Studies in Second Language Acquisition, 2013, **35**, 1–35.
doi:10.1017/S0272263112000915

WHEN GENDER AND LOOKING GO HAND IN HAND

Grammatical Gender Processing In L2 Spanish

Paola E. Dussias
Penn State University, Center for Language Science

Jorge R. Valdés Kroff
University of Pennsylvania

Rosa E. Guzzardo Tamargo
Universidad de Puerto Rico

Chip Gerfen
American University

In a recent study, Lew-Williams and Fernald (2007) showed that native Spanish speakers use grammatical gender information encoded in Spanish articles to facilitate the processing of upcoming nouns. In this article, we report the results of a study investigating whether grammatical gender facilitates noun recognition during second language (L2) processing. Sixteen monolingual Spanish participants (control group)

The writing of this article was supported in part by the National Science Foundation (NSF) grant BCS-0821924 to Paola E. Dussias and Chip Gerfen, by NSF grants BCS-0955090 and OISE-0968369 to Judith F. Kroll and Paola E. Dussias, and by a National Science Foundation graduate research fellowship to Jorge Valdés Kroff. We would like to thank the audience at Penn State's Center for Language Science and the audience at the 2010 Second Language Research Forum for helpful comments. All errors are our own.

Correspondence concerning this article should be addressed to Paola E. Dussias, 237 Burrowes Building, Department of Spanish, Italian & Portuguese, Penn State University, University Park, PA 16802; e-mail: pdussias@psu.edu.

and 18 English-speaking learners of Spanish (evenly divided into high and low Spanish proficiency) saw two-picture visual scenes in which items matched or did not match in gender. Participants' eye movements were recorded while they listened to 28 sentences in which masculine and feminine target items were preceded by an article that agreed in gender with the two pictures or agreed only with one of the pictures. An additional group of 15 Italian learners of Spanish was tested to examine whether the presence of gender in the first language (L1) modulates the degree to which gender is used during L2 processing. Data were analyzed by comparing the proportion of eye fixations on the objects in each condition. Monolingual Spanish speakers looked sooner at the referent on different-gender trials than on same-gender trials, replicating results reported in past literature. Italian-Spanish bilinguals exhibited a gender anticipatory effect, but only for the feminine condition. For the masculine condition, participants waited to hear the noun before identifying the referent. Like the Spanish monolinguals, the highly proficient English-Spanish speakers showed evidence of using gender information during online processing, whereas the less proficient learners did not. The results suggest that both proficiency in the L2 and similarities between the L1 and the L2 modulate the usefulness of morphosyntactic information during speech processing.

Relative to the extensive psycholinguistic literature on gender agreement during lexical and syntactic processing in monolinguals (Barber & Carreiras, 2005; Bates, Devescovi, Hernández, & Pizzamiglio, 1996; Colé & Segui, 1994; Costa, Kovacic, Fedorenko, & Caramazza, 2003; Cubelli, Lotto, Paolieri, Girelli, & Job, 2005; Dahan, Swingley, Tanenhaus, & Magnuson, 2000; Faussart, Jakubowicz, & Costes, 1999; Friederici & Jacobsen, 1999; Grosjean, Dommergues, Cornu, Guillelmon, & Besson, 1994; Gurjanov, Lukatela, Lukatela, Savić, & Turvey, 1985; Jacobsen, 1999; Jakubowicz & Faussart, 1998; Jescheniak, 1999; Schmidt, 1986; van Berkum, 1996), surprisingly little is known about how grammatical gender constrains processing in a second language (L2). Most studies in the SLA tradition have utilized gender agreement not to investigate processing per se but rather as a tool for adjudicating between competing theories within generative approaches to SLA (Bruhn de Garavito & White, 2002; Carroll, 1989; Franceschina, 2005; Hawkins & Chan, 1997; McCarthy, 2007; Montrul, 2004; Montrul, Foote, & Perpiñan, 2008; White, Valenzuela, Kozlowska-MacGregor, & Leung, 2004). Central to this debate is whether adult L2 learners are able to acquire abstract grammatical features in the L2 that are not instantiated in the first language (L1; e.g., Duffield & White, 1999; Epstein, Flynn, & Martohardjono, 1996; Schwartz & Sprouse, 1994, 1996; White, 1989, 2003). The results reveal important asymmetries between native and nonnative speakers, but the task that participants must complete modulates these results. For oral or written production tasks, L2 speakers show persistent problems with gender agreement, consistently performing below the native speaker mark (Carroll, 1989; Fernández-García, 1999; Franceschina, 2005; Hawkins & Franceschina, 2004; McCarthy, 2007). This also appears to be true for speakers who have been immersed in the L2 environment for lengthy periods of time (e.g., Franceschina, 2001). However, with comprehension-based tasks, the results are less clear. Some studies show that performance by L2 speakers is indistinguishable from that of native speakers (e.g., Gess & Herschensohn, 2001; White et al., 2004), whereas others report behavioral differences between the two groups that suggest underlying representational differences (e.g., Montrul et al., 2008).

Nevertheless, one striking generalization stands out: Specifically, across numerous studies, the accuracy rate for gender assignment and gender agreement is quite high, most often ranging between 80 and 90%. This is true for speakers with and without grammatical gender in their L1 (e.g., the learners of Dutch in Sabourin, Stowe, & de Haan, 2006; the learners of Spanish in White et al., 2004). Even those studies that argue most strongly for fundamental representational differences between native and nonnative speakers report accuracy rates above the 75–80% mark. If researchers shift their focus away from evaluating knowledge of grammatical gender in learners solely in terms of native speaker competence, they are left with the realization that learners do, in fact, exhibit a high degree of knowledge of the target language gender system, even if that knowledge is not identical to that of native speakers, who perform at ceiling on the tasks that have been employed. This, in turn, suggests that important questions remain if scholars are to reach a more comprehensive understanding of what L2 speakers can do with the gender system of the L2. Most relevant here is the observation that the relatively high level of accuracy exhibited by L2 learners leads us to ask whether learners access and use the knowledge they have acquired about grammatical gender in their L2 during online processing.

To address this question, we explore whether adult L2 speakers of Spanish process gender-marked articles similarly to native speakers. In a recent study, Lew-Williams and Fernald (2007) showed that native Spanish speakers use grammatical gender information encoded in Spanish articles to facilitate the processing of upcoming nouns. Building on this finding, we asked whether L2 speakers of Spanish whose L1 (English) lacks grammatical gender can exploit gender information in articles to speed up lexical processing. An additional group of Italian learners of Spanish was also tested to investigate whether having gender in the L1 modulates the degree to which gender is used during L2 processing. By expanding the scope of gender agreement research to processing, we argue, it becomes possible to contribute significantly to the debate over whether and how grammatical gender is acquired during adult L2 learning.

GRAMMATICAL GENDER PROCESSING IN A L2

One important area of investigation in the L1 psycholinguistic literature has focused on whether grammatical gender modulates lexical processing under the hypothesis that the presence of overt gender marking may facilitate the processing of subsequent gender-marked target items (e.g., Barber & Carreiras, 2005; Bates et al., 1996; Carello, Lukatela, & Turvey, 1988; Colé & Segui, 1994; Costa et al., 2003; Dahan et al., 2000; Faussart et al., 1999; Grosjean et al., 1994; Gurjanov et al., 1985; Hagoort & Brown, 1999; Jacobsen, 1999; Jakubowicz & Faussart, 1998; Lew-Williams & Fernald, 2007; Schmidt, 1986; van Berkum, 1996). For example, when speakers produce noun phrases (e.g., determiner + noun combinations, such as *la manzana* "the $_{\text{FEM}}$ apple $_{\text{FEM}}$") in response to a target picture, the presence of a distractor word printed on the picture that matches the picture in gender (e.g., *camisa* "shirt$_{\text{FEM}}$") yields shorter naming latencies. This so-called gender congruency effect has also been observed in comprehension studies. For example, Grosjean et al. (1994) found that when a French gender-marked article preceded a noun, listeners needed to hear less of the noun to identify it and were significantly more confident in their choices than when the gender-marked article was absent. A subsequent auditory lexical decision experiment showed that the presence of a prenominal modifier carrying gender (e.g., the article *une* "a" in *une jolie montre* "a pretty watch") resulted in significantly shorter reaction times than did its absence (e.g., *jolie montre*).

Likewise, gender mismatches yield inhibitory effects (e.g., Bates et al., 1996; Carello et al., 1988; Colé & Segui, 1994; Faussart et al., 1999; Gurjanov et al., 1985; Jacobsen, 1999; Jakubowicz & Faussart, 1998; Lew-Williams & Fernald, 2007; Schiller, 2009; Schmidt, 1986; van Berkum, 1996; Wicha et al., 2005). Studies have shown that when the gender of an article and an adjacent noun are incongruent (e.g., *la*$_{\text{FEM}}$ *cuaderno*$_{\text{MASC}}$ "the book"), noun recognition is significantly slowed. Gender congruency effects are robust in studies that use both visual tasks (e.g., Carello et al., 1988; Colé & Segui, 1994; Cubelli et al., 2005; Gurjanov et al., 1985; Jacobsen, 1999; Jescheniak, 1999) and auditory tasks (e.g., Dahan et al., 2000; Faussart et al., 1999; Grosjean et al., 1994; Jescheniak, 1999). They have also been found for languages such as Spanish, which has two grammatical genders (e.g., Barber & Carreiras, 2005; Lew-Williams & Fernald, 2007), as well as for languages with more than two grammatically marked genders, such as German and Serbo-Croatian (e.g., Gurjanov et al., 1985; Jacobsen, 1999; van Berkum, 1996). The general consensus in the literature is that gender marking is one variable among many (e.g., word frequency, word length, neighborhood size) that have already been proposed to account for the time it takes to identify a word (Grosjean et al., 1994).

There remains less consensus regarding how gender agreement affects processing in the L2. A few recent neurophysiological studies that used methods that are highly sensitive to the time course of comprehension have provided evidence for both similarities and differences between native and nonnative speakers' processing of grammatical gender. To contextualize the logic of our own study, we highlight relevant results in studies that investigated the possible contributions of both L1 and language experience to gender processing in a L2.

Sabourin and Stowe (2008) used event-related potentials (ERPs) to investigate whether simply having grammatical gender in the L1 modulated the detection of grammatical gender violations in the L2. Using the so-called P600 as their index of syntactic anomaly detection (Ainsworth-Darnell, Shulman, & Boland, 1998; Gouvea, Phillips, Kazanina, & Poeppel, 2010; Osterhout & Holcomb, 1992), Sabourin and Stowe reported a P600 response during the processing of grammatical gender agreement violations in the L2 only when the gender system of the L2 (Dutch) was very similar to that of the L1 (German). When the two systems were dissimilar—as was the case with Romance speakers proficient in Dutch—they found no P600 effect in the L2 and concluded that the mere presence of grammatical gender in the L1 does not lead to the recruitment of the same type of neurological areas to process gender violations.

