Language Learning ISSN 0023-8333

Second Language Acquisition of Gender Agreement in Explicit and Implicit Training Conditions: An Event-Related Potential Study

Kara Morgan-Short
University of Illinois at Chicago

Cristina Sanz
Georgetown University

Karsten Steinhauer
McGill University

Michael T. Ullman
Georgetown University

This study employed an artificial language learning paradigm together with a combined behavioral/event-related potential (ERP) approach to examine the neurocognition of the

We would like to thank Harriet Wood Bowden, Helen Carpenter, David Birdsong, Mandy Faretta, Ron Leow, Marco Piñeyro, and John Stowe for their invaluable support in various aspects of the project. We are also grateful for the helpful comments offered by Jean-Marc Dewaele and two anonymous reviewers. Any errors or omissions are ours alone. Previous versions of this study were presented at Second Language Research Forum in 2008 and at the annual meeting of the Cognitive Neuroscience Society in 2009. Support for this project was provided to MTU by the National Institute of Health under RO1 MH58189 and RO1 HD049347 and by the National Science Foundation under Doctoral Dissertation Improvement Grant 0446836; to KMS by a Georgetown University Dissertation Fellowship and by the National Institute of Health under Ruth L. Kirschstein National Research Service Award 1 F31 MH67407-01A1 from the NIMH; and to KS by the Canada Research Chair program and the Canada Foundation for Innovation (CRC/CFI; project 201876) and the Natural Sciences and Engineering Research Council of Canada (NSERC; RGPGP 312835).

Correspondence concerning this article should be addressed to Kara Morgan-Short, Departments of Spanish, French, Italian and Portuguese and Psychology, 1706 UH MC 315, University of Illinois at Chicago, Chicago, IL 60607. Internet: karams@uic.edu; Michael T. Ullman, Department of Neuroscience, Georgetown University, Washington, DC 20057. Internet: michael@georgetown.edu

processing of gender agreement, an aspect of inflectional morphology that is problematic in adult second language (L2) learning. Subjects learned to speak and comprehend an artificial language under either explicit (classroomlike) or implicit (immersionlike) training conditions. In each group, both noun-article and noun-adjective gender agreement processing were examined behaviorally and with ERPs at both low and higher levels of proficiency. Results showed that the two groups learned the language to similar levels of proficiency but showed somewhat different ERP patterns. At low proficiency, both types of agreement violations (adjective, article) yielded N400s, but only for the group with implicit training. Additionally, noun-adjective agreement elicited a late N400 in the explicit group at low proficiency. At higher levels of proficiency, noun-adjective agreement violations elicited N400s for both the explicit and implicit groups, whereas noun-article agreement violations elicited P600s for both groups. The results suggest that interactions among linguistic structure, proficiency level, and type of training need to be considered when examining the development of aspects of inflectional morphology in L2 acquisition.

Keywords second language acquisition; event-related potentials; language processing; explicit; implicit; agreement

Aspects of inflectional morphology, including grammatical gender agreement in noun phrases (NPs), seem to be particularly difficult for late second language (L2) learners to acquire (Montrul, 2004; Montrul, Foote, Perpiñán, Thornhill, & Vidal, 2008; White, 2003). Even for learners at advanced levels of proficiency, errors in gender agreement appear to persist (Dewaele & Veronique, 2001; Franceschina, 2005). Although this issue has been addressed from multiple perspectives (Arteaga, Herschensohn, & Gess, 2003; Bartning, 2000; Benati, 2005; De Jong, 2005; Dewaele & Veronique, 2001; Franceschina, 2005; Gass & Alvarez Torres, 2005; Hawkins & Chan, 1997; Keating, 2009; Montrul, 2004; Montrul et al., 2008; White, Valenzuela, Kozlowska-Macgregor, & Leung, 2004), neurocognitive research has only begun to consider the development, representation, and processing of L2 grammatical gender (Sabourin & Haverkort, 2003; Sabourin & Stowe, 2008; Tokowicz & MacWhinney, 2005). The current article aims to contribute to the neurocognitive evidence pertaining to the L2 acquisition and processing of grammatical gender by examining the online processing of noun-phrase gender agreement on both articles and adjectives as affected by two different types of language training conditions—explicit (classroomlike) and implicit (immersionlike)—at both low and high proficiency. After reviewing relevant theoretical perspectives and previous neurocognitive evidence, we will report a behavioral and electrophysiological study of grammatical gender processing in an artificial language and will consider the results in light of neurocognitive models and previous evidence.

Review of Literature

Theoretical Perspectives

Second language learners' acquisition and processing of gender agreement has been addressed from a variety of perspectives. Some researchers account for gender agreement difficulties in L2 by positing either that late L2 learners do not have access to certain aspects of universal grammar that are available in first language (L1) acquisition (Franceschina, 2005; Hawkins & Chan, 1997) or that full access to aspects of universal grammar is retained but is potentially limited by processing or performance constraints (Montrul, 2004; Montrul et al., 2008; White et al., 2004). Other researchers have empirically explored L2 acquisition of gender agreement considering issues such as the effects of input and interaction on acquisition (Gass & Alvarez Torres, 2005), the relationship between acquisition and purported stages of processability (Bartning, 2000; Dewaele & Veronique, 2001), and the effects of the provision or the absence of explicit (metalinguistic) information (Arteaga et al., 2003; Benati, 2005; De Jong, 2005).

A few researchers have also begun to empirically explore the processes involved in L2 gender agreement from psycholinguistic (Keating, 2009) and neurocognitive perspectives (Davidson & Indefrey, 2009; Sabourin & Haverkort, 2003; Sabourin & Stowe, 2008; Tokowicz & MacWhinney, 2005). Neurocognitive theories, however, have thus far not directly addressed gender agreement but have rather taken a broader view.

The declarative/procedural (DP) model (Ullman, 2001, 2005) is of particular interest here, because its predictions can be directly tied to outcomes from event-related potentials (ERPs), the approach used here. This model suggests that lexical/semantic aspects of both the L1 and L2 rely on the same set of neurocognitive mechanisms—specifically, declarative memory, a temporal-lobe-based system that also underlies (explicit as well as implicit) nonlinguistic aspects of semantic and other knowledge. In contrast, the mechanisms underlying the learning, representation, and processing of aspects of grammar initially differ between the L1 and L2. On this view, in the L1, aspects of grammar—in particular, rule-governed structure building—are generally subserved by procedural memory, an implicit memory system rooted in frontal/basal-ganglia circuits that also underlies motor and nonlinguistic cognitive skills, and may be specialized for sequences and rules. In the L2, however—in particular, at lower experience levels—learners do not generally depend on these L1 mechanisms. Instead, they are posited to rely on lexical/semantic processes (in declarative memory) for these same functions. However, with increasing L2 experience, these aspects of grammar may come to rely more and more on the same grammatical/procedural mechanisms as those that underlie L1 grammar—although such "proceduralization" will depend on a number of factors, including the type and the amount of L2 experience and training, as well as individual differences, such as procedural learning abilities.

Other models have made somewhat similar claims, focusing, however, less on the neurocognitive substrates that are of interest here. Thus, both the shallow-structure hypothesis (SSH; Clahsen & Felser, 2006a, 2006b) and the view espoused by Paradis (2004) are similar to the DP model in positing that, at least initially, L2 learners do not rely on L1 mechanisms for grammar, although with increasing experience, L2 learners may increasingly rely on these mechanisms.

In contrast, Indefrey (2006) and Abutalebi (2008) have argued that grammatical processing relies on essentially the same neural substrates in the L2 and L1, although these are sometimes recruited more strongly during L2 processing (depending on tasks, subject groups, and regions; for more on this perspective, see also Hernandez and Li, 2007, and Abutalebi, Cappa, & Perani, 2003). Similarly, the competition model (Hernandez, Li, & MacWhinney, 2005; MacWhinney, 2002, 2005) suggests that L1 and L2 processing rely on the same set of mechanisms.

A variety of behavioral and neurocognitive methods, including electrophysiological and neuroimaging techniques, have been used to examine the neurocognition of L2 and to help distinguish among neurocognitive L2 models. Below, we review electrophysiological evidence from studies using the ERP technique employed here.

ERP Evidence

The acquisition of ERPs is one widely used method for exploring the neurocognition of language processing. ERPs reflect the real-time electrophysiological brain activity of cognitive processes that are time-locked to the presentation of target stimuli (for an overview of ERP components, recording, and analysis, see Luck, 2005). Language-related ERP research often employs a violation paradigm for presenting linguistic stimuli. In this paradigm, the ERP response to a linguistic violation (e.g., lexical, syntactic, morphosyntactic) is compared to the ERP response to a matched control word or structure. Various types of violations (also called difficulties, disruptions, anomalies, etc.) have been shown to elicit particular ERP components in the L1 (for recent comprehensive reviews, see Friederici, 2002; Kaan, 2007; Kutas & Schmitt, 2003; Steinhauer & Connolly, 2008). Of relevance here are three components:

1. *N400s*. Difficulties in lexical/semantic processing elicit central/posterior bilaterally distributed negativities (N400s) that often peak about 400 ms after stimulus onset (Friederici, Steinhauer, & Frisch, 1999; Kutas & Hillyard, 1980). N400s importantly involve bilateral temporal lobe structures (McCarthy, Nobre, Bentin, & Spencer, 1995; Simos, Basile, & Papanicolaou, 1997) and have been posited to depend on lexical/declarative memory (Ullman, 2001, 2004, 2005).
2. *LANs*. Disruptions of rule-governed syntactic, morphosyntactic, and morphophonological processing can yield (although not always; Hagoort & Brown, 1999; Osterhout, Bersick, & McLaughlin, 1997) early (150–500 ms) left anterior negativities (LANs; Friederici, Pfeifer, & Hahne, 1993; Neville, Nicol, Barss, Forster, & Garrett, 1991), which have been linked to rule-based automatic structure-building computations (Friederici, Hahne, & Mecklinger, 1996; Hahne & Friederici, 1999) and left frontal structures (Friederici, von Cramon, & Kotz, 1999). LANs have been posited to depend on the grammatical/procedural memory system (Ullman, 2001, 2004, 2005).
3. *P600s*. Syntactic word-order and morphosyntactic processing difficulties also generally elicit late (600 ms) centro-parietal positivities (P600s; Kaan, Harris, Gibson, & Holcomb, 2000; Osterhout & Holcomb, 1992). The P600 has been associated with controlled processing and structural reanalysis (Friederici et al., 1996; Hahne & Friederici, 1999). Although P600s do not appear to depend on frontal brain structures (Friederici, von Cramon et al., 1999), evidence suggests that they involve the basal ganglia (Friederici & Kotz, 2003; Friederici, Kotz, Werheid, Hein, & von Cramon, 2003), possibly due to the attentional role of these structures (Ullman, 2004, 2006).

