, and Kakigi (2005) used event-related potentials (ERPs) to examine the sensitivity of native Japanese speakers of English as a L2 (ESL) to violations of local verbal number agreement such as the one in (2).

(2) Grammaticality stimulus (Ojima et al., 2005)
a. *Some foreigners eat sushi.*
b. **Some foreigners eats sushi.*

The native participants in the investigation showed a two-phase pattern of brain activation in response to the written agreement errors, including an early left-lateralized negativity followed by a P600 (a positive-going waveform typically associated with syntactic and morphosyntactic anomalies), whereas a group of higher proficiency L2 learners showed a

https://doi.org/10.1017/S027226311400059X Published online by Cambridge University Press

nativelike pattern only for the first phase. A lower proficiency group did not exhibit either of the two phases of the activation pattern. In another ERP experiment (Rossi, Gugler, Friederici, & Hahne, 2006), however, nonnatives did appear to process violations of verbal number agreement in a nativelike fashion. The L1 German-L2 Italian and L1 Italian-L2 German participants in this investigation showed a nativelike two-phase pattern of brain activation in response to written, nonlocal number agreement errors, such as that illustrated in (3), in which a left anterior negativity (LAN) was followed by a P600.

(3) Nonlocal grammaticality stimulus (Rossi et al., 2006)
a. *Der Junge im Kindergarten singt ein Lied.*
b. **Der Junge im Kindergarten singst ein Lied.*
"The boy in the kindergarten sings/*sing a song."

The two-phase pattern was apparent at higher levels of proficiency, but L2 participants with lower proficiency showed only the LAN component. Tanner, Inoue, and Osterhout (2014) also observed a P600 among a subset of L1 Spanish-L2 English participants, although no L1 English group was included for comparison. P600 forms were also observed among the L1 Spanish-L2 English participants studied by Tanner, Nicol, Herschensohn, and Osterhout (2012), as was a nativelike tendency toward smaller P600 effects when a plural versus singular noun intervened between a mismatched singular head noun and plural verb (e.g., **The winner of the big trophies/trophy have . . .*). Taken together, these four ERP studies suggest that divergent processing of verbal number agreement among L2 learners may be a developmental phase that is overcome as acquisition progresses to sufficiently advanced levels of proficiency. In addition, Tanner, Osterhout, and Herschensohn (2009) have shown that a delayed P600 can show up as early as the first year of language study in some individual learners, so proficiency may not be the only factor at play.

More recently, several self-paced reading studies have expanded on the ERP work on the processing of agreement morphosyntax in L2 Spanish, French, and English with reading time data using either error recognition paradigms or the grammatical version of the agreement attraction paradigm (i.e., SSS/SPS, with both conditions grammatical). Foote (2011) found that highly proficient speakers of L2 Spanish had nativelike online sensitivity to both local and nonlocal errors of verbal number agreement. Parenthetically, the same participants also showed nativelike processing of gender agreement, which does not occur in their native English. Sagarra and Herschensohn (2011) found similarly nativelike online sensitivity to violations of local noun-adjective agreement of number and gender in intermediate-level L2 Spanish, but not at the beginning level. Finally, Coughlin and Tremblay (2013) also reported nativelike reading

https://doi.org/10.1017/S027226311400059X Published online by Cambridge University Press

time effects for local and nonlocal number agreement errors with object clitics in L2 French. Only the more proficient of two participant groups recruited by Coughlin and Tremblay showed evidence of nativelike online processing of number, an outcome that was consistent with the previous ERP studies that had suggested that nativelike online sensitivity to agreement can develop over time, as L2 proficiency increases.

Thus, these first three self-paced reading studies contribute to a growing body of evidence that adult L2 learners—at least those whose native language also has overt number agreement—can come to process several different types of number agreement in a nativelike manner. However, one potential limitation among the aforementioned studies that documented nativelike processing is that they all relied on some type of agreement violation paradigm to elicit reading time or ERP effects. In other words, the L2 participants in these studies showed crucial sensitivity to an aspect of grammar that is probably explicitly taught in most formal language-learning contexts. It is therefore possible, at least in principle, that the observed sensitivities can result from some level of activation of explicit metalinguistic knowledge of grammar in addition to the more intuitive and natural type of linguistic knowledge that is commonly assumed by the term *nativelike*. This is not to say that the effects observed in previous research arose entirely from the application of explicit rules or that they do not represent authentic processing effects. After all, native participants show these effects as well, and the P600 in particular is not thought to reflect the conscious application of a basic rule. Rather, the idea is that processing effects that arise during the reading of sentences without errors may be less robust—particularly with reading time measures, which are known to be susceptible to the level of attention to errors demanded by an experimental task (Leeser, Brandl, & Weissglass, 2011)—and thus less likely to occur among L2 participants. Or perhaps reading time effects that do not depend on error recognition appear at higher levels of L2 proficiency than with violation paradigms. As suggested by Foote (2011), this potential limitation of error-based processing paradigms is especially valid for the studies that were conducted in the context of university foreign language departments in the United States, as many or most of the highly proficient nonnative participants were foreign language instructors, who would be expected to possess exceptionally strong explicit knowledge of the L2. Thus, L2 agreement processing research with grammatical stimuli and with more naïve participants represents a novel contribution to this line of study.

To my knowledge, only one previous study employed the more subtle agreement attraction paradigm in its grammatical version (i.e., SSS/SPS) in a study of nonnative comprehension of grammatical number. Jiang (2004) examined the processing of verbal number agreement by native Chinese speakers of ESL in three self-paced reading experiments. In the

https://doi.org/10.1017/S027226311400059X Published online by Cambridge University Press

first experiment, the nonnative participants failed to show nativelike agreement attraction effects. The second and third experiments employed an agreement violation paradigm and revealed nativelike processing of local person agreement, but again not of number agreement. Jiang therefore concluded that the nonnative readers did not process number agreement during online sentence comprehension and that this lack of online sensitivity to verbal number could not be attributed to general L2 proficiency level, given that the participants in the three experiments were all enrolled in university programs in the United States at the time of testing. Thus, the outcome of this study is potentially at odds with those of prior investigations that documented nativelike L2 processing of verbal number agreement. The discrepancy is not likely due to the difference in experimental paradigms (i.e., grammaticality versus agreement attraction), because Jiang's participants also failed to show online sensitivity to verbal number in an agreement violation paradigm. It is more likely that the different outcomes can be traced to participant differences in variables such as proficiency level. Native language may have also been at play, given that Jiang's participants were the only ones whose L1, Chinese, does not have verbal number morphology. This last point is of particular theoretical significance because Jiang's study has been specifically cited as support for the shallow structure hypothesis, which also claims that the deficiencies present in shallow L2 processing cannot be attributed to properties of the L1 (Clahsen et al., 2010).

To summarize, previous ERP and self-paced reading research on number agreement has found nativelike L2 processing at sufficiently high levels of proficiency but only with grammaticality paradigms that could potentially yield different results than paradigms that do not include errors, especially when the L2 participants are recruited in the context of university foreign and second language programs. Only one previous investigation (Jiang, 2004) included the more implicit agreement attraction paradigm, but the participant population in that series of experiments showed a general lack of online sensitivity to verbal number, even when measured with grammar violations, so processing was perhaps limited by either insufficient proficiency or the properties of their native language, Chinese. Given this background, the present study set out to determine whether nativelike L2 processing of verbal number agreement is possible even when tested with the agreement attraction paradigm, which does not rely on ungrammatical stimuli to evoke reading time effects. In an effort to further minimize the impact of explicit knowledge on the experimental results, the L2 participants for the current investigation were not recruited in a university context. To maintain comparability with prior studies, however, the same participants were also tested with an agreement violation paradigm. Finally, to ensure that native language was not a limiting factor, the participants were native English speakers of L2 Spanish, so both of their languages instantiated overt number agreement.

https://doi.org/10.1017/S027226311400059X Published online by Cambridge University Press

THE PRESENT STUDY

The research questions that guided this investigation were the following:

1. Do nonnative readers exhibit agreement attraction effects while processing verbal number in Spanish?
2. Do the same nonnative readers of Spanish exhibit nativelike sensitivity to the grammaticality of nonlocal agreement while processing verbal number in Spanish?
3. Do one or both of these processing behaviors appear to develop over time, as proficiency in the L2 increases?