Other recent ERP evidence, however, suggests that under conditions in which the gender agreement system of the L1 has little in common with that of the L2 (see Foucart & Frenck-Mestre, 2011, for French and German) or even when grammatical gender is entirely absent in the L1 (Dowens, Vergara, Barber, & Carreiras, 2010), L2 speakers with enough proficiency in the L2 show some of the same signature effects associated with native speaker detection of gender-agreement anomaly.

The ERP study in Dowens et al. (2010) is particularly interesting in that it indicates that immersion experience in the L2 modulates sensitivity to grammatical gender in speakers whose L1 does not make grammatical gender distinctions. For within-phrase agreement violations, results showed qualitatively similar ERP patterns between the L2 speakers and a control group of L1 Spanish speakers, consisting of an early negativity followed by a P600 effect. The findings provided important evidence that L1 speakers of a language without grammatical gender can show electrophysiological correlates during the processing of L2 gender agreement violations that are qualitatively similar to those of native speakers (see also Foucart & Frenck-Mestre, 2012, for additional supporting evidence as well as Morgan-Short, Sanz, Steinhauer, & Ullman, 2010, for evidence from an artificial language paradigm for effects of proficiency and immersion-like learning of a L2).

It is important to note, however, that most recent studies on gender processing in the L2 have limited themselves to examining the time course of processing gender agreement violations. Few L2 studies that use techniques that provide a sensitive and continuous measure of real-time processing have investigated grammatical gender processing under conditions of higher ecological validity (e.g., while participants are engaged in tasks that require continuous spoken-language comprehension without the need to perform metalinguistic judgments or to resort to the use of secondary tasks to obtain the dependent measure). One exception is a recent study by Lew-Williams and Fernald (2010) that employed the looking-while-listening technique—an eye-tracking measure of real-time language processing—to examine whether English learners of Spanish used the grammatical gender of articles to facilitate the processing of upcoming nouns. They presented Spanish-speaking adults with two-picture visual scenes in which the pictured objects either matched in gender (e.g., *pelota* "ball$_{\text{FEM}}$" displayed alongside *galleta* "cookie$_{\text{FEM}}$") or differed in gender (e.g., *pelota* "ball$_{\text{FEM}}$" displayed with *carro* "car$_{\text{MASC}}$"). The researchers videotaped participants' gazes with a digital camera while participants heard simple sentences that named one of two pictured objects (e.g., *encuentra la pelota* "find the ball"; *¿dónde está la pelota?* "where is the ball?"). The task was to click on the named object. A series of three experiments showed that when listening to sentences that named both familiar and newly learned objects and words, native speakers were able to orient their eyes toward target objects more quickly on different-gender trials (i.e., when the gender information in the article was informative) than on same-gender trials, eliciting a so-called anticipatory effect. In other words, on different-gender trials, native speakers of Spanish used gender information in the article to identify the correct reference before hearing it in the instruction. Second language speakers of Spanish, on the other hand, waited to hear the noun to initiate a gaze shift. These findings suggested that the presence of a congruent gender-marked article immediately preceding a noun does not speed up lexical processing by L2 speakers.

One potentially crucial limitation of the Lew-Williams and Fernald (2010) study is that the L2 learners were only moderately proficient in Spanish. Results of a language history questionnaire showed that the L2 participants had been exposed to Spanish in instructional settings, including elementary school, middle school, high school, and college, for an average of 5.5 years, and that their mean self-rated proficiency in Spanish was 3.6 out of a possible 5. We previously noted that near-native levels of grammatical gender processing are attainable when L2 speakers have had extensive immersion experience in the L2 environment and are highly proficient speakers of the target language (e.g. Dowens et al., 2010; Foucart & Frenck-Mestre, 2011). It is thus likely that the English-speaking participants in the Lew-Williams and Fernald study were not sufficiently proficient in Spanish to exploit grammatical gender information in articles in a manner resembling that of native speakers of Spanish.

THE PRESENT STUDY

We address two of the central questions related to grammatical gender processing as seen in the review of the recent neurophysiological work in the previous section. First, we ask whether L1 speakers of a language without grammatical gender show effects of prenominal gender marking on the identification of subsequent words when processing a L2 with grammatical gender. Second, because current models of L2 processing demonstrate that processing is influenced by the native language system (Dijkstra, 2005; Kroll & Stewart, 1994), we also examine whether the presence of a gender system in the L1 that overlaps significantly with the gender system of the L2 determines the extent to which grammatical gender processing in the L2 is nativelike. To investigate these questions, we compared L1 speakers of English and Italian who were highly proficient in Spanish to L1 Spanish speakers. To maximize ecological validity in the study of grammatical gender processing in the L2, we employed the eye-tracking technique known as the visual-world paradigm (e.g., Allopenna, Magnuson, & Tanenhaus, 1998; Altmann & Kamide, 1999; Cooper, 1974; Tanenhaus & Spivey-Knowlton, 1996; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995; Trueswell, Sekerina, Hill, & Logrip, 1999). In one variation of the task, auditory material is concurrently presented with a related visual scene containing pictured objects that are displayed on a computer screen. The auditory material plays spoken instructions related to the objects (e.g., *click on the bell*), which participants are asked to follow. During the experiment, participants' eye movements to the objects are recorded while the name of the target object (i.e., the object mentioned in the instruction) unfolds over time. Because eye movements to the visual scene are closely time locked with the auditory input, and because eye movements are recorded while participants are presented with continuous spoken language, the task provides a sensitive, implicit measure of processing in which responses are closely time locked to the input without interrupting the flow of speech (Tanenhaus & Trueswell, 2006).

A third central goal of our study involves increasing the syntactic complexity of the frames in which the target grammatical agreement structures are embedded. Previous studies in which eye movements are examined to investigate the role of gender-marked articles in spoken-word recognition have used spoken instructions in which the target object is embedded in an invariant sentence context (e.g., *encuentra la pelota* "find the ball," in Lew-Williams & Fernald, 2007, 2010; *cliquez sur* *le bouton* "click on the button," in Dahan et al., 2000). Expectation-based accounts of sentence processing (e.g., Levy, 2008) predict that invariant contexts may facilitate the recognition of target nouns, given that words are easier to comprehend in highly predictable contexts. Resource-allocation theories make similar predictions. Invariant sentence frames presumably make fewer demands on cognitive resources than do more elaborate sentence contexts and thus leave comparatively more resources available for alternative processes to be engaged (e.g., when the display contains different-gender objects, participants may use available resources to focus on the gender of the article in the spoken instruction to predict the upcoming noun). One important additional question, then, is to examine whether the anticipatory effect reported when words are embedded in invariant frames is observed when target words appear in richer and more varied sentence contexts. In summary, the study is guided by three principal research questions:

1. Do L1 speakers of a language without grammatical gender show effects of pronominal gender marking on the identification of subsequent words when processing a L2 with grammatical gender?
2. Does the presence of a gender system in the L1 that overlaps significantly with the gender system of the L2 determine the extent to which grammatical gender processing in the L2 is nativelike?
3. Is the anticipatory effect that emerges when words are embedded in invariant frames also present when target words appear in richer and more varied sentence contexts?

METHOD

Participants

Three groups of participants took part in this study: 16 functionally monolingual native speakers of Spanish—who served as native controls—from the University of Granada, Spain; 18 English-speaking learners of Spanish from a large U.S. institution; and 16 Italian learners of Spanish who had learned Spanish during adulthood and were completing a year of university study in Granada. All received monetary compensation for their participation. In a language history questionnaire—described in the Materials section—the native group reported having studied English or French in high school, and none had spent more than one month in a country where the L2 was spoken. The mean self-reported score for overall proficiency in the L2 on a 10-point scale on which 1 was the lowest score and 10 was the highest score was 2.75 (range = 2.0–3.0), indicating that the native speaker group was functionally monolingual. The 18 English-Spanish speakers had learned Spanish during adulthood (i.e., none were heritage speakers of Spanish). This group was divided

Table 1. Mean percentage accuracy and standard deviation on the DELE

Group	*N*	*M*	*SD*
English-Spanish (higher)	9	43.33 (41–45)	1.41
English-Spanish (lower)	9	30.00 (21–38)	5.40
Italian-Spanish	15	26.13 (16–41)	8.23

Note. Range provided in parentheses.

into two proficiency levels on the basis of their performance on a standardized test of Spanish (i.e., *Diploma de Español como Lengua Extranjera* [*Diploma of Spanish as a Foreign Language*], DELE). We provide a detailed description of the test and all other proficiency measures in the Materials section. Because the median correct response was 41 (out of a possible 50 points), we classified the participants whose score was 41 or higher as more proficient and those whose score was less than 41 as less proficient. Table 1 displays mean percentage accuracy, standard deviation, and score ranges for the two English-Spanish groups and for the Italian-Spanish participants. Two-tailed independent-sample t tests revealed a significant difference between the scores of the two groups of English-Spanish speakers, $t(16) = 7.15$, $p < .001$, and between the higher-proficiency English-Spanish group and the Italian-Spanish group, $t(22) = 6.15$, $p < .001$, but no difference between the lower-proficiency English-Spanish group and the Italian-Spanish group, $t(22) = 1.25$, $p = .22$.

As an additional measure of language proficiency, we administered a picture-naming task in which participants produced article + noun fragments to describe pictures displayed on a computer screen. Results of the picture-naming task are presented in Table 2. Two-tailed independent-sample t tests conducted on the correctly named pictures revealed a significant difference between the two proficiency groups of English-Spanish speakers, $t(16) = 6.40$, $p < .001$, and between the higher-proficiency English-Spanish speakers and the Italian-Spanish group, $t(22) = 4.33$, $p < .001$, but no significant differences between the lower-proficiency English-Spanish group and the Italian-Spanish group, $t(22) = 1.42$, $p = .16$.

Table 2. Mean percentage accuracy and standard deviation on the picture-naming task

Group	*N*	*M*	*SD*
English-Spanish (higher)	9	62.55 (60–68)	3.08
English-Spanish (lower)	9	49.29 (45–56)	5.38
Italian-Spanish	15	52.80 (42–65)	6.49

Note. Range provided in parentheses.

Further examination showed that when participants correctly named a picture, they were also highly accurate in supplying the correct article. On average, the higher-proficiency English-Spanish group was 98.9% accurate (*SD* = 1.15), the lower-proficiency English-Spanish group was 92.1% accurate (*SD* = 2.00), and the Italian-Spanish group was 97.6% accurate (*SD* = 2.97). This indicates that all three groups were knowledgeable in gender marking in Spanish.