Event-related potential studies of L2 processing have revealed the following general pattern: Although lexical/semantic processing in the L2 does not differ qualitatively from the L1—in both cases eliciting N400s—aspects of L2 grammatical (syntactic and morphosyntactic) processing seem to show important L1/L2 differences, at least at lower levels of proficiency (for comprehensive reviews, see Mueller, 2005; Osterhout, McLaughlin, Pitkänen, Frenck-Mestre, & Molinaro, 2006; Steinhauer, White, & Drury, 2009; Ullman, 2001, 2005). Specifically, for grammatical processing at lower levels of L2 proficiency LANs are absent, with subjects instead generally showing no negativity at all (Bowden, Sanz, Steinhauer, & Ullman, 2010; Hahne & Friederici, 2001; Mueller, Hahne, Fujii, & Friederici, 2005; Ojima, Nakata, & Kakigi, 2005) or N400s or N400-like posterior negativities (Osterhout et al., 2006, 2008; Weber-Fox & Neville, 1996). However, in recent studies, LANs have been found in higher experience and proficiency L2 speakers (Bowden et al., 2010; Friederici, Steinhauer, & Pfeifer, 2002; Ojima et al., 2005; Rossi, Gugler, Friederici, & Hahne, 2006; but see Chen, Shu, Liu, Zhao, & Li, 2007). Finally, P600s are generally—but not always—(Hahne & Friederici, 2001; Ojima et al., 2005) present in the L2, particularly at higher levels of proficiency (Bowden et al., 2010; Friederici et al., 2002; Mueller et al., 2005; Osterhout et al., 2006; Tokowicz & MacWhinney, 2005; Weber-Fox & Neville, 1996).

We now turn to ERP studies that have specifically examined L1 and L2 grammatical gender agreement in nominal phrases, the focus of the present study. The majority of ERP studies that have examined L1 noun-phrase gender agreement have focused on noun-article agreement violations in sentential contexts. Most such studies have reported a P600 response (Barber & Carreiras, 2005; Davidson & Indefrey, 2009; Gunter, Friederici, & Schriefers, 2000; Hagoort, 2003; Hagoort & Brown, 1999; Molinaro, Vespignani, & Job, 2008; Wicha, Moreno, & Kutas, 2004), which in some cases is preceded by a LAN (Barber & Carreiras, 2005; Gunter et al., 2000; Molinaro et al., 2008; Sabourin & Stowe, 2008). In exceptional cases—in particular, when the violation occurs in isolated word pairs (Barber & Carreiras, 2005) or when the violation is on the last word of the sentence (Hagoort, 2003; Hagoort & Brown, 1999)—an N400 effect may be elicited in addition to a LAN (Barber & Carreiras, 2005) or P600 (Hagoort, 2003; Hagoort & Brown, 1999). Relatively few ERP studies have examined L1 noun-adjective gender agreement violations. Of those that have, results have found that (a) gender agreement violations on attributive adjectives elicit a P600 response in sentential contexts (Davidson & Indefrey, 2009; Sabourin & Haverkort, 2003) and an N400 in isolated word pairs (Barber & Carreiras, 2003, 2005) and (b) violations on predicate adjectives in sentential contexts elicit a P600 (Demestre & Garcia-Albea, 2007) or a biphasic LAN/P600 response (Barber & Carreiras, 2005; Demestre, Meltzer, Garcia-Albea, & Vigil, 1999). Overall, it appears that ERP responses to L1 gender agreement violations are largely consistent with those found for morphosyntactic processing more generally (i.e., a LAN and/or P600 response). These have usually been interpreted in line with broader views of these ERP components—for example, that the LAN in such gender agreement violations may reflect "the detection of a mismatch between morphosyntactic features," whereas the P600 reflects aspects of reanalysis and repair

(Barber & Carreiras, 2005). In the less common cases in which an N400 has been found, interpretations have placed less emphasis on morphosyntactic processing and more on the integration of lexical representations (e.g., between isolated noun-adjective pairs) (Barber & Carreiras, 2003) and semantic end-of-sentence wrap-up effects (Hagoort, 2003; Hagoort & Brown, 1999).

To our knowledge, only a handful of ERP studies have examined L2 gender agreement. In one set of studies, Sabourin and colleagues examined L2 gender agreement in definite and indefinite noun phrases for L2 learners of Dutch with different L1 backgrounds (Sabourin & Haverkort, 2003; Sabourin & Stowe, 2008). Sabourin and Haverkort (2003) examined article-noun (in definite noun phrases) and adjective-noun (in indefinite noun phrases) gender agreement violations in Dutch sentences, in both native and L2 speakers. L2 subjects were native speakers of German who had lived in the Netherlands for approximately 10 years and were highly proficient in Dutch. In the native speakers, both types of violations elicited a P600 effect as well as a late negativity. In the L2 speakers, adjective-noun violations yielded no reported ERP effects, whereas article-noun violations showed P600s only. Interestingly, in a grammaticality judgment task the L2 subjects showed lower accuracy than the L1 subjects on adjective-noun agreement (across correct and violation sentences) but not on article-noun agreement. Given that the ERP analyses in this study were performed only on correctly judged sentences, the behavioral difference between the two types of agreement could help explain the observed ERP differences, as there would have been less power in the ERP analyses on adjective-noun agreement. Additionally, the authors note that gender agreement in definite NPs (as tested by article-noun agreement) but not in indefinite NPs (as tested by adjective-noun agreement) is structurally similar in Dutch and German, and they suggested that the P600 found for article-noun agreement violations may reflect a reliance on "L1 processing strategies to process their L2" (Sabourin & Haverkort, 2003, p. 192). This interpretation was supported by results reported in Sabourin and Stowe (2008), which examined gender agreement in L2 Dutch definite noun phrases for L1 German speakers and L1 Romance language speakers. The L1 German speakers, whose agreement system is similar to Dutch, evidenced P600s, whereas the L1 Romance speakers, whose agreement systems are different than Dutch, did not. Thus, both studies suggest that L2 gender agreement may be processed in an L1-like manner when the gender agreement system in the L1 and L2 is similar. Sabourin and Haverkort (2003) interpreted their findings in terms of the DP model, suggesting that when L1 and L2 are similar, procedural memory may be used by advanced adult L2 learners, whereas when L1 and L2 are different, L2 may be subserved by more general cognitive strategies.

Tokowicz and MacWhinney (2005) examined the processing of L2 article-noun gender agreement in Spanish sentences, among other linguistic structures. L2 subjects were native speakers of English enrolled in any of the first four semesters of beginning university-level Spanish. (No native speaker of Spanish subjects was tested, although the native speakers of English were tested in English on violations other than gender.) The L2 subjects showed P600s in response to article-noun gender agreement violations, even though they did not perform above chance on grammaticality judgments of the same sentences. Note that unlike Sabourin and Haverkort (2003), Tokowicz and MacWhinney performed ERP analyses over both correctly and incorrectly judged sentences. Finally, note that the P600 effect on gender agreement was observed in L2 speakers whose L1 (English) has no gender agreement, indicating that at least in this case L1-L2 transfer is unlikely to explain the P600.

In an L2 training study, Davidson and Indefrey (2009) assessed L1 Dutch speakers' development, over a short period of time, of morphosyntactic aspects of L2 German—specifically, adjective declension and both article-noun and adjective-noun gender agreement. During the first of two experimental sessions, a pretest was administered in which subjects judged the acceptability of German phrases. Subsequently in the same session, subjects were presented with learning tasks that included information about German words and grammatical information about adjective declension and gender. After the learning tasks, subjects completed a training task in which correct/incorrect feedback was provided (unlike in the pretest) as subjects again judged the acceptability of (new) German phrases. One week later, subjects returned for a posttest in which they made acceptability judgments (with no feedback) on the items that had been presented during the pretest and the training task. A group of German native speakers also completed this final task. Behavioral and ERP data were recorded during the judgment tasks in the pretest, the training task, and the posttest. The German native speakers showed a P600 for adjective declension and for both article-noun and adjective-noun gender agreement. Although the L2 learners improved their judgment accuracy during the training and posttest phases for both declension and both types of gender agreement, they developed a P600 response only for declension. It is worth noting that in the posttest, the L2 learners did not differ in judgment accuracy from the L1 speakers on declension but were worse on both types of gender agreement. Overall, the results suggest that such a short training session may be sufficient for the development of a P600 response in L2 learners for adjectival declension but not for either article-noun or adjective-noun gender agreement anomalies.