METHOD

Participants

The baseline or comparison group for this study was composed of 32 native speakers of Spanish (23 women; 9 men), who were recruited and tested at a large public university in Mexico City and were born and raised in that area or in the contiguous states. All had graduated high school and completed some postsecondary coursework, and most were enrolled as undergraduate students at the time of testing. None of the individuals in this group had advanced knowledge of any foreign language and none had been abroad for periods of greater than 2 weeks, with the exception of trips to other Spanish-speaking countries. These participants were therefore assumed to be monolingual speakers of Mexican Spanish.

A total of 44 native speakers of English also participated in the study. These were residents of Mexico City who were late acquirers of Spanish, starting with first exposure after age 12 and not reaching fluency until age 18 or later. All were high school graduates who had completed at least some postsecondary coursework, and most were professionals with advanced university degrees. These nonnative participants were assigned to two groups on the basis of a written test of general language proficiency in Spanish, which is described in more detail in the Materials section. Of the 44 total L2 participants, 24 individuals (16 women; 8 men) had a proficiency test score that fell within the range of scores obtained by the native speakers and were designated *near native*. The remaining 20 nonnative participants (13 women; 7 men) were assigned to a second, less proficient group called *advanced*. Further information regarding the background characteristics of the two L2 participant groups, as assessed via the LEAP-Q language experience questionnaire (Marian, Blumenfeld, & Kaushanskaya, 2007), can be found in Table 1.

https://doi.org/10.1017/S027226311400059X Published online by Cambridge University Press

Table 1. Language background information as reported by L2 participants

	Near native (n = 24)			Advanced (n = 20)		
Characteristic	*M*	*SD*	Range	*M*	*SD*	Range
Age of onset	20.3	6.4	12–32	21.3	9.4	13–41
Years of residency	13.0	12.6	0.9–43	6.0	5.2	1.3–20.5
DELE score	43.7	2.3	40–48	32.8	5.6	21–39
Self-ratings for Spanish (0–10)						
Speaking	8.2	1.0	6–10	7.3	1.3	3–9
Understanding	8.3	0.7	7–10	7.7	0.8	6–9
Reading	7.8	1.3	4–9	6.8	1.6	3–9
Current exposure						
% English	44.6	18.4	10–85	58.3	21.1	19–89
% Spanish	52.8	19.2	20–90	38.6	21.8	10–80
% Other	2.6	7.5	0–36	3.2	5.9	0–20

Note. DELE = *Diplomas de Español como Lengua Extranjera* (Certificates of Spanish as a Foreign Language).

Materials

Self-Paced Reading. The stimuli for the self-paced reading experiment were 40 sentences based on a subset of the 64 Spanish preambles (i.e., speaking prompts) used by Bock et al. (2012) in their study of agreement attraction effects in the production of subject-verb number agreement by native speakers of Spanish and English. The preambles were complex subject NPs such as *The author(s) of the novel(s)* . . . , which could vary with regard to the number of the head noun and the intervening noun in the prepositional phrase. With the head noun and the intervening noun both appearing in both singular and plural form, four different conditions of these preambles were employed in the Bock et al. study. The stimuli set was balanced with regard to the grammatical gender of the head noun and the intervening noun, with equal numbers of each of four permutations: masculine-masculine, masculine-feminine, feminine-masculine, and feminine-feminine. Norming procedures were also conducted by Bock et al. to establish that their speech production stimuli were counterbalanced for additional factors known to affect the rate of attraction errors in speech production, such as notional number, imagery, sensibility, and distributivity (Eberhard, 1999; Vigliocco, Butterworth, et al., 1996; Vigliocco, Hartsuiker, Jarema, & Kolk, 1996). The previously manipulated variable, distributivity, was balanced for the present study, with half the stimuli containing nondistributive or single-token subjects (e.g., *The witness of the lawyers* . . .) and half containing distributive or multiple-token subjects (e.g., *The color of the flowers* . . .).

https://doi.org/10.1017/S027226311400059X Published online by Cambridge University Press

Several changes were made to the Bock et al. (2012) preambles in creating the stimuli for the present experiment, which are illustrated in (4) and (5). First, some of the lexical items in the preamble phrases were modified to adjust for the different dialect of Spanish spoken by the participants (i.e., Mexican vs. Canary Islands Spanish). Second, the preambles were used as the basis for complete sentences, including inflected verbs, as the present study targeted sentence comprehension rather than speech production. Finally, the four conditions for the stimuli were modified to include a manipulation of the head noun in addition to the manipulation of the intervening noun that yields agreement attraction stimulus conditions. More specifically, only the two conditions that typically show the most robust attraction effects, SSS and SPS, were employed to measure this type of effect in the present study. Following Jiang (2004), an additional two conditions were created to test for grammaticality effects that arise from a manipulation of the number of the head noun rather than the intervening noun. Unlike some previous research that has relied primarily on stimuli with copular verbs (e.g., Jiang, 2004), the stimuli for the present study included a wider variety of verbs to maximize the generalizability of the conclusions.

Examples of the two types of stimuli and conditions are given in (4) and (5), where it can be seen that the stimulus conditions differed with regard to the plurality of one of the two preverbal nouns. Backslashes indicate the phrase-by-phrase segmentation for the self-paced reading. For the stimulus in (4), an attraction effect would be evident in relatively longer reading times on the verb in (4b) versus (4a). This is because the processing of verbal number agreement between the singular head noun *el testigo* "the witness" and the singular verb *tiene* "has" can be interrupted by an intervening noun with a contrasting plural number feature, such as *los abogados* "the lawyers."

(4) Agreement attraction stimulus
a. *El testigo* \ *del abogado* \ *tiene* \ *mala fama* \ *en el barrio.* SSS
b. *El testigo* \ *de los abogados* \ *tiene* \ *mala fama* \ *en el barrio.* SPS
"The witness of the lawyer/lawyers has a bad reputation in the neighborhood."

(5) Nonlocal grammaticality stimulus
a. *Los testigos* \ *del abogado* \ *tienen* \ *mala fama* \ *en el barrio.* PSP
b. *El testigo* \ *del abogado* \ **tienen* \ *mala fama* \ *en el barrio.* SSP
"The witnesses/*witness of the lawyer have a bad reputation in the neighborhood."[1]

Stimuli such as the example in (5) also are known to evoke longer reading times on the verb in (5b) relative to the verb in (5a), though in this case the effect is presumably due to the presence of a grammar violation. In other words, readers tend to slow down on the verb in the ungrammatical condition of the stimulus because it is at that point that the sentence

https://doi.org/10.1017/S027226311400059X Published online by Cambridge University Press

becomes ungrammatical, as indicated by the asterisk on the third region of example (5b). Thus, the verb was the primary region of interest for both sets of stimuli.

Each of the 40 critical sentences was used to create both stimulus types, appearing in the four total conditions exemplified in (4) and (5). Four counterbalanced presentation lists were created, with each sentence appearing only once per list, so that (4a) would appear in the first list, (4b) in the second, (5a) in the third, and (5b) in the fourth. The target stimuli were combined with 10 practice items and 120 unrelated distractors (20 of which were ungrammatical) and were presented in pseudorandom order such that no two sentences of the same type appeared in succession.