Finally, to assess knowledge of gender assignment in comprehension, we administered a written picture-identification task. Table 3 provides mean percentage accuracy and standard deviation for the three groups of participants. Two-tailed independent-sample *t* tests revealed a significant difference between the two groups of English-Spanish participants, $t(16) = 5.63$, $p < .001$. There was also a significant difference between the higher-proficiency English-Spanish group and the Italian-Spanish group, $t(22) = 2.66$, $p = .014$, but no significant difference between the lower-proficiency English-Spanish participants and the Italian-Spanish group, $t(22) = 1.54$, $p = .13$. The results indicate that the higher-proficiency English-Spanish participants were more competent in gender assignment in Spanish than the two other groups. Nevertheless, the high mean correct responses of the lower-proficiency English-Spanish group and the Italian-Spanish group suggest that gender assignment in Spanish for these participants proved largely unproblematic. Participants knew the agreement rules in Spanish and applied them with a high degree of accuracy in a production task and a comprehension task. One remaining question is whether these same participants can access this knowledge during online processing of grammatical gender in Spanish.

Materials

Proficiency Measure for the Monolingual Speakers of Spanish: Language History Questionnaire. To assess whether these speakers had knowledge of a L2, the native control group completed a language history questionnaire designed to tap into several aspects of language proficiency and

Table 3. Mean percentage accuracy and standard deviation on the written picture-identification task

Group	*N*	*M*	*SD*
English-Spanish (higher)	9	31.33 (30–32)	0.86
English-Spanish (lower)	9	26.44 (24–32)	2.45
Italian-Spanish	15	28.38 (22–32)	3.23

Note. Range provided in parentheses.

use by self-report (e.g., language dominance; level of proficiency in reading, writing, speaking, and listening; number of years that they studied the L2; length of stay in a country where the L2 was spoken). The self-rated proficiency measure was a 10-point scale with 1 being the lowest score and 10 being the highest score.

Proficiency Measures for the L2 Groups. For the English-Spanish and the Italian-Spanish participants, proficiency in Spanish was assessed via the use of three measures.

DELE. Overall knowledge of Spanish was assessed by administering a section of the DELE. The DELE is a standardized test of Spanish issued by the Ministry of Education, Culture, and Sport of Spain, which tests proficiency in Spanish at seven levels. More information is available at http://diplomas.cervantes.es/en. We administered the *Nivel Superior C2* (*Superior Level C2*), the highest level of accreditation. The test had a maximum score of 50 points and was comprised of three sections: a cloze test, a vocabulary test that included highly specialized uses of the language, and a multiple-choice grammar test. Participants received 1 point for each correct answer and 0 points for incorrect answers.

Picture-naming task. As an additional measure of language proficiency, we administered a picture-naming task. Seventy-two pictures referencing words that range in lexical frequency from low (e.g., *trineo* "sled") to high (e.g., *mesa* "table") appeared, one by one, on a computer screen. The participants' task was to name each picture along with an accompanying definite article (for *trineo* the expected answer was *el trineo*). We chose a picture-naming task because Moreno and Kutas (2005) recently showed that proficiency in vocabulary (as measured by a picture-naming task) affects how quickly L2 speakers recognize and integrate words into a sentence context. Because we required participants to provide article + noun fragments, the picture-naming task also served as a way to assess knowledge of gender assignment in a production task. Responses were scored by one of the experimenters, who assigned a score of 1 to correctly named pictures. All pictures that participants did not name or that they named incorrectly received a 0.

Picture-identification task. Finally, to assess knowledge of gender assignment and agreement in comprehension, we administered a written picture-identification task closely modeled after White et al. (2004) and Montrul et al. (2008). The task, which exploits the availability of nominal ellipsis in Spanish, required participants to complete sentences that were part of a conversation between two individuals who were packing a suitcase for an upcoming vacation. The task included 32 items targeting gender agreement (taken from Appendix A in Montrul et al., 2008)

Figure 1. Sample item from the picture-identification task.

and 16 filler items. Across different varieties of Spanish, objects can vary both in name and grammatical gender. Therefore, we followed Montrul et al. in providing the name of the object below each picture. A sample is given in Figure 1 (objects appeared in color). We assigned a score of 1 to participants' correct responses and a score of 0 to their incorrect responses.

Eye-Tracking Experiment. The eye-tracking portion of the experiment included 112 color pictures of highly familiar concrete objects (see the appendix). Half represented Spanish object names with feminine gender and half with masculine gender. The vast majority of the feminine object names (93%) ended in *-a*, the pattern most typically associated with feminine gender in Spanish. Endings for the masculine object names included the vowels *-o* (69%) and *-e* (12%) as well as a number of consonants (e.g., *-l* [*caracol* "snail"], *-z* [*arroz* "rice"], *-j* [*reloj* "watch"]), reflecting the fact that Spanish masculine phonological endings are less restricted. Lexical frequency for the masculine items did not differ significantly from the feminine items, $t(94.77) = 0.88$, $p = .37$.[1]

Each picture served as the target on one trial and as a distractor on an additional trial. An example is provided in Table 4.

Because readers of left-to-right languages show a bias to view the left side of the screen before viewing the right side, we counterbalanced the presentation side of target items such that each target appeared on the left and on the right side of the screen, as shown in Table 5.

Table 4. Snapshot of the experimental design

Condition 1		Condition 2		Condition 3		Condition 4	
Fem	Fem	Masc	Masc	Masc	Fem	Fem	Masc
Target	Distractor	Target	Distractor	Target	Distractor	Target	Distractor
pera "pear"	*bufanda* "scarf"	*reloj* "watch"	*pozo* "well"	*pozo* "well"	*pera* "pear"	*bufanda* "scarf"	*reloj* "watch"

Note. Fem = feminine, Masc = masculine.

Table 5. Sample of the counterbalance design

Condition 1		Condition 2		Condition 3		Condition 4	
Fem	Fem	Masc	Masc	Masc	Fem	Fem	Masc
Distractor	Target	Distractor	Target	Distractor	Target	Distractor	Target
bufanda "scarf"	*pera* "pear"	*pozo* "well"	*reloj* "watch"	*pera* "pear"	*pozo* "well"	*reloj* "watch"	*bufanda* "scarf"

Note. Fem = feminine, Masc = masculine.

To investigate whether a gender facilitatory effect occurs when participants process sentence contexts, we embedded the picture names in variable sentences, and we distributed the target items evenly so that half appeared in the middle of the sentence (e.g., for *el reloj* "the clock": *El estudiante estaba dibujando* el reloj *que vio ayer* "The student was drawing *the clock* that he saw yesterday") and half at the end (e.g., *El niño miraba a su hermano mientras fotografiaba* el reloj "The boy watched his brother taking a picture of *the clock*"). To conceal the main purpose of the experiment, after listening to each sentence, participants performed a plausibility judgment task. Half of the sentences were plausible (e.g., like those previously exemplified) and half implausible (e.g., *El señor compró* el reloj *para la piedra* "The man bought *the clock* for the rock"). During debriefing, participants confirmed they were not aware that the focus of the experiment was grammatical gender.

A female speaker of Puerto Rican Spanish, who is also a trained linguist, recorded each experimental sentence between three and five times at a comfortable speaking rate in a sound-attenuated chamber with a Shure SM57 microphone on a Marantz Solid State Recorder PMD670 at a sampling rate of 44.1 kHz. The sentences were produced using standard, broad-focus intonation (i.e., no narrow focus or other emphasis was produced on any of the target noun phrases). From the master recordings, one token was selected for inclusion in the experiment. To precisely match the durational properties of the masculine and feminine articles for all of the experimental items, the article preceding the target noun in each selected sentence was edited by hand to a duration of 147 ms ± 3 ms using Praat (Boersma, 2001). This duration was chosen by sampling the master recordings and calculating a mean duration of the masculine and feminine articles. In this way, the duration of the acoustic signal conveying grammatical gender prior to the onset of the target noun was identical across all items.

Picture-Naming Verification Task. To ensure that participants could correctly name the pictures that appeared in the eye-tracking session, we administered a picture-naming verification task. This task included the 112 pictures used in the experiment as well as 10 additional pictures. Half of the 112 experimental pictures comprised the target items in the eye-tracking experiment and, therefore, had been previously named in the auditory presentation of the sentences. The remaining 56 experimental pictures comprised the distractor items in the eye-tracking experiment. Thus, participants had not previously heard those names. Participants saw the pictures one at a time and were asked to say the name of each picture into a microphone. A score of 1 was given to correctly named pictures. One of the experimenters scored correctly named pictures as 1. All pictures that participants did not name or the ones that they named incorrectly received a 0.

Procedure

After providing their consent, participants completed the eye-tracking experiment. Participants' eye movements were recorded using an Eyelink II eye-tracker manufactured by SR Research. Viewing was binocular, but eye movements were recorded from the left eye only.[2] Stimuli were presented on a color 17-in ViewSonic 17PS monitor; participants were seated 65 cm from the monitor and rested their chins comfortably on a chin rest. Calibration was checked on each trial, and spatial resolution was better than 0.5 degrees. To begin each trial, participants looked at a fixation point in the center of the computer screen. Subsequently, two pictures appeared on the screen, and a sentence was played simultaneously in which the name of one of the two pictures was mentioned. Participants clicked on the picture named in the sentence. After selection, the display disappeared, and two squares appeared on the screen, one with the word *plausible* and one with the word *implausible* printed inside. Participants then made a plausibility judgment for each sentence by clicking on one of the squares. Ten practice sentences preceded the experimental items. The session lasted approximately 18 min. After completing the eye-tracking experiment, the participants were administered the language history questionnaire, the picture-naming task, and the DELE. After a short break, participants completed the written identification task and the picture-naming verification task.

Analysis

There is no clear consensus on the best way to analyze proportional data obtained with the visual-world paradigm (see special issue 59 of the *Journal of Memory and Language* for topics on eye-tracking data analysis). In particular, researchers are confronted with a critical decision on how participants are allowed to interact with the visual scene (i.e., free-view or fixed-view presentation). Both methods have their advantages and disadvantages. Allowing free view of the visual scene represents a more ecological task, reflective of what participants would presumably do under nonexperimental settings. Therefore, a free-view presentation offers an ecological advantage over a fixed-visual presentation. However, a free-view presentation aggravates one potentially problematic issue in data analysis that is attenuated in fixed-visual presentations. Specifically, because participants are idiosyncratic in the manner in which they view a visual scene prior to hearing a named object, the free-view presentation greatly increases the likelihood for baseline effects. Briefly, baseline effects are represented on a time-course plot by the y-intercept—the value of y at x = 0. The greater the magnitude of difference between the y-intercept of the target and any distractors, the greater the baseline effect, which subsequently represents a random effect in eye-tracking data. Because, in fixed visual presentations, participants do not begin looking at the visual scene until the onset of the target region of interest, baseline effects are nullified. In other words, all proportional data begin at 0 at the onset of the target region of interest. In the context of the experiments reported here—because the target region is embedded in sentential contexts—a fixed-view presentation mode is not appropriate as it would artificially alert the participant to the target region of interest.