The data from these studies suggest that L2 nominal gender agreement processing is partially, but not entirely, consistent with L1. On the one hand, L2 gender agreement violations have been found to elicit P600s in at least some cases—in particular, on article-noun violations in learners who had at least a few months of exposure to the target language and in which the L1 either has no gender agreement (Tokowicz & MacWhinney, 2005) or has a similar gender agreement structure to the L2 (Sabourin & Haverkort, 2003; Sabourin & Stowe, 2008). However, P600s were not found in several studies, specifically in response to (a) article-noun (Sabourin & Stowe, 2008) and adjective-noun (Sabourin & Haverkort, 2003) gender agreement violations in L2 learners whose L1 has a different gender agreement structure, or (b) article-noun and adjective-noun gender violations in L2 learners who had a very short amount of training (Davidson & Indefrey, 2009). Note also that P600s were not found in either study that examined adjective-noun gender violations (Davidson & Indefrey, 2009; Sabourin & Haverkort, 2003). Finally, no LANs have thus far been reported in L2 nominal gender agreement violations, despite their elicitation by nominal gender agreement violations in L1 (Barber & Carreiras, 2005; Demestre et al., 1999; Gunter et al., 2000; Molinaro et al., 2008; Sabourin & Stowe, 2008), as well as by other morphosyntactic violations in L2 (Ojima et al., 2005; Rossi et al., 2006).

Although this evidence has begun to further our understanding of L2 gender agreement, we do not yet have a complete account of L2 gender agreement processing. In order to develop a fuller account, it will be particularly important to directly compare, ideally within subjects, adjective and article gender agreement, at both low and high proficiency. Additionally, as of yet, it is still unknown whether the type of training under which an L2 is learned—in particular, explicit, form-focused classroomlike instruction versus implicit, more naturalistic immersionlike instruction—affects the neurocognition of L2 gender agreement processing.

Research in the field of second language acquisition (SLA) has extensively explored the distinction between explicit and implicit training. Explicit training conditions are characterized by rule explanation or direction to attend to forms and/or arrive at rules, whereas implicit training conditions are characterized by the lack of rule explanation or direction to attend to forms (DeKeyser, 1995; Norris & Ortega, 2000, 2001). SLA research comparing such learning conditions has shown that explicitly trained groups often outperform implicitly trained groups (see DeKeyser, 2003; Norris & Ortega, 2000; Sanz &

Morgan-Short, 2005). At the same time, implicitly trained groups consistently show development for L2 structures over time and have performed as well as explicitly trained groups (De Jong, 2005; Rosa & O'Neill, 1999; Sanz & Morgan-Short, 2004, 2005; VanPatten & Oikkenon, 1996). Effects of explicit and implicit training conditions, however, may differ for particular linguistic forms, so it is important to investigate the effects of different conditions on different linguistic forms, such as gender agreement. Additionally, to our knowledge there has, to date, been no examination of the neurocognitive substrates resulting from explicit versus implicit language training conditions.

Thus, there are two main motivations for the current ERP study. First, the study aims to further elucidate the neurocognitive processes underlying L2 gender agreement, specifically contrasting the processing of adjective and noun gender agreement at both low and high proficiency. Second, the study will explore whether explicit and implicit training conditions differentially affect the neurocognitive processing of L2 gender agreement, for both adjectives and articles, at both low and high proficiency.

Methods

In order to address these questions, the study used an artificial language paradigm. Subjects received either explicit or implicit training on an artificial language, BROCANTO2, as well as extensive comprehension and production practice. Both the explicitly and implicitly trained subject groups[1] were tested at low proficiency and at the end of practice with both behavioral and neurocognitive (ERP) measures. Processing of noun-article and noun-adjective gender agreement was examined within each group. Comparisons were made between the two agreement types and between the explicitly and implicitly trained groups. (For an overview of the experimental design, see Figure 1.)

Subjects

We tested 41 right-handed (Oldfield, 1971) healthy adults. None of the subjects were fluent in a language other than English, based on self-report. Because the artificial language was structurally similar to Romance languages, the following additional criteria were imposed: (a) Subjects must not have studied any Romance language for more than 1 year in college, and not for more than 3 years total; (b) any classroom exposure to a Romance language must have occurred at least 2 years prior to their participation in the experiment; and (c) they must not have been immersed in a Romance language environment for more than 2 weeks at any time. Subjects were pseudo-randomly assigned

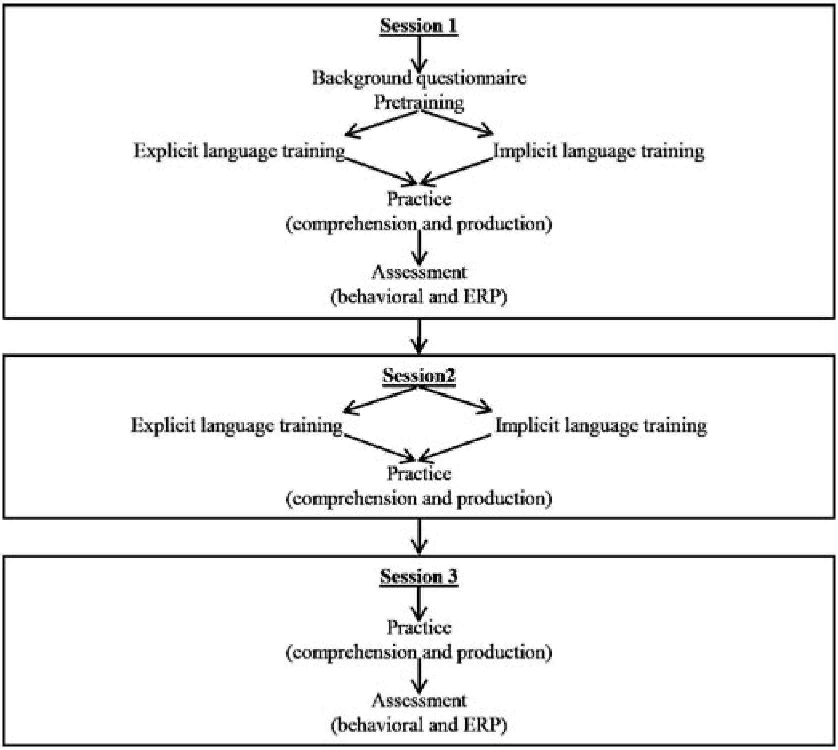

Figure 1 The experimental design consisted of three sessions during which background questionnaires, pretraining, explicit and implicit artificial language training, practice, and assessments were administered. Arrows indicate whether the subsequent experimental procedure was the same (downward and inward pointing arrows) or different (outward pointing arrows) for the explicit and implicit training conditions.

to the explicit or implicit training groups within each sex and were included in analysis if they reached at least a low level of proficiency (see below) during training and completed both ERP test sessions and all behavioral tasks. Of the 31 subjects who met the requirements for inclusion, 1 subject (in the implicit condition) was excluded from analysis because of the large number of artifacts in the second-session ERP data. Thus, data from 30 subjects (explicit: $n = 16$, 7 females, implicit: $n = 14$, 7 females) were analyzed. The explicit and implicit groups did not differ (using independent samples two-tailed t tests with equal variance assumed) on age (explicit: $M = 24.25$ years, $SD = 4.34$; implicit: $M = 24.71$ years, $SD = 5.57$; $t[28] = 0.26$, $p = .800$), years of education (explicit: $M = 16.25$, $SD = 2.82$; implicit: $M = 16.43$, $SD = 2.17$; $t[28] = 0.19$,

$p = .849$), or years of exposure (of any type) to either Romance languages (explicit: $M = 1.51$, $SD = 1.35$; implicit: $M = 1.95$, $SD = 1.30$; $t[28] = 0.91$, $p = .371$), or any other L2s (explicit: $M = 3.45$ years, $SD = 1.71$; implicit: $M = 4.94$ years, $SD = 3.46$; $t[28] = 1.52$, $p = .139$). All subjects gave written informed consent and received monetary compensation for their participation.

Materials

An artificial language rather than an existing natural language was used for several reasons, including the likelihood for subjects to reach high proficiency within a limited time; to control for phonological differences between L1 and L2, which are a common source of difficulty in SLA (Sanders, Neville, & Woldorff, 2002) and have recently been shown to affect the acquisition of morphosyntactic structures (Goad & White, 2006); to control whether the grammatical rules differ from those of the L1; and to meet additional constraints of EEG recording regarding the presentation of stimulus materials (e.g., identical baseline periods; see Friederici et al., 2002). Crucially, previous ERP and functional magnetic resonance imaging (fMRI) studies of a related artificial language (see below) have found brain activity typical of natural language processing (Friederici et al., 2002; Opitz & Friederici, 2003, 2004).