Offline Acceptability Judgment. In addition to the self-paced reading materials testing online processing behavior, there was also a very brief offline acceptability judgment task testing participants' knowledge of Spanish subject-verb number agreement in contexts similar to those in the online task, that is, with complex NP subjects. The acceptability judgment task was administered via computer such that each sentence appeared on the screen for up to 15 s and then the participant advanced to a different display and rated the sentence for correctness on a Likert scale ranging from 1 (*Está mal* "It's bad") to 4 (*Está bien* "It's fine"). Ten critical sentential items (five of which were grammatical) appeared along with five practice items and 50 unrelated distractors. Target ratings were awarded one point each during scoring, and ratings that were one point off (i.e., 2 instead of 1 or 3 instead of 4) were awarded half a point.

Proficiency Test. Following Montrul and Slabakova (2003) and others, the independent measure of general proficiency used to establish the criterion for near-nativeness was an adapted version of a subsection of the *Diplomas de Español como Lengua Extranjera* (Certificates of Spanish as a Foreign Language; DELE), a standardized test of Spanish as a foreign language (Instituto Cervantes, 2007). This assessment included 50 objective items testing superior-level grammar and vocabulary and was administered in written form. This particular version of the test was adapted for learners of Mexican Spanish, as the official version of the exam is geared toward assessment of Peninsular Spanish. After the initial modifications, the test was piloted with 12 native speakers of Mexican Spanish to identify and eliminate any additional dialect-specific items.

Procedure

Participants first completed the background questionnaire and the self-paced reading experiment, both via computer. Next was the proficiency

https://doi.org/10.1017/S027226311400059X Published online by Cambridge University Press

test, which was administered in a pencil-and-paper format, and then finally the computerized acceptability judgment task. All participants were paid approximately 40 U.S. dollars for a 2 hr experimental session.

For the self-paced reading, sentences were presented on a 15 in. laptop computer running SuperLab stimulus presentation software (Version 4.0.5). The display was phrase-by-phrase, left-to-right or noncentered, and noncumulative (i.e., moving window), and participants pressed a green key on a Cedrus RB-730 response pad to advance through the various screens of each trial. A binary-choice comprehension question appeared after every stimulus. The comprehension questions were all in Spanish and targeted sentence meaning rather than grammatical form (e.g., "They say that the witness is bad." a. true b. false). No feedback was given for correct or incorrect answers. Prior to the self-paced reading experiment, participants were instructed to read at a normal speed, as if reading a newspaper or a book, and they were told that the test targeted reading comprehension. Grammatical form was not mentioned prior to or during the experiment. Detailed instructions and 10 practice items were presented prior to the experimental block and an optional 10 min break was offered when the participant had read half of the 160 total sentences.

RESULTS

Mean scores from the 10 verbal agreement items on the acceptability judgment task were 8.1 (SD = 2.1) for the native participants, 8.8 (SD = 1.5) for the near natives, and 7.6 (SD = 1.6) for the advanced group. The scores were submitted to one-way ANOVAs, which showed an effect of group that was marginally significant by subjects and significant by items: $F_1(2, 75) = 2.45$, $p = .093$; $F_2(2, 8) = 12.60$, $p = .003$. Post hoc Tukey's HSD and paired t tests indicated that the difference between the near-native and advanced groups was marginally significant by subjects ($p = .084$) and significant by items ($p = .003$), but neither of the other comparisons was significant. Thus, both L2 groups had roughly nativelike knowledge of subject-verb number agreement, at least as measured by this type of short offline task, although the near-native group was likely at a higher level than the advanced group.

The data from the self-paced reading task were cleaned such that all reading times that were less than 100 ms were treated as missing values, and all reading times greater than 2,000 ms were trimmed to 2,000 (3,000 for the last stimulus region, to account for sentence wrap-up). These steps affected 5.01% of the total reading time data. The mean reaction times for the agreement attraction stimuli for each of the three participant groups are represented in Figures 1, 2, and 3, and the corresponding reaction times and accuracy scores are given in Table 2. The descriptive

https://doi.org/10.1017/S027226311400059X Published online by Cambridge University Press

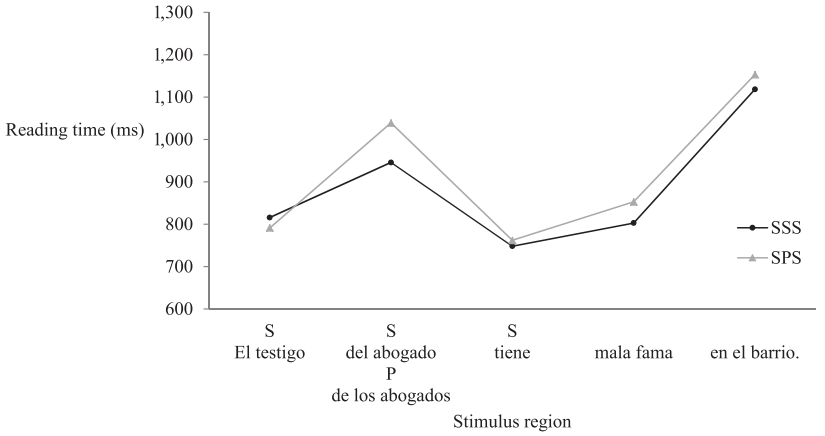


Figure 1. Native reading times for agreement attraction: mean reading times by stimulus condition. SPS was the agreement attraction condition and could yield longer reading times than the SSS condition starting at Region 3.

statistics for the nonlocal grammaticality stimuli are illustrated in Figures 4, 5, and 6 and in Table 3. Trimmed reading times from the attraction and nonlocal grammaticality stimuli were analyzed in separate but similar sets of statistical tests. In both cases, data from the region of interest containing the verb (Region 3), the spillover region (Region 4), the sentence-final region (Region 5), and the comprehension questions (both reaction times and accuracy scores) were submitted to 3 × 2

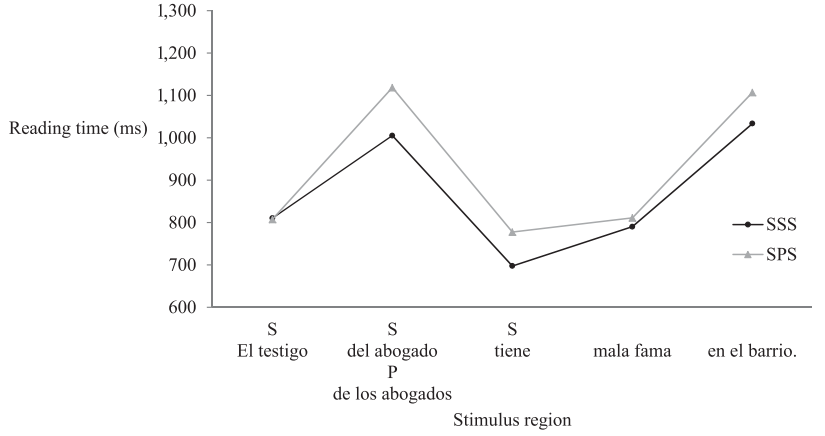


Figure 2. Near-native reading times for agreement attraction: mean reading times by stimulus condition. SPS was the agreement attraction condition and could yield longer reading times than the SSS condition starting at Region 3.

https://doi.org/10.1017/S027226311400059X Published online by Cambridge University Press

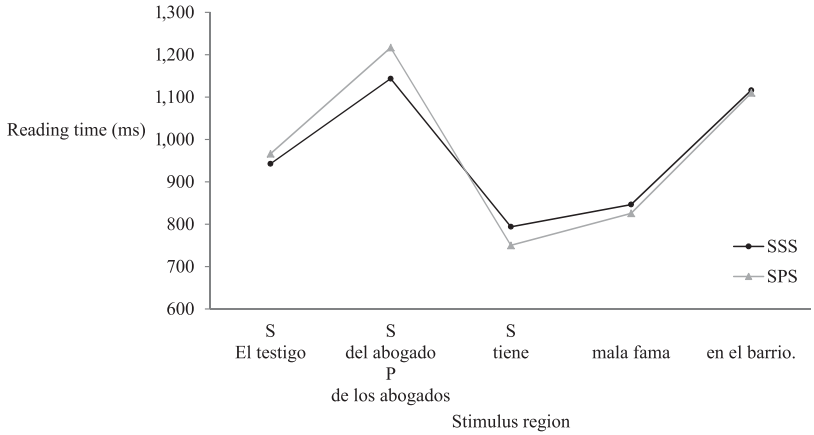


Figure 3. Advanced reading times for agreement attraction: mean reading times by stimulus condition. SPS was the agreement attraction condition and could yield longer reading times than the SSS condition starting at Region 3.