Given this, our analysis needs to account for random baseline effects. One method that has been proposed is known as a contingent-based analysis. This method of data trimming includes only trials in which the participant is not looking at target items at the onset of the critical region, thus removing baseline effects post hoc. Although this method may be a viable option for a traditional four-picture display, we consider it an unviable option for our two-picture display due to high loss of data. Furthermore, some researchers have reported that a contingent-based analysis may be biased, leading to an overestimation of effects (e.g. Barr, Gann, & Pierce, 2010). Thus, our approach would need to attenuate random baseline effects while retaining higher amounts of data. To achieve this, we conducted a change point analysis by implementing a multiphase mixed-effects regression model (Cudeck & Klebe, 2002, see also Baayen, 2008). This type of analysis has been proposed for repeated-measures longitudinal studies to determine whether an experimental treatment has an impact on behavioral change modeled over many days, months, or years. The basic feature of this analysis is that any number of phases, each of which is uniquely modeled by its own function, can be united into a more complex whole (described in more detail in Cudeck & Klebe, 2002). This method allows us to maintain all of the data extracted from the eye-tracker. Another advantage to this approach is that we can estimate a point in the time course (i.e., the change point, or the point at which there is a shift between phases). The change point describes the moment in time when one rate of change changes to a different one.

Although typically applied to longitudinal studies, we apply the logic of this approach to the millisecond timescale for visual-world eye-tracking data. The data that we attain from the eye-tracker is repeated-measures data. Because eye movements are impacted by auditory stimuli (i.e., participants fixate on named objects), the onset of the critical region is the experimental treatment. In terms of the current study, we defined this critical onset as the beginning of the Spanish article and extracted 1,000 ms from this critical onset. Plots of fixations over time to target items were best characterized by an s-shaped curve. Therefore, we modeled a three-phase regression model, with each phase described by a linear function. We term these three phases (a) the preconvergence phase, (b) the convergence phase, and (c) the postconvergence phase. The preconvergence phase corresponds to eye movements that are not directly impacted by the critical region in the auditory stimuli; rather, they include random baseline effects due to participants' free view of the visual scene and the time dedicated to launching eye movements toward target items. The convergence phase represents the period of time whereby participants' eye movements shift toward target items. Finally, the postconvergence phase corresponds to the stage in real-time processing in which participants are no longer uniformly affected by the experimental stimuli; that is, participants begin to return to a random state of free view. Despite these three distinct phases, the data are continuous over time. Thus, we can calculate change points that represent the points in time when a new phase begins.

Experimentally, we are interested in the first change point. The first change point between the preconvergence and convergence phases indicates the point in time when a critical mass of participants begins to shift fixations to the target item. We can then compare change points across conditions. Specifically, by conducting simple paired *t* tests, we then determine whether one change point occurs significantly earlier (or later) than another change point. This method reveals whether an experimental condition induces a facilitatory or delayed effect when compared to a baseline condition. For the current study, we are interested in whether the change point for different-gender conditions happens significantly earlier than for same-gender conditions.

To calculate parameter estimates for the model, we estimated starting values by using the grand averaged data across all participants by condition. These estimates were then included as parameter starting values for the mixed-effects model applied to the individual-level data. Furthermore, the y-intercept was included as a random effect across participants and conditions to allow variability in baseline effects prevalent in spoken-word eye-tracking data (Tanenhaus & Trueswell, 2006). After estimating change points, we conducted paired *t* tests on the first change point. We estimated the difference between same-gender and different-gender conditions for both the masculine and feminine targets to see whether both genders yielded facilitatory effects in the different-gender condition relative to the same-gender condition. A negative difference estimate indicates that the different-gender condition results in a facilitatory effect as compared to the same-gender condition.

RESULTS

Plausibility Judgment

In general, participants were accurate in providing plausibility judgments for the sentences they heard, although, overall, they were more accurate judging implausible sentences than judging plausible ones. This may not be surprising if identification of lexicosemantic fit is more straightforward in implausible sentences given their severely poor fit (e.g., *John used a pump to inflate the carrots*) than in plausible sentences, in which participants could imagine different scenarios that could potentially lead to either judgment (e.g., *John used a fork to cut the carrots*). For plausible sentences, accuracy was 91% for the monolingual Spanish group, 89% for the English-Spanish speakers, and 84% for the Italian-Spanish group. For the implausible sentences, accuracy was 98% for the monolingual group, 93% for the English-Spanish group, and 96% for the Italian-Spanish group. Taken together, these results indicate that participants were, in fact, truly paying attention to meaning and, importantly, understood the sentences.

Picture-Naming Verification Task

Recall that this task assessed whether participants could correctly identify the pictures employed during the eye-tracking session of the study. Pictures named correctly were assigned a score of 1, and those named incorrectly were given a score of 0. Mean percent accuracy and standard deviations are provided in Table 6. Two-tailed independent-samples *t* tests conducted on the correct responses revealed no significant differences among the three groups of participants (between the two groups of English-Spanish speakers, $t(16) = 0.45$, $p = .65$; between the higher-proficiency English-Spanish group and the Italian-Spanish participants, $t(22) = 0.93$, $p = .35$; and between the lower-proficiency English-Spanish

Table 6. Mean percentage accuracy and standard deviation on the picture-naming verification task

Group	*N*	*M*	*SD*
English-Spanish (higher)	9	120.34 (118–122)	1.40
English-Spanish (lower)	9	120.00 (117–122)	1.73
Italian-Spanish	15	119.66 (115–122)	1.87

Note. Range provided in parentheses.

group and the Italian-Spanish participants, $t(22) = 0.44$, $p = .66$). This finding indicates that all three groups could identify the pictures used during the eye-tracking session with a high degree of accuracy.[3]

Eye-Tracking Experiment

Monolingual Speakers of Spanish. The minimum latency to plan and launch a saccade has been estimated to be approximately 200 ms (e.g., Fischer, 1992; Saslow, 1967). Thus, approximately 200 ms after target onset is the earliest point at which one expects to see fixations driven by acoustic information from the target word. Visually, we plot the time course of proportion of fixations toward target items, following Lew-Williams and Fernald (2007, 2010).[4] To delineate where the model calculates the first change point estimates for each condition, we overlay short vertical segments on top of the time-course plots (see Figure 2). Additionally, we plot the change point estimates for each condition,

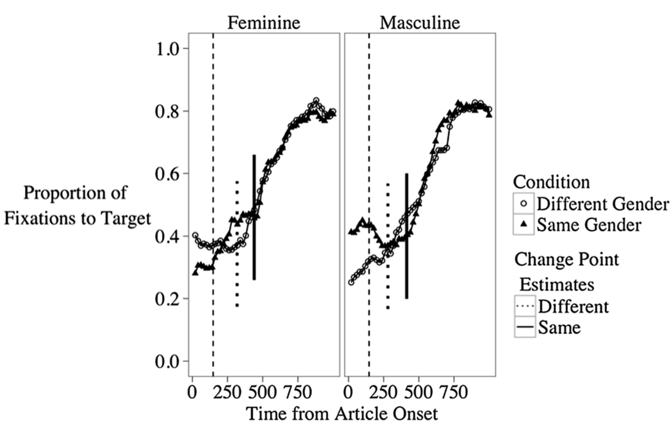

Figure 2. Proportion of fixations to targets over time for different-gender and same-gender trials (monolingual Spanish speakers).

including a standard estimate band to more clearly isolate the estimates on which we conduct our statistical tests (see Figure 3).

For feminine conditions, the change point estimate was 439 ms (*SE* = 41.4) for the same-gender trials. In contrast, different-gender trials have a change point of 317 ms (*SE* = 27.4). Additionally, the y-intercept random effect (i.e., baseline) variance was 0.0072 (*SE* = 0.0027), and the model error was 0.0268 (*SE* = 0.0009). Thus, feminine different-gender trials have an earlier change point (*MD* = 122 ms) than same-gender trials (*SE* = 49.65, paired $t(15) = -2.45$, $p = .027$), indicating that Spanish monolinguals use the feminine article as a facilitatory cue in real-time speech. For masculine conditions, same-gender trials reveal a change point of 414 ms (*SE* = 22.94). For the different-gender trials, the change point is 280 ms (*SE* = 48.66). Here, the y-intercept random effect variance was 0.0061 (*SE* = 0.0023), and the model error was 0.0221 (*SE* = 0.0007). As in the feminine conditions, masculine different-gender trials have an earlier change point (*MD* = 134 ms) than same-gender trials (*SE* = 53.79, paired $t(15) = -2.5$, $p = .025$). Spanish monolinguals also use masculine as a facilitatory cue in real-time processing.

Higher-Proficiency English-Spanish Speakers. For feminine targets, the high-proficiency group has a change point estimate of 466 ms (*SE* = 24.83) for same-gender trials. On different-gender trials, the change point estimate is 375 ms (*SE* = 31.34) (see Figures 4 and 5). The y-intercept random effect variance was 0.0055 (*SE* = 0.0027), and the model error was 0.0292 (*SE* = 0.0013).

The associated paired *t* test for the difference magnitude reveals a marginally significant faster change point (*MD* = 91 ms) for different-gender

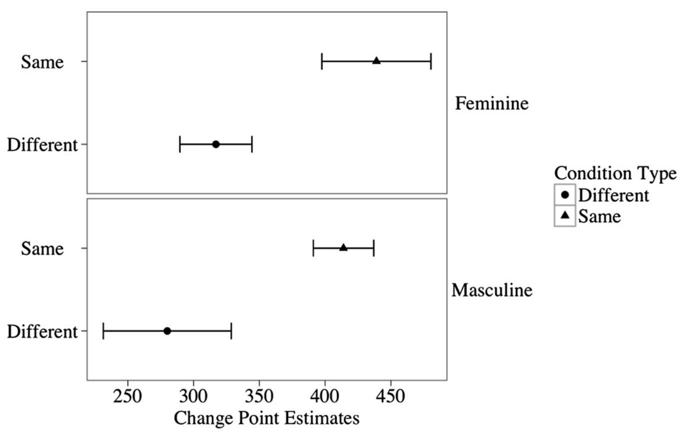

Figure 3. Change point estimates plotted with standard errors (monolingual Spanish speakers).