Subjects in the current study learned BROCANTO2, a modified version of BROCANTO (Friederici et al., 2002; Opitz & Friederici, 2003, 2004). BROCANTO2 is based on universal requirements of natural language, is fully productive, and was designed to have a structure with specific similarities to Romance languages and dissimilarities to English. Each of the 1,404 possible BROCANTO2 sentences was meaningful in that it described a move of a computer-based board game (Figure 2), which provided a context for the subjects to use the artificial language. The game, which was originally developed and successfully utilized for training subjects in BROCANTO (Friederici et al., 2002; Opitz & Friederici, 2003, 2004), is based on rules that are entirely independent of the artificial language. The lexicon of BROCANTO2 consists of 14 pronounceable nonce words: 2 articles, marked for gender (*li*, masculine; *lu*, feminine); 2 adjectives (*trois*-, *neim*-), each marked for gender (masculine *troise/neime*; feminine *troiso/neimo*); 4 nouns (*pleck, neep, blom, vode*, 2 of which are feminine and 2 masculine; the nouns are not overtly marked for gender, but their articles and adjectives must agree with them); 4 verbs (*klin, nim, yab, praz*); and 2 adverbs (*noyka, zayma*). Thus, the grammar consists of nominal phrases, in which articles and attributive adjectives are (a) morphologically marked so as to agree in gender with the noun that they refer to and

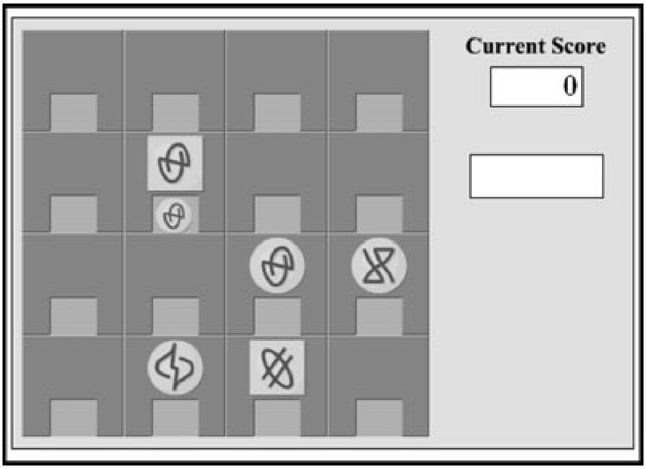

Figure 2 Game tokens are represented by visual symbols, which correspond to nouns in BROCANTO2. The tokens can further be distinguished by their background shape (square or round), each of which corresponds to a BROCANTO2 adjective. Players can move, swap, capture, and release tokens, with these actions corresponding to BROCANTO2 verbs, and they can move them either horizontally or vertically (corresponding to BROCANTO2 adverbs).

(b) postnominal, with an article following an adjective when an adjective is present. Verbal phrases have a fixed subject-object-verb word order and have no morphological features. Adverbs, when used, immediately follow the verb. See Table 1 for an example sentence.

Table 1 Sample control and violation BROCANTO2 sentences

Sentence type	BROCANTO2 stimuli					
Example	*Blom*	*neim-o*	*lu*	*neep*	*li*	*praz*
control	Blom-piece-*F*	square-*F*	the-*F*	neep-piece-*M*	the-*M*	switch
sentence	"The square blom switches the neep."					
Noun-adjective	*Blom*	*neim-e	*lu*	*neep*	*li*	*praz*
agreement	Blom-piece-*F*	*square-*M*	the-*F*	neep-piece-*M*	the-*M*	switch
violation						
Noun-article	*Blom*	*neim-o*	*li	*neep*	*li*	*praz*
agreement	Blom-piece-*F*	square-*F*	*the-*M*	neep-piece-*M*	the-*M*	switch
violation						

Note. * = violation; M = masculine; F = feminine.

Procedure

Before beginning explicit or implicit language training (see Figure 1), subjects completed (a) a background questionnaire; (b) pretraining, during which they received a brief, self-paced introduction to the computer-based game and learned the names of the four game tokens; and (c) a brief pretraining assessment, in which subjects had to successfully name all four game tokens several times. At this point, one of two computer-based aurally presented language training conditions, each lasting approximately 13 min, was presented. Both conditions (a) presented simple phrases initially and gradually moved to full and complex sentences, (b) contained equal proportions of the different words in each category, (c) provided no English translations, and (d) utilized computer-controlled timing of the presentation of the training.

The two training conditions crucially differed in that the explicit condition provided metalinguistic explanations and meaningful examples (phrases and sentences along with their corresponding game constellations and moves), whereas the implicit condition provided only meaningful examples. Specifically, in the explicit condition, explicit metalinguistic explanations regarding the functions and rules related to the nouns, articles, adjectives, verbs and adverbs (in that order), as well as 33 corresponding meaningful examples, were presented (see Appendix A for a sample section). The implicit condition presented the same 33 meaningful examples with an additional 94 meaningful examples (which were interspersed with and followed the same general type and order as the 33 examples) in order to balance for the time that was required for the metalinguistic explanation to be given in the explicit training condition (see Appendix B for a sample section). Overall, the training conditions were designed to approximate real-life language learning settings (e.g., traditional, form-focused L2 classroom settings vs. immersion settings) while controlling as tightly as possible other intervening variables such as time-on-task and order of input presentation (simple to complex linguistic input), which would have been difficult, if not impossible, to control for outside of a laboratory context.

After the initial explicit or implicit training period, subjects in both conditions practiced using BROCANTO2 by playing comprehension and production versions of the computer-based board game. The practice provided to the explicitly and implicitly trained subjects was identical. There were 22 blocks of each version, with 20 items (moves on the game board) in each block, for a total of 44 blocks (half comprehension, half production) and 880 items. Words within each grammatical class were presented with approximately equal frequency within blocks and throughout the total number of blocks.

The comprehension and production versions alternated every two blocks. For the comprehension version, subjects viewed a game board on the computer screen and were asked to respond to prerecorded aural statements in BROCANTO2 by attempting to make the stated move. For production blocks, subjects watched a move displayed on the screen and had to describe it with a single oral BROCANTO2 sentence. The experimenter then indicated the correctness of subjects' statement with a key press. In both the comprehension and production versions, the computer program displayed "correct" or "incorrect" as appropriate (no other feedback was given), added or deducted 10 points to a running score that was tabulated and displayed, and proceeded to the next item. In this way, the entire training set of moves remained the same for all subjects.[2] Subjects continued to play until they achieved a level of low proficiency, which was defined as performance on comprehension practice that was significantly higher than chance. When subjects reached this level, a first round of ERP assessments was administered (see below).

After the first round of ERP testing was completed, subjects returned for a second session (minimum: immediately after the first session; maximum: 4 days afterward; mean days between sessions $M = 1.53$, $SD = 1.25$) during which additional training and then practice was provided. The training was identical to the original training session, with the exact same input and examples as earlier. As earlier, practice involved alternating comprehension and production blocks, although with novel sentences (i.e., sentences that had not been presented previously). Subjects completed all blocks up to block 36 in the second session. The remaining eight blocks were completed in the third session (minimum: 1 day after the second session; maximum: 5 days afterward; mean days between second and third sessions $M = 2.35$, $SD = 1.41$), after which a second round of ERP testing was administered. Note that whereas the criterion for the first round of assessment was proficiency based (above-chance performance), the criteria for the second round of assessment was based on completing a certain number of practice blocks.

ERP Assessment

Event-related potential assessment was carried out with 240 BROCANTO2 sentences, crucially including 48 sentences each with an agreement violation and 48 matched correct control sentences. Additional violation and control sentences examining word order and argument structure are not discussed here. Agreement violation sentences were created from each of the 48 correct sentences by changing the morphosyntactic gender marking of either an adjective or an article (see Table 1). Thus, the correct and violation sentences differed only in the gender marking of either an adjective or an article. The violations were equally distributed between adjectives ($n = 24$) and articles ($n = 24$), as well as between the first and second noun phrases. Violations of articles occurred with and without intervening (nonviolated) adjectives with approximately equal frequency.[3] Both novel (not presented during training or practice) and repeated correct control sentences were selected, with the different words within each category (e.g., the four nouns) occurring with approximately equal frequency across the sentences of each type.[4] Violations and correct sentences were pseudo-randomly intermixed (with maximal distance between them) and distributed across four blocks of 62 trials each, with the presentation order of the blocks balanced across groups and the two sexes. Each subject was exposed to the entire set of stimuli, including the 48 agreement violation sentences and the matched 48 control sentences.

Event-related potential recording occurred in a dark, quiet testing room. Subjects sat in a comfortable chair 70 cm from a 16-in. CRT monitor. Prior to ERP recording, subjects were instructed to minimize eye and body movements during the acoustic presentation of the sentences, and they were given instructions on how and when to respond. During ERP data collection, the following presentation sequence occurred for each sentence: First, a fixation cross appeared in the center of the computer monitor, immediately followed by a BROCANTO2 sentence, which was presented auditorily via ER-4 insert earphones (Etymotic Research, Inc.). Words were separated by a 50-ms interval of silence, following Friederici et al. (2002). Immediately after the 50-ms interval following the last word of each sentence, the fixation cross on the monitor was replaced by the prompt “Good?” Subjects had up to 5 s to make a judgment about whether the sentence was good or bad, indicated with the buttons of a computer mouse (left for good, right for bad). After the subject responded, the program proceeded immediately to the next sentence. Scalp EEG was continuously recorded in DC mode at a sampling rate of 500 Hz from 64 electrodes (extended 10-20 system) mounted in an elastic cap (Electro-Cap International, Inc.) and analyzed using EEProbe software (Advanced Neuro Technology). Scalp electrodes were referenced to the left mastoid, and impedances were kept below 5 kΩ. The EEG was amplified by Neuroscan SynAmps[2] amplifiers and filtered online with a bandpass filter (DC to 100 Hz, 24-dB/octave attenuation). Eye-blinks (determined by reference to vertical EOG) as well as additional artifacts exceeding 75 Hz at any given scalp electrode were removed from the raw data. Offline, the EEG was filtered with a 0.16- to 30-Hz bandpass filter.