(Group × Agreement or Grammaticality) mixed between-within ANOVAs, with either subjects (F_1 and t_1) or items (F_2 and t_2) as a random factor. Data from all trials were included, even those that corresponded to incorrect comprehension question responses, because removal of the inaccurate trials would have resulted in a high number of missing values (20% in the case of the advanced group). Alpha was set at .05 for all analyses. Effects that reached this criterion both by subjects and by items were treated as significant, and those that reached it only by subjects or by items were treated as approaching significance.

The results of the ANOVAs for the experimental sentences that tested for agreement attraction effects are provided in Table 4. At Region 3, the region with the verb, the Group × Agreement interaction was significant by subjects and marginally significant by items. Post hoc paired *t* tests conducted to probe the possible interaction revealed that the three

Table 2. Attraction stimuli: Mean accuracy and reaction time (RT) for comprehension questions (*SD*s in parentheses)

	Accuracy		RT (ms)	
Participant group	SSS	SPS	SSS	SPS
Native	.878 (.126)	.844 (.122)	4,518 (1,309)	4,509 (1,407)
Near native	.900 (.102)	.908 (.110)	4,067 (1,038)	3,904 (973)
Advanced	.885 (.166)	.820 (.167)	4,120 (842)	4,164 (934)

Note. RT = reaction time; SSS = Singular-Singular-Singular; SPS = Singular-Plural-Singular.

https://doi.org/10.1017/S027226311400059X Published online by Cambridge University Press

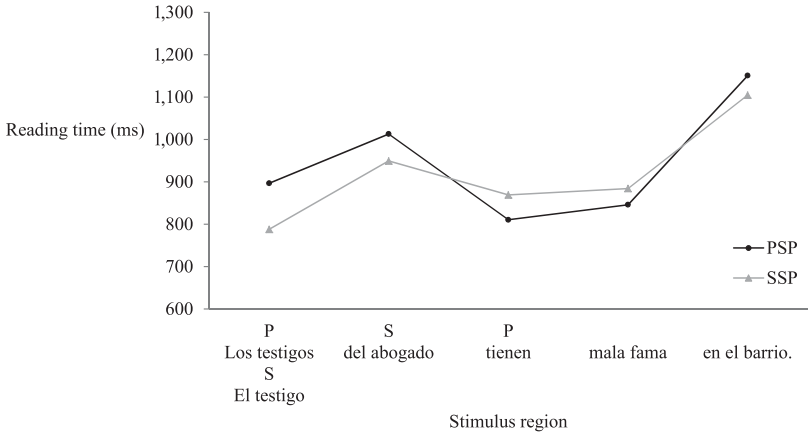


Figure 4. Native reading times for nonlocal grammaticality: mean reading times by stimulus condition. The SSP condition was ungrammatical and could yield longer reading times than the PSP condition starting at Region 3.

groups did indeed differ with regard to the within-subjects effects; only the near-native group showed significantly longer reading times for the SPS versus SSS stimulus condition at Region 3, $t_1(23) = 2.95$, $p = .007$, $t_2(39) = 2.14$, $p = .039$. No such effect was observed among the L1 group, $t_1(31) = 0.68$, $p = .503$, $t_2(39) = 0.48$, $p = .636$, or among the advanced group, $t_1(19) = 1.48$, $p = .154$, $t_2(39) = 0.98$, $p = .332$. At Region 4, the post-verbal or spillover region, the interaction was marginally significant by

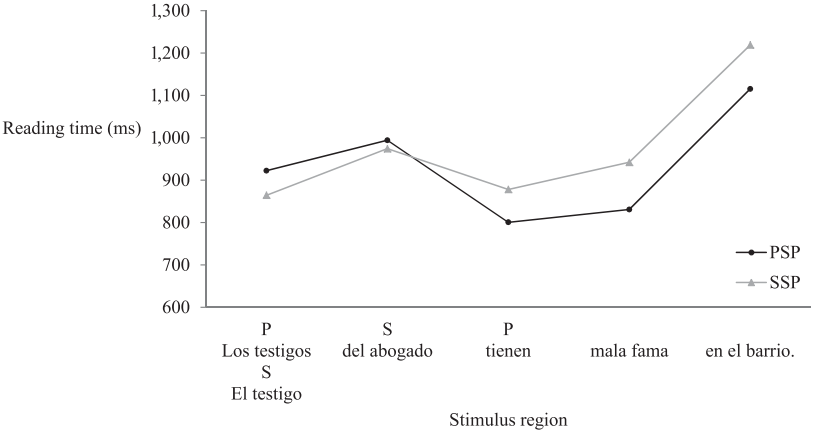


Figure 5. Near-native reading times for nonlocal grammaticality: mean reading times by stimulus condition. The SSP condition was ungrammatical and could yield longer reading times than the PSP condition starting at Region 3.

https://doi.org/10.1017/S027226311400059X Published online by Cambridge University Press

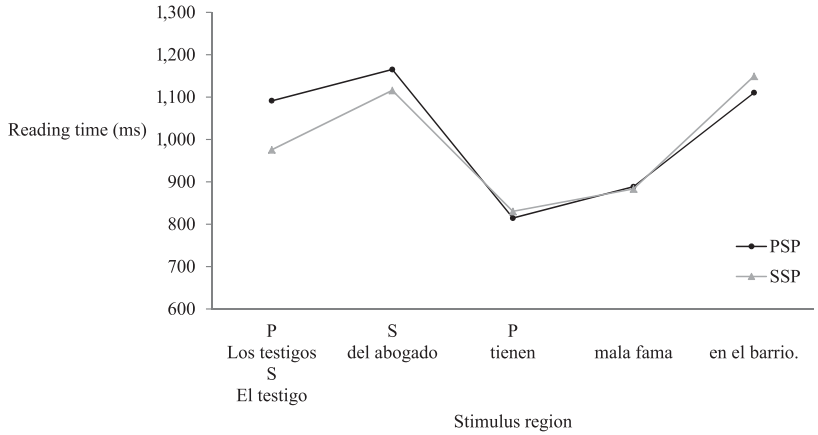


Figure 6. Advanced reading times for nonlocal grammaticality: mean reading times by stimulus condition. The SSP condition was ungrammatical and could yield longer reading times than the PSP condition starting at Region 3.

items only. Post hoc paired *t* tests conducted to explore the potential interaction revealed significantly longer reading times for the SPS condition at Region 4 only among the L1 group, $t_1(31) = 2.41$, $p = .022$, $t_2(39) = 1.94$, $p = .060$, but not among the near-native L2 group, $t_1(23) = 0.44$, $p = .665$, $t_2(39) = 0.62$, $p = .537$, or the advanced L2 group, $t_1(19) = 0.59$, $p = .563$, $t_2(39) = 0.75$, $p = .457$. At Region 5, the sentence-final region, there were no significant effects. Finally, the reaction times and accuracy scores from the comprehension questions that followed the attraction stimuli both showed main effects of group that were significant by items only (i.e., were approaching significance), indicative of a trend toward faster and more accurate responses among the near-native group, particularly as compared to the L1 group.