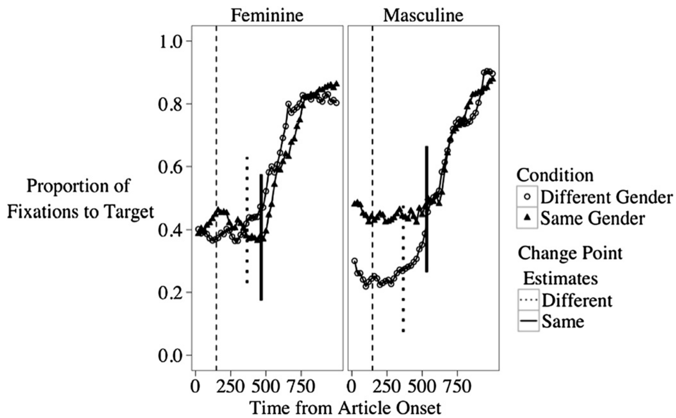

Figure 4. Proportion of fixations to targets over time for different-gender and same-gender trials (high-proficiency English-Spanish learners).

trials (*SE* = 39.99, paired $t(8) = -2.28$, $p = .05$). For masculine targets, the change point estimate for same-gender trials is 532 ms (*SE* = 30.48). For different-gender trials, the change point estimate is 366 ms (*SE* = 30.56). The different-gender trial change point is significantly faster (*MD* = 166 ms) than the same-gender trial change point (*SE* = 43.16, paired $t(8) = -3.85$, $p = .005$). Additionally, the y-intercept random effect variance was 0.0063 (*SE* = 0.0031), and the model error was 0.0317 (*SE* = 0.0014). The results suggest that for both gender types, high-proficiency English-Spanish bilinguals are capable of using grammatical gender as a facilitatory cue in real-time processing.

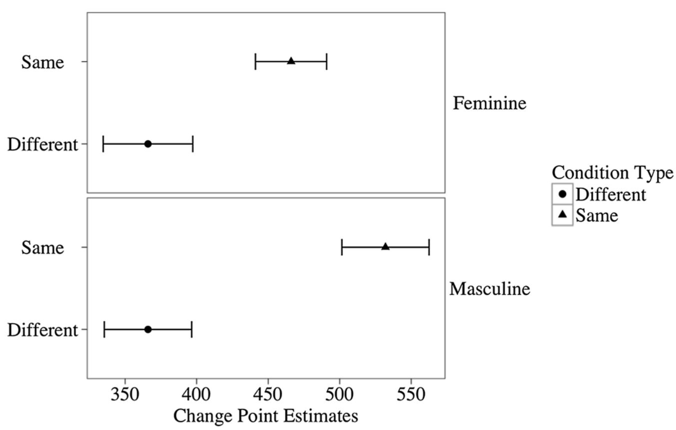

Figure 5. Change point estimates plotted with standard errors (high-proficiency English-Spanish learners).

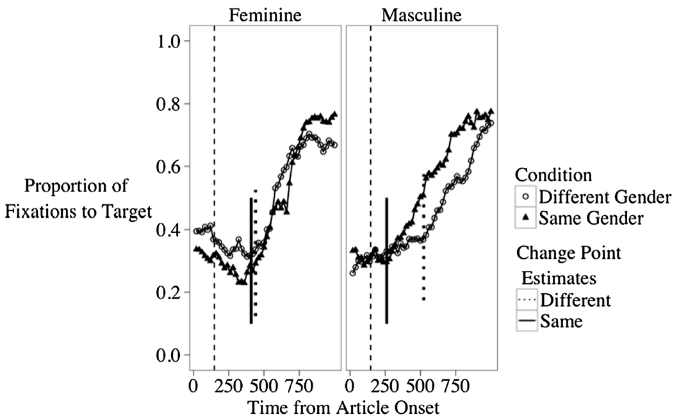

Figure 6. Proportion of fixations to targets over time for different-gender and same-gender trials (low-proficiency English-Spanish learners).

Lower-Proficiency English-Spanish Speakers. For feminine targets, the change point estimate for the low-proficiency group is 406 ms (*SE* = 25.67) on same-gender trials. Different-gender trials have a change point estimate of 439 ms (*SE* = 27.02) (see Figures 6 and 7). The y-intercept random effect variance was 0.0218 (*SE* = 0.0104), and the model error was 0.0333 (*SE* = 0.0014).

The difference between the two change points is not significant, difference = 31 ms (*SE* = 37.27, paired $t(8) = 0.82$, $p = .434$). For masculine targets, same-gender trials have a change point estimate of 261 ms (*SE* = 36.78). The change point estimate for different-gender trials is 523 ms (*SE* = 52.64). Here, the difference estimate is significantly different, but

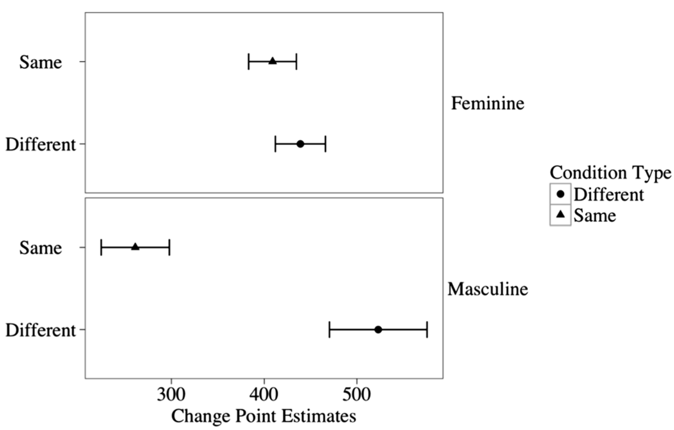

Figure 7. Change point estimates plotted with standard errors (low-proficiency English-Spanish learners).

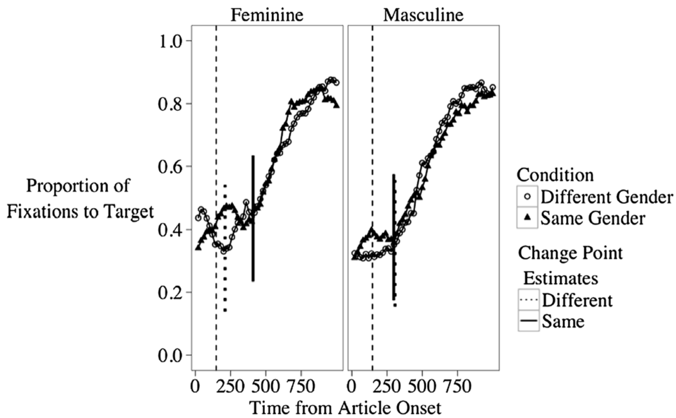

Figure 8. Proportion of fixations to targets over time for different-gender and same-gender trials (Italian-Spanish learners).

in the unanticipated direction. That is, the same-gender change point is significantly faster (*MD* = 262 ms) than the different-gender change point (*SE* = 64.22, paired $t(8) = 4.08$, $p < .005$). The y-intercept random effect variance was 0.0148 (*SE* = 0.0071), and the model error was 0.242 (*SE* = 0.001).

Italian-Spanish Speakers. For feminine targets, the same-gender change point is estimated at 408 ms (*SE* = 21.38) from article onset. In contrast, the change point is 211 ms (*SE* = 29.54) for different-gender trials (see Figures 8 and 9). The y-intercept random effect variance was 0.0072 (*SE* = 0.0027), and the model error was 0.0268 (*SE* = 0.0009).

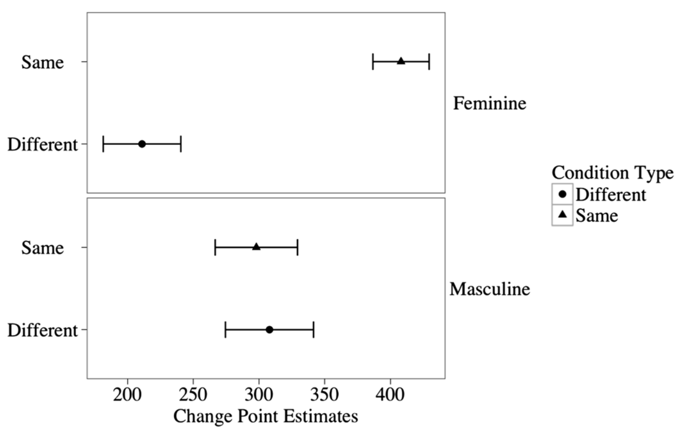

Figure 9. Change point estimates plotted with standard errors (Italian-Spanish learners).

The difference estimate and associated paired t test indicate that Italian-Spanish bilinguals use the feminine article as a facilitatory cue in real-time processing of Spanish ($MD = 197$ ms, $SE = 36.46$, paired $t(14) = -5.4$, $p < .0001$). For masculine targets, the change point estimate is 298 ms ($SE = 31.32$) for same-gender trials. For different-gender trials, the change point estimate is 308 ms ($SE = 33.53$). The y-intercept random effect was 0.0061 ($SE = 0.0023$), and the model error was 0.0221 ($SE = 0.0007$). The difference estimate and associated paired t test is 10 ms ($SE = 45.88$, paired $t(14) = 0.88$, $p = .827$). In contrast to the manner in which they exploit gender cue provided by the feminine article, Italian-Spanish bilinguals do not appear to make use of masculine articles as a facilitatory cue.

DISCUSSION

As we note in the introduction, much of the traditional work on L2 gender agreement in the SLA literature has taken as its point of departure differences between native and nonnative speakers with respect to ultimate attainment—differences largely examined in the context of competing competence-based theoretical accounts of L2 acquisition in the generative tradition. Although insightful, a consequence of this discourse is that it minimizes the surprisingly accurate overall performance (generally 80–90% accuracy) exhibited by learners on tasks involving gender agreement. This kind of performance across numerous studies suggests that L2 learners have integrated the target language gender system into their developing L2, and it thus becomes interesting to examine whether and how L2 speakers can exploit gender during online processing tasks—an approach that provides an excellent means of furthering our understanding of the similarities and differences between native speakers and learners while also illuminating our understanding of the mechanisms underlying the architecture of the language-processing system.

Our study addressed three particular questions. First, can and do native speakers of a language without grammatical gender use grammatical gender information in the L2 to facilitate the processing of upcoming nouns? Second, does the presence of grammatical gender in the L1 facilitate gender agreement processing in L2 as indexed by anticipatory eye movements to noun referents? Third, does the anticipatory effect reported in previous gender-processing studies of spoken-word recognition obtain when target nouns are embedded in variable sentence contexts and participants are engaged in a secondary task (i.e., providing a semantic judgment on the sentence)?

Unlike most recent studies on grammatical gender processing in the L2, which have focused on the processing of gender agreement violations, we examined grammatical gender processing while participants were engaged in a task that required attention to meaning. This was important because past research has shown that the allocation of visual attention during spoken-language processing is task dependent and that task can critically affect linguistically driven eye movements during online spoken-language comprehension (Kreysa & Knoeferle, 2011). Therefore, an important additional goal was to examine whether the facilitatory effect reported during the processing of grammatical gender marking in articles generalized beyond tasks in which participants respond to simple (and invariant) instructions.