Statistical Analysis

In order to determine subjects' ability to discriminate correct and violation stimuli, their behavioral responses were transformed to d-prime scores. The d-prime value is an unbiased statistic that represents the ability to correctly discriminate between two types of stimuli—in this case, sentences that either are correct or contain a noun-adjective or a noun-article agreement violation (Macmillan & Creelman, 2005). Differences in the ability to discriminate correct and violation sentences were examined by submitting d-prime scores for each subject to a 2 × 2 × 2 ANOVA with Level (low proficiency, end-of-practice) and Agreement Type (noun-adjective, noun-article) as repeated factors and with Group (explicit, implicit) as the between-subjects factor. In order to clarify the nature of any interactions, all significant interactions from the model were followed up with specific step-down ANOVAs (based on the factors included in the interaction) and motivated comparisons of simple effects.

For ERP analysis, EEG data time-locked to the onset of the violation or matched control article or adjective (i.e., the onset of the target word) were averaged for each subject for a subset of 15 electrodes (12 lateral: F7, F3, F4, F8, T3, C3, C4, T4, T5, P3, P4, T6; 3 midline: Fz, Cz, Pz), using a 200-ms prestimulus baseline. Individual ERPs were entered into separate grand ERP averages for each of the two agreement types. Time-windows were selected on the basis of previous research and visual inspection of the grand averages, as well as predicted effects: 350–600 ms for the LAN and N400, and 600–900 ms and 900–1,200 ms for the P600.

To examine the main aims of this study—namely, to elucidate the neurocognition of gender agreement in (a) adjective versus article agreement type, and (b) explicit versus implicit training conditions, at both low and high proficiency—we ran two sets of analyses. First, in order to reveal any differences or commonalities between the two agreement types, we directly compared the two types of agreement (adjective, article), separately for each group (implicit, explicit), first at low proficiency and then at end-of-practice. In each case, the global ANOVA contained the within-subject factors Agreement Type (noun-adjective agreement, noun-article agreement) and Violation (correct, violation), and the distributional factors Laterality (lateral, medial), Hemisphere (right, left), and Anterior/Posterior (anterior, central, posterior). When evaluating the Anterior/Posterior factor (which included more than one degree of freedom), the Greenhouse-Geisser correction was applied (corrected p-values are reported). Second, we directly compared the explicit and implicit training groups, separately for each agreement type (adjective, article), first at low proficiency and then at end-of-practice. In each case, we performed a global ANOVA with the between-subject factor Group (explicit, implicit) and the within-subject factors Violation as well as Laterality, Hemisphere, and Anterior/Posterior. In all cases, any global ANOVA that yielded a significant interaction ($p < .05$), including the factor Violation and any distributional factor, was followed up with step-down ANOVAs to clarify the nature of the interaction. Similar analyses were also carried out for the midline electrodes, except without the factors Laterality and Hemisphere. Results of the midline analysis are reported only when they revealed effects that were not evidenced in the lateral analyses.

Results

Behavioral Results

The d-prime scores are displayed in Figure 3 for the explicit and implicit groups at low proficiency and end-of-practice. The repeated measures ANOVA analysis on the judgment data acquired during the ERP task as reflected by d-prime scores revealed a main effect of Level, $F(1, 28) = 57.70$, $p < .001$, $\eta^2 = .673$, indicating the fact that, overall, learners performed better at end-of-practice than at low proficiency. This main effect, however, was qualified by a significant Level × Group interaction, $F(1, 28) = 8.01$, $p = .009$, $\eta^2 = .222$, indicating that the main effect of Level differed by Group. Follow-up analyses by Group showed that both explicitly and implicitly trained learners made significant gains from low proficiency to end-of-practice, $F(1, 15) = 14.05$, $p = .002$, $\eta^2 = .484$ and $F(1, 13) = 44.13$, $p < .001$, $\eta^2 = .772$, respectively, with the implicit group making larger gains. The results from the two-way Level × Group interaction were, in turn, qualified by a significant Level × Group × Agreement Type interaction, $F(1, 28) = 7.78$, $p = .009$, $\eta^2 = .217$, which reflected the fact that the Level × Group interaction differed by Agreement Type (see Figure 3). Follow-up analyses by Agreement Type indicated that the Level × Group interaction reached significance only for noun-adjective agreement, $F(1, 28) = 17.12$, $p < .001$, $\eta^2 = .379$. Comparisons of main effects by Group for noun-adjective agreement revealed that only the implicit group showed a significant effect of Level, $F(1, 13) = 43.97$, $p < .001$, $\eta^2 = .772$. Thus, when discriminating between correct and violation sentences, both the explicit and implicit groups evidenced significant gains between low proficiency and end-of-practice for noun-article agreement, but only the implicit group showed a significant gain for noun-adjective agreement. Note that there were no significant group differences, at either low proficiency

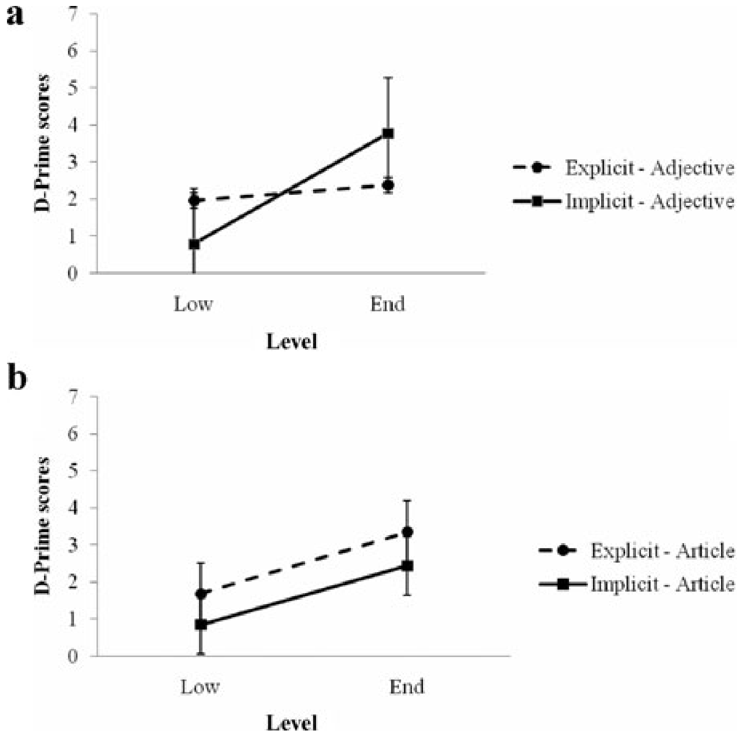

Figure 3 D-Prime scores reflecting the explicit and implicit groups' ability to judge violation and control sentences, represented separately for noun-adjective agreement (**a**) and noun-article agreement (**b**). Note that a complete inability to discriminate would yield a d-prime score of zero and perfect discrimination ability would yield a d-prime score of 6.93, with the effective limit being 4.65 (Macmillan & Creelman, 2005). Error bars represent standard error.

or at end-of-practice (or over both these levels), for either noun-adjective or noun-article agreement.

ERP Results

Grand average ERP waveforms for noun-adjective and noun-article agreement processing by the implicit and explicit groups at low proficiency and at end-of-practice are displayed in Figure 4.

Implicit Group: Analyses by Agreement Type

For the implicit group at low proficiency, the global ANOVA (with Agreement Type as a factor) and follow-up analyses revealed two effects, both of which

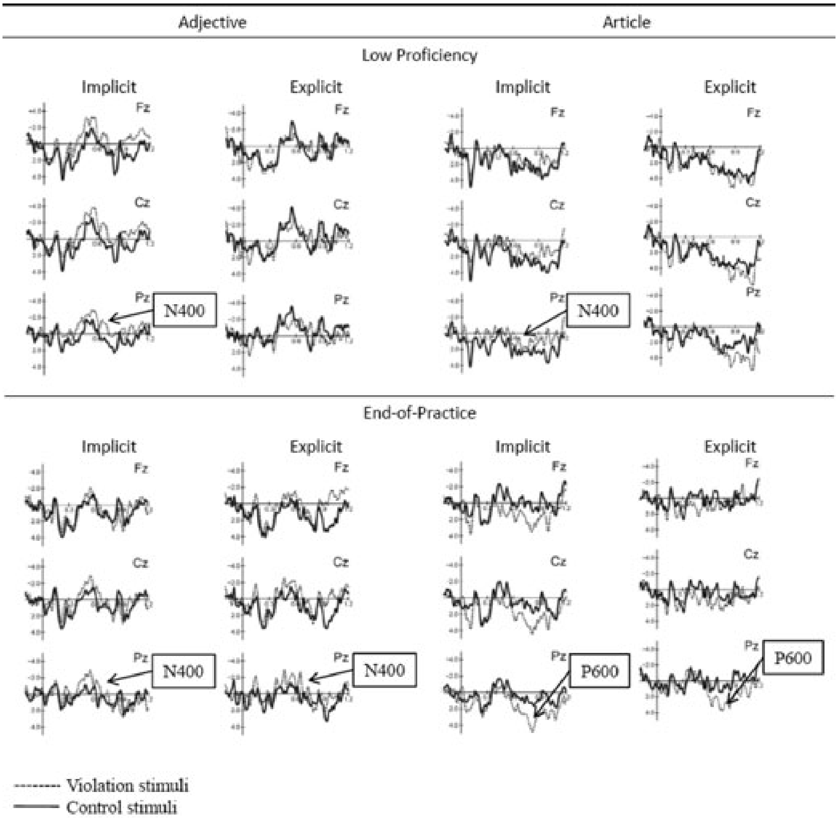

Figure 4. Grand average ERP waveforms showing all classical ERP language components, by agreement type at low proficiency and end-of-practice by group. Frontal (Fz), central (Cz), and posterior (Pz) electrodes, which are located along the midline, are displayed as representative electrodes.

were common to the two agreement types: an N400 (350–600 ms) as well as a later negativity (600–900 ms and 900–1,200 ms) (see Table 2). The global ANOVA yielded no interactions with the factor Agreement Type in any of these time windows, suggesting no reliable ERP differences between processing noun-adjective and noun-article agreement violations at low proficiency for the implicit group. At end-of-practice, the analysis for the implicit group revealed a positivity in the 600- to 900-ms time-window across both agreement types, although an Agreement Type × Violation interaction and follow-up analyses showed that the positivity, consistent with a P600, was reliable only for violations of noun-article agreement.