To summarize the results from the agreement attraction stimuli, an immediate reading time effect on the verb in Region 3 was evident only

Table 3. Nonlocal grammaticality stimuli: Mean accuracy and RT for comprehension questions (*SD*s in parentheses)

	Accuracy		RT (ms)	
Participant group	PSP	SSP	PSP	SSP
Native	.856 (.101)	.878 (.113)	4,412 (1,154)	4,311 (996)
Near native	.891 (.102)	.900 (.102)	3,855 (792)	3,840 (857)
Advanced	.805 (.176)	.820 (.136)	4,125 (562)	4,350 (1,124)

Note. RT = reaction time; PSP = Plural-Singular-Plural; SSP = Singular-Singular-Plural.

https://doi.org/10.1017/S027226311400059X Published online by Cambridge University Press

Table 4. Attraction stimuli: Mixed design ANOVAs

	By subjects			By items		
Effect	*df*	F_1	*p*	*df*	F_2	*p*
Region 3						
Group	2, 73	0.14	.869	2, 38	0.52	.595
Agreement	1, 73	1.30	.259	1, 39	0.69	.411
Group × Agreement	2, 73	5.37	**.007**	2, 38	2.74	.071
Region 4						
Group	2, 73	0.17	.842	2, 38	0.13	.876
Agreement	1, 73	1.18	.281	1, 39	0.63	.434
Group × Agreement	2, 73	1.80	.172	2, 38	2.68	.082
Region 5						
Group	2, 73	0.34	.714	2, 38	1.33	.277
Agreement	1, 73	1.76	.188	1, 39	0.71	.406
Group × Agreement	2, 73	0.73	.486	2, 38	0.94	.401
RTs comprehension questions						
Group	2, 73	1.93	.153	2, 38	8.47	**.001**
Agreement	1, 73	0.14	.708	1, 39	0.09	.772
Group × Agreement	2, 73	0.28	.754	2, 38	0.80	.458
Comprehension accuracy						
Group	2, 73	1.46	.240	2, 38	3.46	**.042**
Agreement	1, 73	3.48	.066	1, 39	0.94	.338
Group × Agreement	2, 73	1.55	.218	2, 38	2.34	.111

Note. Effects significant at α = .05 appear in boldface. RT = reaction time.

among the near-native L2 group, and a spillover effect on the postverbal Region 4 was exhibited only by the L1 group. The advanced L2 participant group, on the other hand, did not show the effect of attraction at any point in the stimuli or on the poststimulus comprehension questions. The results of the ANOVAs for the experimental sentences that tested for nonlocal grammaticality effects are provided in Table 5. At Region 3, the region with the verb, only the main effect of grammaticality reached significance, which indicates that reading times in general were slower for ungrammatical sentences at this point. At Region 4, the postverbal or spillover region, the main effect of grammaticality was significant only by subjects, and the Group × Grammaticality interaction was marginally significant only by items (i.e., neither effect was fully significant). Post hoc paired *t* tests conducted to explore the possible interaction showed signs of a spillover grammaticality effect that was significant by items only for the L1 group, $t_1(31) = 1.40$, $p = .172$, $t_2(39) = 3.28$, $p = .002$, and by subjects only for the near-native group, $t_1(23) = 2.47$, $p = .021$, $t_2(39) = 1.05$, $p = .301$. The effect was not significant for the advanced group, $t_1(19) = 0.13$, $p = .900$, $t_2(39) = 0.50$, $p = .619$. At Region 5, the sentence-final region, the Group × Agreement interaction was again significant by items

https://doi.org/10.1017/S027226311400059X Published online by Cambridge University Press

Table 5. Nonlocal grammaticality stimuli: Mixed design ANOVAs

	By subjects			By items		
Effect	*df*	F_1	*p*	*df*	F_2	*p*
Region 3						
Group	2, 73	0.05	.954	2, 38	0.82	.444
Grammaticality	1, 73	8.76	**.004**	1, 39	7.72	**.008**
Group × Grammaticality	2, 73	1.00	.372	2, 38	1.53	.222
Region 4						
Group	2, 73	0.07	.933	2, 38	0.40	.672
Grammaticality	1, 73	4.91	**.030**	1, 39	2.49	.122
Group × Grammaticality	2, 73	2.25	.112	2, 38	3.11	.056
Region 5						
Group	2, 73	0.12	.887	2, 38	1.20	.314
Grammaticality	1, 73	0.95	.334	1, 39	1.65	.207
Group × Grammaticality	2, 73	1.95	.149	2, 38	3.24	**.050**
RTs comprehension questions						
Group	2, 73	2.60	.081	2, 38	21.44	**.000**
Grammaticality	1, 73	0.13	.719	1, 39	0.57	.457
Group × Grammaticality	2, 73	0.88	.418	2, 38	1.90	.164
Comprehension accuracy						
Group	2, 73	4.69	**.012**	2, 38	5.26	**.010**
Grammaticality	1, 73	0.65	.422	1, 39	1.62	.210
Group × Grammaticality	2, 73	0.05	.952	2, 38	0.10	.904

Note. Effects significant at α = .05 appear in boldface. RT = reaction time.

only (i.e., approaching significance). Post hoc paired *t* tests conducted to further examine the potential interaction revealed no significant effects among any of the three participant groups (L1 group: $t_1[31] = 1.06$, $p = .298$, $t_2[39] = 0.91$, $p = .370$; near-native group: $t_1[23] = 1.68$, $p = .107$, $t_2[39] = 1.64$, $p = .109$; advanced group: $t_1[19] = 0.55$, $p = .586$, $t_2[39] = 1.52$, $p = .137$). Finally, analyses performed on the reaction time and accuracy data from the comprehension questions that followed the nonlocal grammaticality stimuli revealed main effects of group, which were marginally significant by subjects and significant by items for the reaction times and fully significant by subjects and by items for the accuracy scores. One-way ANOVAs were conducted to examine the simple effects of group, but these proved significant only for the accuracy scores, $F(2, 80) = 4.69$, $p = .012$, in which a post hoc Scheffé test showed that the near-native group had significantly higher accuracy scores than the advanced group for the comprehension questions that followed the grammaticality stimuli, $p = .013$, with neither of the other two comparisons reaching significance.

To summarize the results for the nonlocal grammaticality stimuli, the participants as a whole showed immediate effects of grammaticality at

https://doi.org/10.1017/S027226311400059X Published online by Cambridge University Press

the verb in Region 3. In addition, the near-native and native groups also showed signs of a spillover effect at the postverbal Region 4, although these were not fully significant by subjects and by items. There was no significant extended effect of grammaticality at sentence wrap-up in Region 5, although the near natives had a numerical trend in that direction.

DISCUSSION AND CONCLUSION

The first research question for the present study asked whether nonnative readers exhibited nativelike agreement attraction effects while processing verbal number in Spanish. Interestingly, the near-native participants in the present study appeared to be even more immediately sensitive to the attraction manipulation than the native speakers. Specifically, the near-native group showed an immediate reading time effect on the verb, whereas the native speakers showed only a delayed or spillover effect. Compared with previous research that had documented immediate sensitivity to the attraction manipulation among native speakers of English (Jiang, 2004; Pearlmutter et al., 1999) and Spanish (Acuña-Fariña et al., 2014), the native Spanish readers in the present study were slightly slower in reacting to the manipulation of number agreement, whereas the results from the near-native participants were actually more consistent with these prior L1 studies of agreement attraction. One interpretation of this outcome is that it reflects L1 interference and crosslinguistic variation, in which agreement attraction is not as likely to affect native processing in a rich agreement language such as Spanish as it would be in English. Such an interpretation is consistent with the notion of *morphological filtering* proposed for agreement phenomena in speech production, in which richer agreement morphology appears to diminish attraction and other agreement aberrations (Eberhard, Cutting, & Bock, 2005; Foote & Bock, 2012; Vigliocco, Butterworth, & Semenza, 1995; Vigliocco & Franck, 1999, 2001). It is also consistent with the noisy channel model proposed by Bergen and Gibson (2012), which also would predict that multiple markers of number on a head noun (i.e., *el testigo*/*los testigos* in Spanish versus *the witness*/*the witnesses* in English) would be less likely to be erroneously perceived via noise than a single marker and therefore less likely to be corrected for by the parser. Still, even though crosslinguistic variation in agreement attraction may well lead to different processing tendencies in native English versus native Spanish, it seems unlikely that the difference between the native and near-native groups in the present study was due to L1 influence, as the reading time effect observed among the near natives on the region with the verb was very similar to that previously observed among native speakers of Spanish using the eye-tracking method (Acuña-Fariña et al., 2014).