To address these questions, we compared the performance of two groups of native speakers of English and a group of native speakers of Italian to that of native speakers of Spanish. Participants saw two pictured objects on a computer screen while hearing a sentence naming one of the two objects. We collected data using the visual-world paradigm, an eye-tracking technique that capitalizes on the dynamics of speaker gaze during online sentence processing to provide information about the time course of spoken-language processing with millisecond precision. In this technique, moment-by-moment allocation of visual attention to target and distractor pictures is recorded to yield a dependent measure comprised of proportion of gaze shifts to the pictured objects in response to the unfolding speech signal.

In line with previous findings (Dahan et al., 2000; Lew-Williams & Fernald, 2007), results for the native Spanish-speaking group showed evidence of the use of gender marking on articles to anticipate upcoming nouns in contexts in which two pictured objects belonged to different gender classes. Native speakers launched anticipatory eye movements to target items in different-gender trials (when information in the article was informative) but waited to hear the target nouns in the same-gender trials before shifting their gaze. It is useful, in particular, to highlight three aspects of our results in the context of Lew-Williams and Fernald's (2007) work on Spanish. Specifically, we have replicated their finding of an anticipatory effect with a more precisely time-locked methodology (i.e., eye-tracking), in varied carrier phrases, and with a secondary task that adds cognitive load. Additionally, in contrast to Lew-Williams and Fernald (2007), we have broken out the effect for both genders and provide evidence of anticipatory eye movements for both masculine and feminine agreement structures independently.

One particularly noteworthy feature of our findings is that this anticipatory effect was observed even when the target nouns were embedded in rich sentence contexts, as opposed to invariant sentences structures such as those employed in the Lew-Williams and Fernald studies (which can potentially focus participants' attention on article + noun processing alone), and when participants were engaged in a task that required attending to other properties of the sentence—in this case its plausibility. To our knowledge, our study is the first to provide empirical evidence demonstrating the rapid use of gender-marked information in articles to speed up noun recognition when attention is directed to other features of processing (i.e., in-depth semantic processing), and it suggests that (a) previous findings in the gender-processing literature are not artifacts of the experimental setup and (b) native listeners do attend to gender cues and exploit them to quickly make sense of a speaker's message.

Turning to the learner data, results for the two groups of late English-Spanish learners revealed sensitivity to gender marking on Spanish articles similar to that found in native speakers, but this sensitivity was affected by the level of proficiency. Specifically, the higher-proficiency English-Spanish group was quicker to orient to both feminine and masculine target pictures when the article was informative (i.e., in different-gender trials) than when it was not (i.e., in same-gender trials). This result is crucial in that it clearly indicates that late L2 learners of Spanish are able use the information they have acquired about grammatical gender in Spanish quickly and efficiently to establish reference in a task that was resource demanding, both because it required the secondary task of clicking on a picture and because it required listeners to process the auditorily presented stimuli for semantic plausibility. Even if it is not surprising that the mean processing time—as measured by the timing of the change point—is slower than that of native Spanish speakers, the results for the high-proficiency learners are qualitatively the same. In this sense, our eye-tracking results for the high-proficiency English learners are congenial with recent electrophysiological results, such as those of Dowens et al. (2010) and Foucart and Frenck-Mestre (2011). Although these studies were different in that they examined learner sensitivity to gender violations within different constituency structures and across phrases, when taken together, their early electrophysiological results and our later behavioral measures provide converging evidence that, with sufficient proficiency, late adult learners can process grammatical gender in a manner highly comparable to that of native speakers. Such converging results have important theoretical implications in that they support an experience-based view of acquisition and provide evidence for continued plasticity in the system, even for L2 learners such as our participants whose L1 lacks grammatical gender (Foucart & Frenck-Mestre, 2012).

The data from the low-proficiency English learners present a more complicated picture. In sentences with feminine targets, the low-proficiency participants exhibited no anticipatory effects for gender; that is, we found no significant difference when we compared the change points for scenes in which the items pictured were both feminine nouns versus scenes in which the two items differed in grammatical gender. If these participants were exploiting gender cues on the article in the processing of an upcoming noun, we should have expected an anticipatory effect. Again, this result accords with what has been found in the electrophysiological literature by Dowens et al. (2010) for English learners of Spanish. Their nonimmersed participants did not exhibit evidence in the ERP signal of sensitivity to gender violations, whereas our low-proficiency participants did not show behavioral evidence of using the gender cue to facilitate the processing of an upcoming noun.

The complication for this group resides in the masculine condition. As reported previously, we find a significant difference in the masculine condition, but the results indicate a facilitatory effect for scenes in which both pictures have masculine gender in Spanish. This is a reversal of what would be expected if the article's gender facilitates processing. That is, we should expect the masculine determiner to speed the processing of a masculine noun in scenes in which the two pictures do not both have masculine gender. It is not clear to us at this point why same-gender scenes should lead to faster identification (i.e., an earlier change point) in the masculine condition. What is extremely clear, however, is that our low-proficiency group shows no evidence of nativelike behavior in the task. For these participants, gender does not serve as a cue that is exploited to facilitate online processing.

One admittedly speculative explanation of what may be happening in this condition is that our learner participants may be evidencing the initial stages of incorporating an awareness of grammatical gender during processing that becomes manifest as a reverse effect under the particular task conditions here. Recall that we allow our participants to freely view the scene prior to the onset of each trial, and we know from our posttest that the participants could name the pictures used in the experiment with a high degree of accuracy. One potential explanation for the results in this condition could be that in same-gender trials, participants made no attempt whatsoever to incorporate gender processing into their response to the task, given the fact that when two same-gendered pictures appeared on screen, there would be no reason to engage resources in processing the gender of the article. However, when different-gender items appeared in trials, it is plausible that these speakers began to attempt to process the gender of the masculine article as the speech signal unfolded. However, at this point in their development, rather than providing a beneficial cue, the cost of attempting to process masculine gender online, under these task demands, yielded an inhibitory effect in terms of the time needed to converge on the named picture. Although this reasoning is speculative, it does lead to two possible predictions. The first is that the time course of learning to process gender for late L2 English-speaking learners of Spanish may be different for the masculine versus feminine gender. The second is that the cognitive demands of gender processing may yield inhibitory effects in lower\-proficiency learners and facilitative effects in higher-proficiency learners.

Further studies must examine this issue. However, stepping back from the particular details of the masculine conditions in our experiment, our results clearly suggest that high-proficiency English-speaking learners of Spanish can process grammatical gender agreement in article + noun structures in a manner qualitatively equivalent to that of native Spanish speakers, whereas low-proficiency late learners do not exhibit the ability to do so.

The Italian learners provide an interesting contrast to the native English-speaking participants. In the feminine condition, the Italian participants exhibited an anticipatory effect when viewing scenes in which one picture was a masculine noun and the other a feminine noun. That is, the change point in the different-gender displays was significantly earlier than in the same-gender displays. This result is comparable to both the results we found for native Spanish speakers and for the high-proficiency English-speaking participants. For the feminine condition, at least, our results suggest that the Italian participants do exploit grammatical gender on the article as a means of predicting the identity of an upcoming noun. By contrast, there is no significant difference between the change points for the same- and different-gender displays in the masculine article conditions, which indicates that the Italian participants did not exploit gender as a cue in predicting the identity of a following noun when the determiner carried masculine grammatical gender.

The results of the Italian participants are particularly useful to consider in the context of the performance of the high- and low-proficiency groups of English-Spanish speakers. On the one hand, unlike the English speakers, the Italian learners may be advantaged by the presence of grammatical gender in the L1, especially given the similarities between Italian and Spanish. This would appear to be the case in the feminine conditions, in which the Italian participants capitalize on the gender of the feminine article to facilitate the processing of the noun. This result is compatible with Sabourin and Stowe's (2008) results that learners of a L2 with a similar gender system as their L1—in their case, German learners of Dutch—can exhibit nativelike processing of gender, whereas speakers of languages with highly distinct gender systems do not (i.e., Romance languages and Dutch).

On the other hand, the relatively low proficiency of the Italian participants might be expected to mitigate their having sufficient resources available to efficiently capitalize on the presence of the gender on the article as a means of predicting the identity of the upcoming noun while processing the unfolding acoustic stream. This might have influenced the results for the masculine conditions, in which we found no difference in time to convergence between same- and different-gender displays. For the masculine article, the native Italian participants did not exploit gender to facilitate the processing of the following noun. If this view is correct, the low-proficiency Italian participants provide an illustrative example of the balancing of factors that may advantage learners (e.g., having a L1 with a very similar two-gender system and a large degree of overlap in cognate vocabulary) and that may hinder learners (e.g., low proficiency) in their allocation of resources when processing gender in a L2. Such a finding is compatible with results that display that proficiency is a critical factor in predicting whether late L2 learners can exhibit nativelike processing of grammatical gender in the L2 (e.g., Foucart & Frenck-Mestre, 2011).

Nevertheless, if the results for the Italian speakers may be modulated by a combination of factors that should both facilitate (i.e., having gender in their L1 that is very similar to the Spanish gender system) and hinder (i.e., their relatively low Spanish proficiency) the use of grammatical gender in processing the L2, the question arises as to why the masculine, but not the feminine, condition should be affected by proficiency level. After all, the gender agreement systems in Italian and Spanish share many features. Both languages have a two-gender system (masculine and feminine), and, like Spanish, Italian words that end in *-o* generally correlate with masculine gender (e.g., *il tavolo* "the table," *lo scoiattolo* "the squirrel," *il treno* "the train") and words that end in *-a* are generally feminine gender (e.g., *la bicicletta* "the bicycle," *la chitarra* "the guitar," *la sedia* "the chair"). In the context of our study, approximately 84% of the words used in the experiment shared the same gender between the two languages. This high degree of convergence—coupled with past findings that show that, when agreement rules are similar between the L1 and the L2, L2 speakers can process gender in a similar way to native speakers (Foucart & Frenck-Mestre, 2011; Sabourin & Stowe, 2008)—predicted that anticipatory effects would be manifested in both the feminine and the masculine different-gender trials. Yet, our Italian-Spanish learners used gender information encoded in the article anticipatorily only in the feminine condition.

One reason may have to do with the nature of the experimental items. Most pictures in the experiment had transparent gender in Spanish (e.g., ended in either *-a* or *-o*), but a little more than 17% of the pictures had opaque gender (i.e., they ended in a vowel other than *-a* or *-o* or ended in a consonant). Of these opaque words, approximately 80% were masculine gender. It is possible, then, that for this group of low-proficiency Italian-Spanish learners, deploying resources to process words with opaque gender may have been effortful, resulting in the lack of an anticipatory effect for the masculine different-gender trials. This explanation is congruent with recent findings in the functional imaging literature showing that gender processing for opaque words in Spanish requires deeper and more effortful processing than gender processing for transparent words (Hernandez et al., 2004). The difference in behavior for the masculine and feminine gender trials may also have to do with differences in the masculine articles in the two languages.