Table 2 Results of the global ANOVA for the implicit group with Agreement Type as the comparative factor

	Time-windows	
350–600 ms	600–900 ms	900–1,200 ms
	Low proficiency: Effects across agreement type	
V (13.65): N**	V (7.81): N*	V (19.42): N***
V × L (15.13)**	V × L (17.10)**	V × L (13.76)**
Medial	*Medial*	*Medial*
V (18.02): N**	*V (12.37): N***	*V (27.54): N****
Lateral		*Lateral*
V (6.90): N*		*V (10.09): N***
	Low proficiency: Agreement type differences	
	End-of-practice: Effects across agreement type	
Midline		
V × AP (5.76)*	V (7.44): P*	V × AP (5.60)*
Follow-ups		*Follow-ups*
not significant		*not significant*
	End-of-practice: Agreement type differences	
AGR × V × L (7.59)**	AGR × V (5.71)**	
Follow-ups	*Article*	
not significant	*V (22.64): P****	

Note. Results are reported for lateral repeated measures ANOVAs, except when the midline is indicated. *F* values are reported where $p < .05$. Step-down analyses are reported in *italics*. V = Violation; L = Laterality; H = Hemisphere; AP = Anterior/Posterior; AGR = Agreement Type; N = Negativity; P = Positivity.
$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

Explicit Group: Analyses by Agreement Type

The analyses for the explicit group at low proficiency revealed no effects either across or between the two agreement types (see Table 3). At end-of-practice, the global ANOVA yielded a significant Violation × Anterior/Posterior interaction across both agreement types in the 900- to 1,200-ms time-window. Step-down analyses revealed a significant negative effect over only anterior sites. However, a significant Agreement Type × Violation interaction and follow-up analyses in the same time-window suggested that this late negativity was reliable only for violations of noun-adjective agreement.

Table 3 Results of the global ANOVA for the explicit group with Agreement Type as the comparative factor

Time-windows		
350–600 ms	600–900 ms	900–1,200 ms
Low proficiency: Effects across agreement types		
Low proficiency: Agreement type differences		
AGR × V × AP × H (4.43)*		
Follow-ups		
not significant		
End-of-practice: Effects across agreement types		
		V × AP (4.83)*
		Anterior
		*V (6.56): N**
End-of-practice: Agreement type differences		
	AGR × V × L (9.11)**	AGR × V (7.66)*
	Follow-ups	*Adjective*
	not significant	*V (13.47): N***
		AGR × V × L (5.15)*
		Follow-ups
		not significant

Note. Results are reported for lateral repeated measures ANOVAs, except when the midline is indicated. F values are reported where $p < .05$. Step-down analyses are reported in *italics*. V = Violation; L = Laterality; H = Hemisphere; AP = Anterior/Posterior; AGR = Agreement Type; N = Negativity; P = Positivity.
$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

Noun-Adjective Agreement: Analyses by Group

The analyses for noun-adjective agreement at low proficiency revealed one effect common to both the implicit and explicit groups: a medial negativity in the 600- to 900-ms time-window (see Table 4). The nature of this negativity is unclear, although its medial distribution and onset are consistent with a late N400; see just below for analysis and further detail. The global ANOVA yielded Group × Violation and Group × Violation × Laterality interactions in the 350- to 600-ms time-window. Follow-up analyses (Table 4) revealed an N400 for the implicit group only. However, the lack of a higher amplitude negativity for violation compared to correct adjectives in the explicit group was not due to an absence of N400s for the two types of forms but rather to N400s of equivalent,

Table 4 Results of the global ANOVA for noun-adjective agreement with Group as the comparative factor

Time-windows		
350–600 ms	600–900 ms	900–1,200 ms
	Low proficiency: Effects across groups	
V × L × H (5.08)* *Follow-ups* *not significant*	V (4.67): N* V × L (5.01)* *Medial* *V (5.96): N**	
	Low proficiency: Group differences	
G × V (5.26)* *Implicit* *V (7.51): N** G × V × L (5.08)* *Implicit* *Lateral* *V (5.55): N** *Medial* *V (8.72): N**		G × V × AP × H (3.82)* *Follow-ups* *not significant*
	End-of-practice: Effects across groups	
V × L (7.39)* *Medial* *V (4.60): N** Midline V (4.85): N*	V × L × AP (4.49)* *Follow-ups* *not significant*	V (4.94): N* V×AP (6.53)* *Anterior* *V (8.37): N*** *Central* *V (5.09): N**
	End-of-practice: Group differences	
	G × V × H (4.78)* *Follow-ups* *not significant*	G × V (4.80)* *Explicit* *V (13.47): N***

Note. Results are reported for lateral repeated measures ANOVAs, except when the midline is indicated. *F* values are reported where $p < .05$. Step-down analyses are reported in *italics*. V = Violation; L = Laterality; H = Hemisphere; AP = Anterior/Posterior; G = Group; N = Negativity; P = Positivity.
*$p < .05$. **$p < .01$. ***$p < .001$.

$F(1, 15) = 0.60$, $p = .450$, and rather elevated amplitudes in both conditions (see Figure 4). Together, the data suggest that the negativity across the implicit and explicit groups in the 600- to 900-ms time-window indeed reflects an N400, due to an ongoing N400 in the implicit group and, in the explicit group, to the divergence of waveforms between the violation and control conditions from N400s of equivalent amplitudes for these two conditions in the earlier time-window. At end-of-practice, the analyses revealed an N400 effect (350- to 600-ms time-window) common to both groups, as well as a late anterior negativity (900–1,200 ms) common to both groups, although follow-up analyses on a Group × Violation interaction suggested that the negativity was reliable for the explicit group only.

Noun-Article Agreement: Analyses by Group

For noun-article agreement at low proficiency, the analyses revealed no effects either across or between the implicit and explicit groups (see Table 5). At end-of-practice, the analyses revealed a P600 effect common to both groups, beginning in the 350- to 600-ms time-window and continuing through the 900- to 1,200-ms time-window, with the positivity more anterior in the 350- to 600-ms time-window and more posterior in the 900- to 1,200-ms time-window. The analyses indicated no differences between the two groups.

Discussion

This study aimed to provide insights into the neurocognitive underpinnings of L2 gender agreement processing. In particular, it compared and contrasted the processing of noun-adjective and noun-article gender agreement anomalies, at both low and high experience/proficiency. In addition, it examined how explicit and implicit training conditions might affect the processing of such gender agreement anomalies.

To summarize the behavioral results, the analysis of subjects' ability to discriminate between correct sentences and sentences containing either a noun-adjective or a noun-article agreement violation showed a main effect of Level (better performance at end-of-practice than at low proficiency, over both agreement types and both training conditions) as well as two significant interactions: a Group × Level interaction and a Group × Level × Agreement Type interaction. These interactions reflect the fact that although both the explicit and implicit groups evidenced significant gains from low proficiency to end-of-practice for noun-article agreement, only the implicit group showed a significant gain for noun-adjective agreement (see Figure 3).

Table 5 Results of the global ANOVA for noun-article agreement with Group as the comparative factor

Time-windows		
350–600 ms	600–900 ms	900–1,200 ms
Low proficiency: Effects across groups		
Low proficiency: Group differences		
		G × V × L × AP (4.42)* *Follow-ups not significant*
End-of-practice: Effects across groups		
Midline V (4.91): P* V × AP (4.86)* *Anterior* *V (7.22): P**	V (9.43): P** V × L (16.26)*** *Medial* *V (15.43): P****	V × L (9.54)** *Medial* *V (4.83): P** V × AP (4.63)* *Posterior* *V (7.51): P**
End-of-practice: Group differences		

Note. Results are reported for lateral repeated measures ANOVAs, except when the midline is indicated. F values are reported where $p < .05$. Step-down analyses are reported in *italics*. V = Violation; L = Laterality; H = Hemisphere; AP = Anterior/Posterior; G = Group; N = Negativity; P = Positivity.
$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

The ERP findings can be summarized as follows. At low proficiency, both adjective and article gender agreement violations yielded N400s (350- to 600-ms time-window) but only for the implicitly trained group. Additionally, in the subsequent time-window (600–900 ms) adjective violations in both the explicitly and implicitly trained groups elicited an apparent N400, which likely represented a continuation of the N400 in the implicit group, and the divergence of waveforms between the violation and correct conditions from equivalent N400s for these two conditions in the earlier time-window. At high proficiency, noun-adjective agreement violations elicited N400s for both the explicit and implicit groups, whereas noun-article agreement violations elicited P600s for both groups.

These results can be contrasted with the pattern observed in previous L1 ERP studies of nominal gender agreement violations in sentence contexts (see the Review of Literature section). In all such L1 studies, article-noun violations have elicited P600s, often preceded by LANs. The only studies examining attributive adjectives in adjective-noun violations (Davidson & Indefrey, 2009; Sabourin & Haverkort, 2003) found a P600, with no preceding LAN. None of the studies examining either article-noun or adjective-noun violations in nonfinal positions in sentence contexts has reported N400s. Note that no L1 study has examined effects of explicit versus implicit training.