https://doi.org/10.1017/S027226311400059X Published online by Cambridge University Press

The observed effects for attraction among both the native and near-native participants could also potentially be accounted for by the entirely alternative explanation proposed by Wagers et al. (2009). Under their interpretation, increased reading times on the verb in the SPS versus the SSS condition would be due not to the complication of agreement processes by the number feature of the intervening noun but rather to a spillover effect of the previous region. In other words, the reading time effect would arise from the lower level processing of a plural versus singular NP, rather than from higher level computations of agreement between subject and verb. It is certainly true that the plural NP took longer to read than the singular NP; the difference can be seen in Region 2 among all three participant groups in Figures 1, 2, and 3, and the main effect of agreement was also highly significant at Region 2 both by subjects and by items (ps < .001, Fs > 15.00, no interaction with group). However, spillover of the effect of agreement from the intervening NP to the region with the verb does not seem likely here, particularly in the case of the L1 participants, who showed the effect at the preverbal Region 2, then showed no effect for agreement on the verb in Region 3, and then again showed an effect for agreement on the postverbal Region 4. In other words, it would be unusual for a reading time effect to appear immediately on the critical region, then disappear on the following region, and then resurface as a delayed spillover on the next region. Spillover effects in self-paced reading are more typically continuous, and reading time effects usually resurface during sentence wrap-up, if at all. Spillover from the plural NP in Region 2 clearly did not occur among the advanced L2 group in the present study, who showed no reliable reading time effects after Region 2. In the case of the near-native L2 group, on the other hand, it does seem that spillover from the intervening NP is more plausible because the effect of agreement was present at the NP in Region 2 and on the verb in Region 3. However, the same near-native participants did not show any such spillover effect with the grammaticality stimuli, as seen in Figure 5, even though a similarly significant main effect of NP plurality was seen across all groups on Region 1 (ps < .001, Fs > 15.00, no interaction with group). Furthermore, Wagers et al. (2009) proposed that the agreement attraction effect observed on the verb during self-paced reading by Pearlmutter et al. (1999, Experiment 1) was actually a spillover effect from the preceding region with the intervening noun in part because there was no reading time effect evident on that preceding region, which was unexpected. What is more, in the current investigation, the expected reading time difference was immediately evident on the NP that varied with regard to number in Region 2. Thus, although the results from the present study are not entirely conclusive in refuting the retrieval-based account of Wagers et al. (2009), they do strongly suggest that the reading time effect that was observed among the native and near-native participants reflects higher level processing of subject-verb

https://doi.org/10.1017/S027226311400059X Published online by Cambridge University Press

number agreement, as opposed to a spillover of lower level processing effects from a plural versus singular intervening noun in Region 2, and that the near-native L2 group exhibited nativelike processing of agreement. Particularly when taken in conjunction with the number attraction effects observed during native Spanish comprehension by Acuña-Fariña et al. (2014), the outcome of the present study is more consistent with either a representational account (Bock & Eberhard, 1993; Nicol et al., 1997) or the noisy channel account (Bergen & Gibson, 2012).

One final alternative interpretation of the agreement attraction data as it pertains to the first research question is that, although the reading time patterns between the L1 and near-native participants were similar, the two groups actually employed different processing strategies. Such an interpretation was proposed by Jiang (2004). In the attraction paradigm, reading time effects on (or after) the verb in a SPS stimulus such as *The witness of the lawyers has a bad reputation in the neighborhood* are assumed to arise because the number feature of the intervening NP *the lawyers* interferes in the computation of number agreement between the head noun and the verb. On the other hand, a similar effect might obtain in shallow or otherwise limited L2 processing if the nonnative reader were limited to the computation of only local agreement. In the previous example, longer reading times could also be evident on SPS sequences such as *The witness of the lawyers has* if the nonnative processer were to employ some type of local noun strategy, computing number agreement with the nearest NP *the lawyers* rather than with the more distant head noun *the witness*, because in the SSS condition this intervening noun agrees with the verb, and in the SPS condition it does not. Such a reading time effect would arise from the perception of a grammatical violation where none exists. To put it another way, the stimulus manipulation for the attraction paradigm only involves the number feature of the immediately preverbal NP, so there is no guarantee that the resulting reading time effects involve the processing of nonlocal agreement. However, the nonlocal grammaticality stimuli that were also included in the present study did involve a manipulation of a number feature that was not immediately adjacent to the verb in the linear structure of the stimuli, and they thus serve to corroborate the interpretation that the processing strategy of the near-native participants was nativelike and not based on shallow representations of grammar. This outcome is broadly consistent with previous observations of nativelike L2 speech production with regard to agreement attraction (Foote, 2010; Hoshino, Dussias, & Kroll, 2010).

The second research question guiding the current investigation asked if nonnative speakers show nativelike sensitivity to the grammaticality of nonlocal verbal number agreement during online sentence comprehension. As previously discussed, the main purpose of this second set of stimuli was to complement the first set with a manipulation of verbal

https://doi.org/10.1017/S027226311400059X Published online by Cambridge University Press

number that affected the head noun rather than the embedded local noun to ensure the computation of agreement over linear distance. The results from the native and near-native participants were again quite similar overall, as both groups displayed sensitivity to the grammaticality of number agreement immediately (i.e., at the stimulus region containing the verb) and a trend toward a spillover effect in the following sentence region.

The third and final research question for the present study sought to examine whether the processing of verbal number agreement develops over time, in conjunction with L2 proficiency. The results from the current investigation do support the claim that processing behavior changes as proficiency in the L2 increases because there was a clear difference in the behavior of the near-native and advanced participant groups; whereas the near-native group showed nativelike effects for attraction and the grammaticality of nonlocal agreement, the advanced group showed no effects with the attraction stimuli and only a minimal trend with the grammaticality stimuli that was statistically similar to the other two groups in the omnibus ANOVA but probably would not hold on its own. It can be inferred from this outcome that online sensitivity to verbal number agreement varies greatly according to proficiency level and possibly also according to the type of sentence processing paradigm used to test agreement. Moreover, it can also be inferred that the advanced participants did not simply read over the few letters that distinguished plural from singular; they had nativelike longer reading times on plural NPs than on singular NPs in both Region 2 of the attraction stimuli and Region 1 of the nonlocal grammaticality stimuli, because each of the regions in question was longer in terms of characters than in the plural condition (e.g., *el testigo* vs. *los testigos*) and/or because plural nouns in general take longer to process. Thus, at the level of proficiency of the advanced participants, there is basic, lower level processing of number morphology, minimally at the level of basic orthography, but higher level linguistic processes are not fully engaged until the learner reaches a higher level of proficiency. In terms of grammatical knowledge, the advanced-level subjects demonstrated nativelike knowledge of verbal agreement in the offline acceptability judgment task (although even there they scored statistically lower than the near-native group and numerically lower than the native group) but did not appear to engage in the type of online computations of nonlocal number agreement that would give rise to agreement attraction effects or unmistakable grammaticality effects during an online task such as self-paced reading.