Italian has two masculine definite articles (i.e., *il* and *lo*), which are phonologically conditioned, and only one feminine definite article (i.e., *la*). Spanish, on the other hand, only has two definite articles (i.e., *el* for masculine and *la* for feminine). Sabourin and Stowe (2008) showed that when there is not sufficient congruence between the L1 and the L2 systems, transfer of information from one system to the other can fail. Hence, the inability of the Italian-Spanish learners to use the presence of the gender on masculine articles as a means to predict upcoming nouns may also be at least partially related to differences in the definite article systems of the two languages. Clearly, both of these explanations remain largely speculative because our design does not allow us to pinpoint the cause for the difference in behavior between the masculine and feminine conditions. In future research, it may be useful to carefully control list composition as a way of testing for the potential contributions of subtle differences between the gender systems of the two languages when Italians process gender in Spanish.

In conclusion, these results contribute to our understanding of L2 gender processing and, thus, to our growing understanding of the approximation of L2 learners to native speaker performance in a number of ways. First, methodologically, we have shown that despite our task's higher cognitive load, we replicate the findings of a facilitatory effect of gender cue on articles in the processing of upcoming nouns for native speakers of Spanish (e.g., Lew-Williams & Fernald, 2007) and high-proficiency learners. Additionally, we go beyond the Lew-Williams and Fernald (2007) results by showing these effects separately for both the masculine and feminine genders. Second, we provide clear behavioral evidence, congruent with previous electrophysiological studies, that high-proficiency learners of Spanish can achieve qualitatively similar patterns of performance as native monolingual Spanish speakers, even if they are L1 speakers of a language that lacks grammatical gender. This contrasts with the Lew-Williams and Fernald (2010) results, a difference we attribute to the lower proficiency of their participants.

Our results also give rise to a number of issues that warrant further investigation. One of these involves the performance of our low-proficiency English learners of Spanish. Given both the previous work in generative approaches to SLA previously cited and the reviewed electrophysiological studies, it is not surprising that these learners showed no facilitatory effect of feminine gender. We were surprised, however, to see that in the cases of masculine agreement, same-gender trials were significantly faster than different-gender trials. We hypothesize that this may actually be due to an emerging effect of gender processing—that is, an attempt to exploit masculine gender that leads to a slowing down of processing, given the lack of automatized gender processing. In future research, one way to begin testing this hypothesis is to exploit the potential for task effects by simplifying the task—for example, by using an invariant syntactic frame, such as that of Lew-Williams and Fernald (2007), to probe the question of whether a task involving a lower cognitive load would allow for the facilitative use of the masculine gender. The assumption would be that resources would be freed up to allocate to the task of processing gender. Another remaining question involves potential differences in the time course of acquisition regarding the ability to exploit gender on the masculine and feminine articles, respectively. Whether the two genders in Spanish are acquired on the same time course and what the potential modulation of the L1 of that acquisition time course may be remain open questions. AQ6

Regarding the Italian participants, two important questions remain concerning the interplay of L1 and proficiency. What is clear from our results is that the Italian learners look like neither the high-proficiency nor the low-proficiency English learners of Spanish. We reason that this finding derives from the two principal variables in play in this study: (a) the presence of a gender system in the L1 and its similarity to the gender system of the target language, and (b) the target language proficiency of the speakers. In future work, we will begin to tease these factors apart both by testing high-proficiency Italian learners of Spanish and by manipulating the list composition of the stimuli to gain a more granular perspective on the effects of the particular similarities and differences between the gender systems of Spanish and Italian.

NOTES

1. Lexical frequency data were obtained using the *Diccionario de frecuencias de las unidades lingüísticas del castellano* (*Dictionary of the Linguistic Units of Spanish*) (Alameda & Cuetos, 1995).

2. An anonymous reviewer asked why viewing was binocular, whereas recording was monocular. The decision to record just one eye is largely driven by practical issues. The Eyelink 1000 records with a default monocular setting; this setting provides a number of advantages. First, it allows for a higher sampling rate, which results in more precise data per participant. Second, monocular recording speeds up the process of camera setup and calibration and also allows more freedom of head movement on the part of the participant (i.e., binocular recording reduces the allowed head movement to approximately 25 mm for horizontal and vertical movement, which makes the experimental session uncomfortable for participants). Third, data extraction and data analysis for monocular recording are significantly more efficient than for binocular recording. For example, minuscule differences typically arise between the two eyes, which are difficult to reconcile. While one eye may be in saccadic movement, the other eye may be in fixation. These differences must be reconciled, but averaging the results of the two eyes is not an appropriate solution because the measures are binary (i.e., the eye is either in saccadic movement or is not; the eye is either blinking or is not). Additionally, the data files produced from one eye alone are quite large. Having twice as much data considerably slows down the processing time of the system. In short, binocular recording adds a number of complications to the data-collection process and does not produce any clear benefits. For this reason, most eye-tracking studies allow binocular viewing but analyze data from one eye only.

3. An anonymous reviewer asked whether participants could conceivably complete the task if they did not know the meaning of the nouns. In this task, participants were shown pictures and were asked to name them. Because picture naming minimally requires

accessing the concept represented by the picture and retrieving the word that names the object (Potter, So, Von Eckhardt, & Feldman, 1984), we believe that participants could not complete the task unless they knew the meaning of the nouns represented in the pictures.

4. Although we followed the basic experimental design (i.e., a two-picture display) of Lew-Williams and Fernald (2007, 2010), there are a number of differences between our study and their studies. First, the experimental procedures employ different equipment and means of data coding. Specifically, the Lew-Williams and Fernald studies use the looking-while-listening technique, a procedure that was primarily developed for use with children (Fernald, Perfors, & Marchman, 2006). Although this procedure is similar to a visual-world design, participants are seated in a booth with two monitors presented side by side. A video camera is embedded between the two monitors and records the children as they turn to look at either monitor. In a highly labor-intensive procedure, data coders who are blind to the objectives of the experiment hand-code the location of eye movements of the video recorded session frame by frame with a standard refresh rate of 33 ms. In contrast, the participants in the experiments reported here were shown a visual scene on one monitor alone. The software that accompanies the eye-tracker determines the vertical and horizontal position of the eye and whether the eye is in fixation, blinking, or in saccadic movement by way of an internal algorithm. To our knowledge, no study has directly compared the results of an eye-tracking study using the visual-world paradigm with the looking-while-listening procedure. Second, as the focus of the original Lew-Williams and Fernald study (2007) was very young children, the experimental materials and procedure also differ from our experiments. The original study, which was followed in the 2010 study, included eight experimental objects that were repeated eight times, four times as the target and four times as the distractor (Lew-Williams & Fernald, 2007). Furthermore, target objects were embedded in sentence-final position in simple Spanish sentences. In contrast, the stimuli used in the experiments reported here were not repeated. Participants viewed targets and distractors only once during the experimental session. Additionally, target items were embedded in variable sentential contexts that were distinct in each trial. Due to these methodological and experimental differences, a replication of the results found in Lew-Williams and Fernald (2007) was warranted to determine whether our experimental procedure would be comparable. Ultimately, we replicated the Lew-Williams and Fernald study with a Spanish monolingual control group, which leads us to interpret the differences that we found in the bilingual groups as attributable to proficiency.

REFERENCES

Ainsworth-Darnell, K., Shulman, H. G., & Boland, J. E. (1998). Dissociating brain responses to syntactic and semantic anomalies: Evidence from Event-Related Potentials. *Journal of Memory and Language*, *38*, 112–130.

Alameda, J. R., & Cuetos, F. (1995). *Diccionario de las unidades lingüísticas del castellano: Volumen I: Orden alfabético / Volumen II: Orden por frecuencias* [Dictionary of the linguistic units of Spanish: Volume I: Ordered alphabetically / Volume II: Ordered by frequencies]. Oviedo, Spain: Servicio de publicaciones de la Universidad de Oviedo.

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition: Evidence for continuous mapping models. *Journal of Memory and Language*, *38*, 419–439.

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*, 247–264.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. New York: Cambridge University Press.

Barber, H., & Carreiras, M. (2005). Grammatical gender and number agreement in Spanish: An ERP comparison. *Journal of Cognitive Neuroscience*, *17*, 137–153.

Bates, E., Devescovi, A., Hernández, A., & Pizzamiglio, L. (1996). Gender priming in Italian. *Perception & Psychophysics*, *85*, 992–1004.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International*, *5*, 341–345.

Bruhn de Garavito, J., & White, L. (2002). L2 acquisition of Spanish DPs: The status of grammatical features. In A. T. Pérez-Leroux & J. M. Liceras (Eds.), *The acquisition of Spanish morphosyntax: The L1/L2 connection* (pp. 151–176). Dordrecht: Kluwer.

Carello, C., Lukatela, G., & Turvey, M. T. (1988). Rapid naming is affected by association but not by syntax. *Memory & Cognition*, *16*, 187–195.

Carroll, S. (1989). Second-language acquisition and the computational paradigm. *Language Learning*, *39*, 535–594.

Colé, P., & Segui, J. (1994). Grammatical incongruency and vocabulary types. *Memory & Cognition*, *22*, 387–394.

Cooper, R. (1974). The control of eye-fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory and language processing. *Cognitive Psychology*, *6*, 84–107.

Costa, A., Kovacic, D., Fedorenko, E., & Caramazza, A. (2003). The gender congruency effect and the selection of free-standing and bound morphemes: Evidence from Croatian. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *29*, 1270–1282.

Cubelli, R., Lotto, L., Paolieri, D., Girelli, M., & Job, R. (2005). Grammatical gender is selected in bare noun production: Evidence from the picture-word interference paradigm. *Journal of Memory and Language*, *53*, 42–59.

Cudeck, R., & Klebe, K. J. (2002). Multiphase mixed-effects models for repeated measures data. *Psychological Methods*, *7*, 41–63.

Dahan, D., Swingley, D., Tanenhaus, M. K., & Magnuson, J. S. (2000). Linguistic gender and spoken-word recognition in French. *Journal of Memory and Language*, *42*, 465–480.

Dijkstra, T. (2005). Bilingual visual word recognition and lexical access. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 179–201). Oxford: Oxford University Press.

Dowens, M. G., Vergara, M., Barber, H. A., & Carreiras, M. (2010). Morphosyntactic processing in late second-language learners. *Journal of Cognitive Neuroscience*, *22*, 1870–1887.

Duffield, N., & White, L. (1999). Assessing L2 knowledge of Spanish clitic placement: Converging methodologies. *Second Language Research*, *15*, 133–160.