The results from the current study suggest the following. At low proficiency, neither article nor adjective gender agreement processing, in either implicitly or explicitly trained L2 learners, depend on neurocognitive processes that are implemented in L1 gender agreement processing, as evidenced by an absence of P600s at low proficiency. Instead, as evidenced by N400s, implicitly trained learners at low proficiency rely on lexical/semantic processes—and likely on declarative memory—for both adjective and article gender agreement processing. Explicitly trained learners also appear to rely on lexical/semantic processes at low proficiency, but only for adjective (not article) agreement, and the increase in lexical/semantic processing expected for violation versus correct conditions is delayed in comparison to the implicitly trained group. In contrast, gender agreement processing at end-of-training does not depend on the type of training: For both training groups, noun-article agreement processing depends, at least to some extent, on L1 processing mechanisms, as evidenced by P600s, whereas noun-adjective agreement processing depends on lexical/semantic processes, as evidenced by N400s.

The findings appear to be largely consistent with results from previous L2 ERP studies of gender agreement and other grammatical structures, although some differences and new patterns can be seen. First, the P600s found in response to noun-article gender agreement violation at end-of-practice jibe with previous ERP studies of gender agreement. These studies reported P600 effects for article-noun gender agreement in L2 learners who had at least a few months of exposure to the target language and whose L1 either has no gender agreement (Tokowicz & MacWhinney, 2005) or has a similar gender agreement structure to the L2 (Sabourin & Haverkort, 2003; Sabourin & Stowe, 2008). Similarly, in the current study, noun-article agreement violations elicited P600 effects in learners who did not have an L1 grammatical gender agreement system and who reached higher levels of proficiency after more than minimal exposure (training and practice) to an artificial language. Thus, our findings strengthen the already-observed pattern (Tokowicz & MacWhinney, 2005) that learners whose L1 does not have gender agreement and who are provided with some reasonable amount of L2 exposure show L1-like aspects of gender agreement processing between articles and nouns. In addition, the data extend this pattern from contexts in which the article precedes the noun, which contains the violation, to contexts in which the article contains the violation and follows the noun. The lack of P600s among low proficiency learners and for noun-adjective violations is also consistent with previous findings, specifically both with the observation that grammatical violations tend to elicit P600s less reliably in lower than higher proficiency L2 learners (Steinhauer et al., 2009) and with the finding that adjective-noun gender agreement violations have not elicited P600s (Davidson & Indefrey, 2009; Sabourin & Haverkort, 2003). As a final note regarding P600s, the current study extends previous observations by suggesting that the presence or absence of P600s in response to gender agreement violations seems to occur independently of the type of training—specifically, explicit versus implicit training. Second, the absence of LANs in this study is consistent with an absence of LANs in previous studies of nominal gender agreement—although this pattern seems to differ from findings of LANs in response to other (morpho)syntactic violations in high proficiency L2 learners of both artificial and natural languages (Bowden et al., 2010; Friederici et al., 2002; Ojima et al., 2005; Rossi et al., 2006). Third, N400s, which were observed here, have not been reported in other L2 studies of nominal gender agreement violations. Moreover, although clear N400s have been found in response to other grammatical violations in L2 (Osterhout et al., 2006), this is the first report, to our knowledge, of N400s elicited by grammatical violations in higher as well as lower proficiency L2 learners.

The pattern of classical language ERP effects observed here—specifically, the P600s, the absence of LANs, and the N400s—may be explained by a number of factors. First, we discuss the pattern of P600s. Their presence at higher but not lower proficiency and exposure—and similar patterns of P600 elicitation found in other L2 studies of both gender agreement and other grammatical structures (see the Review of Literature section)—is consistent with views suggesting that higher levels of L2 proficiency and/or experience are associated with a greater involvement of L1 grammatical mechanisms (Paradis, 2004; Steinhauer et al., 2009; Ullman, 2001, 2005). The elicitation of P600s by noun-article but not noun-adjective violations, both here and in previous studies, could be partly explained by the fact that, at least in L1, articles are widely posited to play more important grammatical roles than adjectives and other content words (Radford, Atkinson, Britain, Clahsen, & Spencer, 1999). Thus, articles would be more likely than adjectives to elicit P600s—in particular, at higher levels of proficiency, at which there may be an increased dependence on L1 grammatical mechanisms.

Second, as discussed earlier, although LANs have not been found by previous studies of L2 nominal gender agreement, they have been found in response to other grammatical violations in high proficiency L2 learners, although less reliably than P600s. The absence of LANs in this and other studies of grammatical gender could be at least partly explained by insufficiently high levels of proficiency, which may be necessary to achieve a dependence on the L1 mechanisms underlying LANs (Ullman, 2001, 2005). Indeed, in the present study, performance when discriminating between correct sentences and sentences containing noun-article violations, which elicited P600s but not LANs, did not closely approach the effective limit for d-prime scores (i.e., 4.65; Macmillan & Creelman, 2005) (see Figure 3), suggesting that learners had not reached the upper bounds of discrimination. More generally, the particular difficulty of gender agreement for L2 learners (see the Introduction) may explain the lack of LANs in response to gender agreement violations, in this and other studies. It is also possible that the absence of LANs accompanying the P600s on the noun-article violations may be partly explained by the view that aspects of morphosyntax—in particular, when non-local dependencies are involved—may be relatively unlikely to be proceduralized by L2 learners and, thus, less likely to elicit LANs (Clahsen & Felser, 2006a, 2006b; Ullman, 2001, 2005): In the current study, noun-article pairs were presented half of the time adjacently and half of the time nonadjacently (i.e., nonlocally), both during training and during ERP testing (see Note 2).

Third, the pattern of N400s may be explained by several factors. N400 effects were found for noun-adjective gender agreement violations at end-of-practice over both training groups, as well as for violations of both adjective and article agreement at low proficiency in the implicit group. Additionally, the data suggest a late N400 for violations of adjective agreement in the explicit group at low proficiency.

The N400 pattern observed here may be partly explained by the claim made by the DP model that aspects of grammatical processing should tend to depend on declarative memory at lower levels of experience and proficiency (Ullman, 2001, 2005). This might help account for the elicitation of N400s in the implicit group for noun-article violations at low proficiency but not at end-of-practice, as the former yielded lower performance. However, this explanation does not appear to capture the full set of results, given the lack of any N400 in the explicit group for noun-article violations at low proficiency. One possibility is that variability in explicit cognitive strategies and/or the timing of any ERP components in the explicit group at low proficiency could wash out any clear components in the waveforms and lead to a lack of reliable statistical differences. This could explain the lack of N400s—as well as any other reliable effects[5]—on articles, as well as the absence of typical N400 effects for adjectives, in the explicit group at low proficiency.

However, this still does not appear to be the whole story for adjectives in the explicit group at low proficiency. The data suggest that N400s are present at equivalent amplitudes for both violation and correct adjectives in a typical N400 time period (350–600 ms) but diverge, yielding a violation N400 effect, somewhat later (600–900 ms). This, in turn, suggests that the differential lexical/semantic processing of violation and correct adjectives occurs quite late in the explicitly trained group at low proficiency compared to the implicitly trained group. One possible explanation for this pattern is that at low proficiency the explicit group processes violation and correct adjectives as similarly unexpected in the typical N400 period, and only in a subsequent period do they process a difference in expectancy between the forms. Indeed, this would be consistent with evidence suggesting that explicit knowledge tends to be processed more slowly than implicit knowledge (Domangue, Mathews, Sun, Roussel, & Guidry, 2004).

The pattern of N400s in this study might also be partly explained by basic differences between adjectives and articles. In particular, the finding that N400s were elicited more reliably for noun-adjective agreement processing (for which typical or late N400s were found in all four conditions; i.e., in both the implicit and explicit groups at both low proficiency and end-of-practice) than for noun-article agreement processing (for which N400s were found only in one condition: at low proficiency for the implicit group) may be partly explained by the fact that adjectives and other content words have richer semantic content than articles and, thus, may be easier for declarative memory to process, resulting in more reliable N400s for noun-adjective than noun-article violations.

Finally, the finding of more reliable N400s on noun-adjective than noun-article violations is also consistent with the view that local dependencies should be easier to learn (e.g., as chunks) than nonlocal dependencies in lexical/declarative memory (Ullman, 2001, 2005), because noun-adjective pairs (unlike noun-article pairs) were always presented adjacently in this study (both in training and in practice). Note that although local dependencies are also posited to be more easily proceduralized than nonlocal dependencies, the absence of LANs here is not surprising given that noun-adjective violations did not even yield P600s. Thus, although the adjacent presentation of noun-adjective pairs could have facilitated chunking in declarative memory, other factors, such as lower performance or rich semantic context, may have precluded a shift of dependence from N400s to L1-like grammatical processes.

Several aspects of this study warrant further discussion. First, we consider the results in light of the field of SLA. The finding that implicitly trained learners performed as well (at both low proficiency and end-of-practice) as explicitly trained learners does not appear to be consistent with the majority of SLA research on this issue, which has generally reported advantages for L2 development under explicit conditions (see Norris & Ortega, 2000). Rather, the behavioral results from the current study add to the growing body of evidence that implicitly trained learners may show the same level of L2 development as explicitly trained learners (De Jong, 2005; Rosa & O'Neill, 1999; Sanz & Morgan-Short, 2004, 2005; VanPatten & Oikkenon, 1996). These studies suggest that equivalent performance between explicitly and implicitly trained groups may be found particularly when the training or practice requires the learner to process the form in order to accomplish a task (Sanz & Morgan-Short, 2005), as did the game-based comprehension and production practice completed by our subjects. Additionally, the present finding that explicit and implicit types of training do not appear to have differential effects on the neurocognitive processes underlying L2 gender agreement at higher proficiency, even if they might at lower proficiency, may have interesting implications for SLA research. Furthermore, the results from the current study make a new contribution to this line of research, in that they suggest a learning advantage for the implicitly trained learners: Recall that although the implicit group showed significant gains on noun-adjective and noun-article agreement, the explicit group evidenced significant gains *only* for noun-article agreement.