The overall outcome of the present study is thus consistent with several previous self-paced reading and ERP investigations that had documented nativelike sensitivity to errors of number agreement, manifest either as longer reading times or as P600 effects, among highly proficient nonnative readers (Coughlin & Tremblay, 2013; Foote, 2011; Rossi et al., 2006;

https://doi.org/10.1017/S027226311400059X Published online by Cambridge University Press

Sagarra & Herschensohn, 2011; Tanner et al., 2012; Tanner et al., 2009). The current investigation thereby expands the range of nativelike agreement processing to include a context in which there is no grammar violation. Still, this outcome also stands in contrast with the results of the one previous L2 study that had also employed the agreement attraction paradigm and had found an overall insensitivity to English verbal number morphology during L2 processing (Jiang, 2004). Differences in either the participants' native languages (Chinese vs. English) or their proficiency levels might account for the discrepancy between Jiang's results and those from the current experiment. The potentially limiting role of the L1 remains an open question that can only be resolved for certain by further research on subject-verb number agreement, but a self-paced reading study by Jiang, Novokshanova, Masuda, and Wang (2011) of the processing of plural morpheme errors (e.g., *They discovered one of the dinosaurs/*dinosaur* . . .) in ESL showed that L1-L2 morphological congruency can affect online sensitivity to number morphemes, so there is good reason to think the L1 is a factor. In addition, the outcome of the present study underscores the critical role of proficiency in the processing of verbal number agreement. Jiang's (2004) participants were fairly characterized as "proficient" (p. 612), on the basis of TOEFL scores and enrollment at a university in the United States, but they appear to have been potentially less proficient than even the advanced group in the present study, on the basis of L2 self-proficiency ratings of 5.3–5.8 versus 6.8–7.7 (maximum 10) and mean length of residency of 1.4 versus 6.0 years (although the Jiang et al., 2011, participants had residency similar to the near-native group in the present study). Therefore, there remains the possibility that learners with native languages such as Chinese could acquire more nativelike online sensitivity to subject-verb number agreement at a more advanced stage of L2 development.

A second interesting outcome of the current experiment was that the nonlocal grammaticality stimuli provoked slightly more robust effects than the agreement attraction stimuli, meaning that the effect was more immediate than with the attraction stimuli, in the case of the L1 group, and that the effect also appeared to spill over into the postverbal region with both the native and near-native groups. It seems that a grammar error somehow presents a greater challenge in processing, perhaps because the parser has difficulty finding a subject when neither of the preverbal NPs agrees with the verb. Another explanation for the asymmetry between the two stimulus types is that the processing of the grammar violation may not be more difficult in and of itself, but the presence of an easily identifiable error is distracting to the reader. Especially among L2 learners, an error has the potential to foster conscious reflection, which would entail the activation of explicit metalinguistic knowledge that may otherwise be inactive (or relatively less active) during a meaning-based comprehension task such as that employed

https://doi.org/10.1017/S027226311400059X Published online by Cambridge University Press

during the self-paced reading experiment. The results thus appear to corroborate the a priori proposal of the present study that the agreement attraction paradigm is more subtle than the grammar violation paradigms that are more often employed for L2 research and therefore can yield somewhat different results. It is important to note, however, that this difference may well affect only reading time measures and not ERPs, as conscious attention to L2 agreement errors while reading has been shown to cause a self-paced reading time effect that otherwise did not occur (Leeser et al., 2011), but the P600 effect seen in ERP research on agreement may be less dependent on concurrency with a metalinguistic task (Osterhout, Allen, McLaughlin, & Inoue, 2002).

Finally, the results of the current experiment are also relevant to grammar-based L2 processing theories such as the shallow structure hypothesis, which specifically predicts that the type of behavior observed herein would not occur. To be sure, the data from either of the two stimuli sets viewed in isolation could be interpreted as consistent with some type of grammatically deficient processing, if not entirely consistent with shallow structure in particular. To illustrate, the slower reading times observed on the verb in attraction stimuli with sequences such as *The witness of the lawyers has* versus *The witness of the lawyer has* could still occur if the parser were to employ a local noun strategy in shallow processing, attempting to compute verbal number agreement only between the adjacent elements in *the lawyers has*. An alternative account of the reading time effects observed with the nonlocal grammaticality stimuli is that the parser could have employed a first-noun strategy, correctly computing number and reacting to grammatical errors in stimuli with sequences such as *The witness of the lawyer have* because the first noun also happens to be the head noun of the complex NP subject that carries the number feature that must agree with the verb. However, when the results from both stimuli sets are taken together, these explanations are no longer plausible because the nonnative parser would have to employ wholly different compensatory processing strategies on two types of sentences that are syntactically quite similar. It is therefore difficult to imagine how a shallow or otherwise grammatically deficient L2 processing strategy could yield the reading time patterns observed in this experiment.

There are some methodological limitations to the present investigation that are worth noting. First, the participant groups were not closely matched with regard to age and level of education. The nonnative groups were, in general, older and included many individuals with graduate degrees, whereas the native group was composed almost entirely of undergraduate students. This difference in education level might explain why the native readers did not exhibit generally faster reading times than the L2 participants[2] (as is typically the case in studies of L2 processing), if the nonnative participants had more reading experience

https://doi.org/10.1017/S027226311400059X Published online by Cambridge University Press

and therefore were more fluent readers than the native participants. More important, a difference in reading fluency could also explain why the native readers appeared to be less immediately sensitive to the experimental manipulation of verbal number in the attraction stimuli than the near-native readers. The age difference in the participants, on the other hand, does not appear to have affected the results, as the older L2 participants would have been expected to read more slowly, which was not the case. Finally, the fact that the L1 comparison group was monolingual introduced a second participant variable that could potentially be confounded with age of acquisition (i.e., nativeness). In other words, given that the L1 group was monolingual and acquired Spanish in childhood, whereas the L2 groups were multilingual and acquired Spanish in adulthood, it is possible that the observed differences are due not to differences in age of acquisition but rather to the difference in multilingual status or perhaps to a combination of both variables. However, the overall similarity between the L1 and near-native L2 groups suggests that the effects of bilingualism (either crosslinguistic influence from the L1 or reduced processing efficiency because of the cognitive effort required to suppress the L1 during L2 processing) either are not important to the type of processing measured in the present study or cease to become important at sufficiently high levels of proficiency. Further investigation is needed to resolve these limitations.

To conclude, the overall outcome of the present investigation suggests that nonnative readers can come to process agreement in a nativelike manner, even over linear distance and even as measured by a stimulus manipulation that is potentially less evident to L2 learners than error recognition, and that this behavior develops only after reaching a sufficiently high level of proficiency in the L2. More broadly speaking, the results reported herein, particularly when considered in conjunction with previous psycholinguistic study of L2 number agreement phenomena, suggest that nonnative processing is not necessarily shallow or lacking syntactic detail and that morphosyntactic aspects of processing are not necessarily the primary source of L1-L2 differences. Although further investigation on other grammatical aspects of sentence comprehension is still needed to develop a more complete description of how processing is accomplished in a nonnative language, it seems that the other two potential sources of L2 divergence currently under consideration (i.e., cognitive limitations such as working memory and interference from the L1) show the potential to be especially informative with regard to L1-L2 differences and thus should also be the target of considerable future research in the area of L2 psycholinguistics.

Received 28 March 2014
Accepted 20 June 2014
Final Version Received 30 June 2014

https://doi.org/10.1017/S027226311400059X Published online by Cambridge University Press

NOTES

1. It is worth noting that Spanish does not have an equivalent of the synthetic genitive construction with *'s* that is used in English, so constructions such as *the witness of the lawyer* do not sound awkward or unusual in Spanish because the analytic genitive is the only option.

2. The similarity of overall reading speed between the L1 and near-native groups was shown both by a lack of main effect of group in the primary ANOVAs and by no difference in global reading times calculated for each individual participant on the basis of the entire self-paced reading experiment, $t(55) = 0.78$, $p = .437$.

REFERENCES

Acuña-Fariña, J. C., Meseguer, E., & Carreiras, M. (2014). Gender and number agreement in comprehension in Spanish. *Lingua*, *143*, 108–128.

Bergen, L., & Gibson, E. (2012, March). *Agreement errors as rational encoding errors*. Poster presented at the 25th annual CUNY Conference on Human Sentence Processing, New York, NY.