Epstein, S. D., Flynn, S., & Martohardjono, G. (1996). Second language acquisition: Theoretical and experimental issues in contemporary research. *Behavioral and Brain Sciences*, *19*, 677–758.

Faussart, C., Jakubowicz, C., & Costes, M. (1999). Gender and number processing in spoken French and Spanish. *Rivista di Linguistica*, *11*, 75–101.

Fernald, A., Perfors, A., & Marchman, V. A. (2006). Picking up speech in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental Psychology*, *42*, 98–116.

Fernández-García, M. (1999). Patterns of gender agreement in the speech of second language learners. In J. Gutiérrez-Rexach & F. Martínez-Gil (Eds.), *Advances in Hispanic linguistics: Papers from the 2nd Hispanic Linguistics Symposium* (pp. 3–15). Somerville, MA: Cascadilla Press.

Fischer, B. (1992). Saccadic reaction time: Implications for reading, dyslexia and visual cognition. In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading* (pp. 31–45). New York: Springer Verlag.

Franceschina, F. (2001). Against an L2 morphological deficit as an explanation for the differences between native and non-native grammars. In S. H. Foster-Cohen & A. Nizegorodcew (Eds.), *EUROSLA Yearbook: Vol. 1* (pp. 143–158). Amsterdam: Benjamins.

Franceschina, F. (2005). *Fossilized second language grammars: The acquisition of grammatical gender*. Amsterdam: Benjamins.

Friederici, A. D., & Jacobsen, T. (1999). Processing grammatical gender during comprehension. *Journal of Psycholinguistic Research*, *28*, 467–484.

Foucart, A., & Frenck-Mestre, C. (2011). Grammatical gender processing in L2: Electrophysiological evidence of the effect of L1-L2 syntactic similarity. *Bilingualism*, *14*, 379–399.

Foucart, A., & Frenck-Mestre, C. (2012). Can late L2 learners acquire new grammatical features? Evidence from ERPs and eye-tracking. *Journal of Memory and Language*, *66*, 226–248.

Gess, R., & Herschensohn, J. (2001). Shifting the DP parameter: A study of Anglophone French L2ers. In J. Camps & C. R. Wiltshire (Eds.), *Romance syntax, semantics and their L2 acquisition* (pp. 105–119). Amsterdam: Benjamins.

Gouvea, A. C., Phillips, C., Kazanina, N., & Poeppel, D. (2010). The linguistic processes underlying the P600. *Language and Cognitive Processes*, *25*, 149–188.

Grosjean, F., Dommergues, J.-Y., Cornu, E., Guillelmon, D., & Besson, C. (1994). The gender-marking effect in spoken word recognition. *Perception & Psychophysics*, *56*, 590–598.

Gurjanov, M., Lukatela, G., Lukatela, K., Savić, M., & Turvey, M. T. (1985). Grammatical priming of inflected nouns by the gender of possessive adjectives. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *11*, 692–701.

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

Hawkins, R., & Chan, C. Y. (1997). The partial availability of Universal Grammar in second language acquisition: The "failed functional features hypothesis." *Second Language Research*, *13*, 187–226.

Hawkins, R., & Franceschina, F. (2004). Explaining the acquisition and non-acquisition of determiner-noun gender concord in French and Spanish. In P. Prévost & J. Paradis (Eds.), *The acquisition of French in different contexts* (pp. 175–205). Amsterdam: Benjamins.

Hernandez, A. E., Kotz, S. A., Hofmann, J., Valentin, V. V., Dapretto, M., & Bookheimer, S. Y. (2004). The neural correlates of grammatical gender decisions in Spanish. *Brain Imaging*, *15*, 863–866.

Jacobsen, T. (1999). Effects of grammatical gender on picture and word naming: Evidence from German. *Journal of Psycholinguistic Research*, *28*, 499–514.

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

Jescheniak, J. D. (1999). Gender priming in picture naming: Modality and baseline effects. *Journal of Psycholinguistic Research*, *28*, 729–737.

Kreysa, H., & Knoeferle, P. (2011). Effects of speaker gaze on spoken language comprehension: Task matters. In L. Carlson, C. Hölscher, & T. F. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 1557–1562). Austin, TX: Cognitive Science Society.

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, *33*, 149–174.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*, 1126–1177.

Lew-Williams, C., & Fernald, A. (2007). Young children learning Spanish make rapid use of grammatical gender in spoken word recognition. *Psychological Science*, *18*, 193–198.

Lew-Williams, C., & Fernald, A. (2010). Real-time processing of gender-marked articles by native and non-native Spanish speakers. *Journal of Memory and Language*, *63*, 447–464.

McCarthy, C. (2007). *Morphological variability in second language Spanish* (Unpublished doctoral dissertation). McGill University, Montreal, Canada.

Montrul, S. (2004). Subject and object expression in Spanish heritage speakers: A case of morphosyntactic convergence. *Bilingualism: Language and Cognition*, *7*, 125–142.

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

Moreno, E. M., & Kutas, M. (2005). Processing semantic anomalies in two languages: An electrophysiological exploration in both languages of Spanish-English bilinguals. *Cognitive Brain Research*, *22*, 205–220.

Morgan-Short, K., Sanz, C., Steinhauer, K., & Ullman, M. T. (2010). Acquisition of gender agreement in second language learners: An event-related potential study. *Language Learning*, *60*, 154–193.

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, *31*, 785–806.

Potter, M. C., So, K.-F., Von Eckardt, B., & Feldman, L. B. (1984). Lexical and conceptual representation in beginning and proficient bilinguals. *Journal of Verbal Learning and Verbal Behavior*, *23*, 23–38.

Sabourin, L., & Stowe, L. A. (2008). Second language processing: When are first and second language processed similarly? *Second Language Research*, *24*, 397–430. **AQ7**

Sabourin, L., Stowe, L. A., & de Haan, G. J. (2006). Transfer effects in learning a second language grammatical gender system. *Second Language Research*, *22*, 1–29.

Saslow, M. G. (1967). Latency for saccadic eye movement. *Journal of the Optical Society of America*, *57*, 1030–1033.

Schiller, N. (2009, July). *Grammatical gender in L2 from a processing perspective.* Paper presented at the 7th International Symposium on Bilingualism, Utretch, Netherlands.

Schmidt, R. (1986). Was weiss der Artikel vom Hauptwort? Ein Beitrag zur Verarbeitung syntaktischer Beziehungen beim Lesen [What does the article know about the noun? A study of syntactic agreement processing during reading]. *Zeitschrift für Experimentelle und Angewandte Psychologie*, *23*, 150–163.

Schwartz, B. D., & Sprouse, R. A. (1994). Word order and nominative case in non-native language acquisition: A longitudinal study of (L1 Turkish) German interlanguage. In T. Hoekstra & B. D. Schwartz (Eds.), *Language acquisition studies in generative grammar* (pp. 317–368). Amsterdam: Benjamins.

Schwartz, B. D., & Sprouse, R. A. (1996). L2 cognitive states and the Full Transfer/Full Access model. *Second Language Research*, *12*, 40–72.

Tanenhaus, M. K., & Spivey-Knowlton, M. J. (1996). Eye-tracking. *Language and Cognitive Processes*, *11*, 583–588.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *286*, 1632–1634.

Tanenhaus, M. K., & Trueswell, J. C. (2006). Eye movement and spoken language comprehension. In M. J. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (pp. 835–862). Amsterdam: Elsevier.

Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition*, *73*, 89–134.

van Berkum, J. J. A. (1996). *The psycholinguistics of grammatical gender*. Nijmegen, Netherlands: Nijmegen University Press.

White, L. (1989). Linguistic universals, markedness and learnability: Comparing two different approaches. *Second Language Research*, *5*, 127–140.

White, L. (2003). *Second language acquisition and Universal Grammar.* New York: Cambridge University Press.

White, L., Valenzuela, E., Kozlowska-MacGregor, M., & Leung, Y.-K. I. (2004). Gender and number agreement in nonnative Spanish. *Applied Psycholinguistics*, *25*, 105–133.

Wicha, N. Y. Y., Orozco-Figueroa, A., Reyes, I., Hernández, A., Gavaldón de Barreto, L., & Bates, E. (2005). When zebras become painted donkeys: Grammatical gender and semantic priming interact during picture integration in a spoken Spanish sentence. *Language and Cognitive Processes*, *20*, 553–587.

APPENDIX

NAMES OF PICTURED OBJECTS USED IN THE EYE-TRACKING EXPERIMENT

abanico	cementerio	gusano	oreja
abeja	cerebro	harina	pájaro
abrigo	chaleco	hidrante	pato
árbol	clavo	hormiga	pega
arroz	columpio	huella	pera
ballena	copa	hueso	perro
banco	corbata	huevo	piedra
bandera	corona	iglesia	piscina
barco	correa	jamón	pistola
bolígrafo	cuchillo	jaula	plato
bolsa	cuerno	ladrillo	playa
bomba	diamante	lámpara	pozo
botella	diente	lavamanos	puente
bufanda	enchufe	leche	puerta
búho	escoba	lengua	rana
caballo	escorpión	libélula	reloj
cabello	escudo	libro	remolacha
cadena	espada	llave	rosa
caja	espárrago	lupa	rueda
cama	espátula	maíz	semáforo
camión	espina	mano	silla
camisa	estatua	medalla	sopa
campana	falda	media	teclado
candado	flauta	mosca	tenedor
cangrejo	flecha	muñeca	trampa
caracol	fósforo	murciélago	uña
carro	fuego	nariz	vaso
casa	gato	nido	venado
castillo	grapadora	nube	vestido
castillo	guante	ojo	zapato

Author Queries

AQ1	Change "*p* = .22" to "*p* = .220" (added zero), for consistency with most other *p* values, which have 3 digits past the decimal? See here and throughout.
AQ2	Pls check format of *SE* values. Note that the 2 examples here have 1 digit past the decimal, but most have 2 digits or 4 digits. One example later on has 3 digits. Should these be consistent?
AQ3	Change "$t(15) = -2.5$" to "$t(15) = -2.50$" (added zero), for consistency with most other *t* values, which have 2 digits past the decimal?
AQ4	Change "$t(14) = -5.4$" to "$t(14) = -5.40$" (added zero), for consistency with most other *t* values, which have 2 digits past the decimal?
AQ5	Edits ok? Or change to "and it is the first to suggest that…"?
AQ6	Consider rewording for clarity? E.g., "and what modulation effects the L1 might have on that acquisition time course…"
AQ7	Changed "our" to "are" (When our first and second…) in Sabourin and Stowe 2008. Correct?
AQ8	Note repetition of "castillo" in appendix. Delete?
AQ9	Spelled Von Eckardt (rather than Eckhardt) in Refs. Which is correct?