Bock, J. K., Carreiras, M., & Meseguer, E. (2012). Number meaning and number grammar in English and Spanish. *Journal of Memory and Language*, *66*, 17–37.

Bock, J. K., & Eberhard, K. M. (1993). Meaning, sound, and syntax in English number agreement. *Language and Cognitive Processes*, *8*, 57–99.

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

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

Clahsen, H., & Felser, C. (2006b). How native-like is nonnative language processing? *Trends in Cognitive Sciences*, *10*, 564–570.

Clahsen, H., Felser, C., Neubauer, K., Sato, M., & Silva, R. (2010). Morphological structure in native and nonnative language processing. *Language Learning*, *60*, 21–43.

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

Eberhard, K. M. (1999). The accessibility of conceptual number to the processes of subject-verb agreement in English. *Journal of Memory and Language*, *41*, 560–578.

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

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, 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.

Frenck-Mestre, C., & Pynte, J. (1997). Syntactic ambiguity resolution while reading in second and native languages. *Quarterly Journal of Experimental Psychology: Section A*, *50*, 119–148.

Hopp, H. (2006). Syntactic features and reanalysis in near-native processing. *Second Language Research*, *22*, 369–397.

Hopp, H. (2010). Ultimate attainment in L2 inflection: Performance similarities between non-native and native speakers. *Lingua*, *120*, 901–931.

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.

Instituto Cervantes. (2007). *Diplomas de Español como Lengua Extranjera: Nivel Superior* [Certificates of Spanish as a foreign language: Superior level] [Spanish language test]. Retrieved from http://diplomas.cervantes.es/en/general-information/level-c2.html

Jackson, C. N. (2010). The processing of subject-object ambiguities by English and Dutch L2 learners of German. In B. VanPatten & J. Jegerski (Eds.), *Research in second language processing and parsing* (pp. 207–230). Amsterdam, the Netherlands: Benjamins.

https://doi.org/10.1017/S027226311400059X Published online by Cambridge University Press

Jiang, N. (2004). Morphological insensitivity in second language processing. *Applied Psycholinguistics*, *25*, 603–634.
Jiang, N. (2007). Selective integration of linguistic knowledge in adult second language learning. *Language Learning*, *57*, 1–33.
Jiang, N., Novokshanova, E., Masuda, K., & Wang, X. (2011). Morphological congruency and the acquisition of L2 morphemes. *Language Learning*, *61*, 940–967.
Juffs, A. (1998). Some effects of first language argument structure and syntax on second language processing. *Second Language Research*, *14*, 406–424.
Leeser, M., Brandl, A., & Weissglass, C. (2011). Task effects in second language sentence processing research. In P. Trofimovich & K. McDonough (Eds.), *Applying priming methods to L2 learning, teaching, and research: Insights from psycholinguistics* (pp. 179–198). Amsterdam, the Netherlands: Benjamins.
Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The language experience and proficiency questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research*, *50*, 940–967.
Montrul, S., & Slabakova, R. (2003). Competence similarities between native and near-native speakers: An investigation of the preterite-imperfect contrast in Spanish. *Studies in Second Language Acquisition*, *25*, 351–398.
Nicol, J., Forster, K. I., & Veres, C. (1997). Subject-verb agreement processes in comprehension. *Journal of Memory and Language*, *36*, 569–587.
Ojima, S., Nakata, H., & Kakigi, R. (2005). An ERP study of second language learning after childhood: Effects of proficiency. *Journal of Cognitive Neuroscience*, *17*, 1212–1228.
Osterhout, L., Allen, M. D., McLaughlin, J., & Inoue, K. (2002). Brain potentials elicited by prose-embedded linguistic anomalies. *Memory & Cognition*, *30*, 1304–1312.
Pearlmutter, N. J., Garnsey, S. M., & Bock, K. (1999). Agreement processes in sentence comprehension. *Journal of Memory and Language*, *41*, 427–456.
Pienemann, M. (1998). *Language processing and second language development: Processability theory*. Amsterdam, the Netherlands: Benjamins.
Rossi, S., Gugler, M. F., Friederici, A. D., & Hahne, A. (2006). The impact of proficiency on syntactic second-language processing of German and Italian: Evidence from event-related potentials. *Journal of Cognitive Neuroscience*, *18*, 2030–2048.
Sagarra, N., & Herschensohn, J. (2011). Asymmetries in gender and number agreement processing in late bilinguals. In L. A. Ortiz-López (Ed.), *Selected proceedings of the 13th Hispanic Linguistics Symposium* (pp. 169–177). Somerville, MA: Cascadilla Proceedings Project.
Segalowitz, N. (2003). Automaticity and second languages. In C. Doughty & M. Long (Eds.), *The handbook of second language acquisition* (pp. 382–388). Oxford, UK: Blackwell.
Sorace, A. (2005). Syntactic optionality at interfaces. In L. Cornips & K. Corrigan (Eds.), *Syntax and variation: Reconciling the biological and the social* (pp. 46–111). Amsterdam, the Netherlands: Benjamins.
SuperLab (Version 4.0.5) [Computer software]. San Pedro, CA: Cedrus Corporation.
Tanner, D., Inoue, K., & Osterhout, L. (2014). Brain-based individual differences in on-line L2 grammatical comprehension. *Bilingualism: Language and Cognition*, *17*, 277–293.
Tanner, D., Nicol, J., Herschensohn, J., & Osterhout, L. (2012). Electrophysiological markers of interference and structural facilitation in native and nonnative agreement processing. In A. K. Biller, E. Y. Chung, & A. E. Kimball (Eds.), *Proceedings of the 36th Boston University Conference on Language Development* (pp. 594–606). Somerville, MA: Cascadilla Press.
Tanner, D., Osterhout, L., & Herschensohn, J. (2009). Snapshots of grammaticalization: Differential electrophysiological responses to grammatical anomalies with increasing L2 exposure. In J. Chandlee, M. Franchini, S. Lord, & G.-M. Rheiner (Eds.), *Proceedings of the 33rd Boston University Conference on Language Development* (pp. 528–539). Somerville, MA: Cascadilla Press.
VanPatten, B. (1996). *Input processing and grammar instruction: Theory and research*. Norwood, NJ: Ablex.
VanPatten, B. (2007). Input processing in adult second language acquisition. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction*. Mahwah, NJ: Erlbaum.
Vigliocco, G., Butterworth, B., & Garrett, M. F. (1996). Subject-verb agreement in Spanish and English: Differences in the role of conceptual constraints. *Cognition*, *61*, 261–298.

https://doi.org/10.1017/S027226311400059X Published online by Cambridge University Press

Vigliocco, G., Butterworth, B., & Semenza, C. (1995). Constructing subject-verb agreement in speech: The role of semantic and morphological factors. *Journal of Memory and Language*, *34*, 186–215.
Vigliocco, G., & Franck, J. (1999). When sex and syntax go hand in hand: Gender agreement in language production. *Journal of Memory and Language*, *40*, 455–478.
Vigliocco, G., & Franck, J. (2001). When sex affects syntax: Contextual influences in sentence production. *Journal of Memory and Language*, *45*, 368–390.
Vigliocco, G., Hartsuiker, R. J., Jarema, G., & Kolk, H. H. (1996). One or more labels on the bottles? Notional concord in Dutch and French. *Language and Cognitive Processes*, *11*, 407–442.
Wagers, M. W., Lau, E. F., & Phillips, C. (2009). Attraction in comprehension: Representations and processes. *Journal of Memory and Language*, *61*, 206–237.
Witzel, J., Witzel, N., & Nicol, J. (2012). Deeper than shallow: Evidence for structure-based parsing biases in second-language sentence processing. *Applied Psycholinguistics*, *33*, 419–456.

https://doi.org/10.1017/S027226311400059X Published online by Cambridge University Press