As discussed earlier, SLA research has generally found that explicitly trained groups outperform implicitly trained groups. However, as pointed out by Norris and Ortega (2000) and by Sanz and Morgan-Short (2005), experimental treatments in SLA research are often relatively short and, thus, might favor learning under explicit conditions, as learning under implicit conditions may take longer (Ellis, 1994). Moreover, assessments of language development often tap into more explicit types of knowledge (Ellis et al., 2009; Norris & Ortega, 2000). The current study attempted to address these methodological concerns through a longitudinal within-subject design that examined both online (ERP) and offline (judgment) measurements. The implications of the results reported here extend beyond theoretical issues to methodological issues and suggest that results from study designs that are not (as) biased toward explicit conditions might not support advantages for explicitly trained groups.

Second, the current study can help distinguish between competing neurocognitive models of L2. The data presented here are consistent with the view that higher levels of L2 experience and proficiency lead to a greater dependence of L2 grammatical processing on L1 neurocognitive mechanisms (Clahsen & Felser, 2006a, 2006b; Paradis, 2004; Steinhauer et al., 2009; Ullman, 2001, 2005). The results are also consistent with the view that at lower experience and proficiency, L2 grammatical processing may rely on lexical/semantic processing mechanisms and declarative memory, as hypothesized by the DP model (Ullman, 2001, 2005). In contrast, it is not clear how the data reported here could be accounted by the view that grammatical processing relies on the same set of neurocognitive mechanisms in the L2 as in the L1, regardless of L2 proficiency level (Abutalebi, 2008; Hernandez et al., 2005; Indefrey, 2006; MacWhinney, 2002, 2005). In particular, whereas N400s were elicited by a number of conditions in the current study, N400s have not been reported for L1 in response to nominal gender agreement violations. Moreover, the same condition (article violations in the implicit group) yielded an N400 at low proficiency and a P600 at end-of-practice, a pattern that does not seem consistent with a single neurocognitive mechanism across both the L2 and L1, regardless of L2 proficiency.

Third, in addition to the classical language ERP components discussed earlier (N400 and P600 effects), late negativities were observed. First, a negativity common to adjectives and articles was found in the implicit group at low proficiency in both the 600- to 900-ms and 900- to 1,200-ms time-windows (see the Results section and Table 2). Given its medial distribution in both of these time-windows, it is not unlikely that this reflects a continuation of the N400 effects that was observed in these conditions in the 350- to 600-ms time window. Such an elongated N400 is consistent with previous findings that L2 speakers often show long-lasting ERP effects (Hahne & Friederici, 2001; Hahne, Mueller, & Clahsen, 2006; Osterhout et al., 2006; Osterhout et al., 2008). Second, a late negativity was elicited by the explicit group at end-of-practice in response to noun-adjective violations (Table 4). This negativity, which was found in the 900- to 1200-ms but not the 600- to 900-ms time-window and was anteriorly distributed, does not appear to reflect a continuation of the N400 in the 350- to 600-ms time-window. Rather, it appears to be an independent effect. Late anterior negativities have been found previously for both L1 and L2 grammatical processing (e.g., Mueller et al., 2005; Sabourin & Haverkort, 2003; Sabourin & Stowe, 2008) but are not yet well understood, although it has been suggested that they may be elicited by increased involvement of working memory (Mueller et al., 2005; Sabourin & Stowe, 2008). It is beyond the scope of this article to interpret this negativity, which may be further elucidated by studies specifically designed to examine these effects.

In conclusion, this study, which further validates the use of both an artificial language and a combined behavioral/ERP approach to examine the acquisition and processing of L2, elucidated various aspects of the behavioral and neural underpinnings of noun-article and noun-adjective gender agreement processing at both lower and higher L2 experience/proficiency levels, under both explicit (classroomlike) and implicit (immersionlike) training conditions.

Revised version accepted 15 September 2009

Notes

1 It is important to emphasize that this study examined neurocognitive outcomes of explicit and implicit *training* conditions, *not* whether the resulting learning or knowledge may have been (partly or wholly) explicit or implicit.

2 This tightly controlled comparison of the explicit and implicit training conditions may have come at the expense of other factors, such as ecological validity. For example, it seems unlikely that feedback is identical in all classroom and immersion training contexts, which the explicit and implicit conditions are respectively designed to approximate. On the other hand, neither the classroom training context nor the immersion training context is entirely homogeneous, and feedback does in fact occur in both contexts, even in immersion, for example, with repetition and clarification requests (Lyster & Ranta, 1997; Lyster & Mori, 2006).

3 Although this factor (i.e., adjacent vs. nonadjacent noun-article pairs) was controlled for in the experimental design, it was not examined as an independent variable, both because it had not been previously coded for analysis and because of a lack of power.

4 As with adjacent and nonadjacent noun-article pairs (see Note 2), this factor (i.e., novel vs. repeated sentences) was not examined as an independent variable, both because it had not been previously coded for analysis and because of a lack of power.

5 Although visual inspection of the waveforms (Figure 4) suggested a possible P600 in the 900- to 1,200-ms time-window for article agreement violations in the explicit group at low proficiency, statistical analyses indicated that this was not a reliable effect. First, the analyses discussed in the main text (see Results) did not reveal a reliable effect. However, to increase our confidence, we examined the possibility that a weak P600 in this condition might lead to a reliable common P600 effect for article violations across low proficiency and end-of-practice in the explicit group. We performed an ANOVA that examined the influence of Level (low proficiency vs. end-of-practice), Violation, and distributional factors on the ERPs of articles in the explicit group in the 900- to 1,200-ms time-window. However, the ANOVA yielded no significant Violation effects common to the two levels. Moreover, it yielded one interaction that included both Level and Violation—namely, a Level × Laterality × Violation interaction, $F(1, 15) = 10.85$, $p = 0.005$, which in fact suggested a P600

at end-of-practice (Laterality × Violation, $F[1, 15] = 5.22, p = 0.037$) but not at low proficiency (Laterality × Violation, $F[1, 15] = 0.91, p = 0.355$). Thus, the apparent late positivity for article violations in the explicit group at low proficiency does not appear to be reliable.

References

Abutalebi, J. (2008). Neural aspects of second language representation and language control. *Acta Psychologica, 128*(3), 466–478.

Abutalebi, J., Cappa, S. F., & Perani, D. (2003). What can functional neuroimaging tell us about the bilingual brain? In A. M. B. De Groot & J. F. Kroll (Eds.), *Handbook of bilingualism: Psycholinguistic perspectives*. London: Oxford University Press.

Arteaga, D., Herschensohn, J., & Gess, R. (2003). Focusing on phonology to teach morphological form in French. *Modern Language Journal, 87*(1), 58–70.

Barber, H., & Carreiras, M. (2003). Integrating gender and number information in Spanish word pairs: An ERP study. *Cortex, 39*(3), 465–482.

Barber, H., & Carreiras, M. (2005). Grammatical gender and number agreement in Spanish: An ERP comparison. *Journal of Cognitive Neuroscience, 17*(1), 137–153.

Bartning, I. (2000). Gender agreement in L2 French: Pre-advanced vs. advanced learners. *Studia Linguistica, 54*(2), 225–237.

Benati, A. (2005). The effects of processing instruction, traditional instruction and meaning-output instruction on the acquisition of the English past simple tense. *Language Teaching Research, 9*(1), 67–93.

Bowden, H. W., Sanz, C., Steinhauer, K., & Ullman, M. T. (2010). *Effects of experience and proficiency on second-language neurocognition: An ERP study of Spanish*. Manuscript submitted for publication.

Chen, L., Shu, H., Liu, Y., Zhao, J., & Li, P. (2007). ERP signatures of subject-verb agreement in L2 learning. *Bilingualism: Language and Cognition, 10*(2), 161–174.

Clahsen, H., & Felser, C. (2006a). Grammatical processing in language learners. *Applied Psycholinguistics, 27*(1), 3–42.

Clahsen, H., & Felser, C. (2006b). How native-like is non-native language proceessing? *Trends in Cognitive Science, 10*(12), 564–570.

Davidson, D. J., & Indefrey, P. (2009). An event-related potential study on changes of violation and error responses during morphosyntactic learning. *Journal of Cognitive Neuroscience, 21*(3), 433–446.

De Jong, N. (2005). Can second language grammar be learned through listening? An experimental study. *Studies in Second Language Acquisition, 27*(2), 205–234.

DeKeyser, R. M. (1995). Learning second language grammar rules: An experiment with a miniature linguistic system. *Studies in Second Language Acquisition, 17*(3), 379–410.

DeKeyser, R. M. (2003). Implicit and explicit learning. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 313–348). Malden, MA: Blackwell.

Demestre, J., & Garcia-Albea, J. E. (2007). ERP evidence for the rapid assignment of an (appropriate) antecedent to PRO. *Cognitive Science, 31*(2), 343–354.

Demestre, J., Meltzer, S., Garcia-Albea, J. E., & Vigil, A. (1999). Identifying the null subject: Evidence from event-related brain potentials. *Journal of Psycholinguistic Research, 28*(3), 293–312.

Dewaele, J.-M., & Veronique, D. (2001). Gender assignment and gender agreement in advanced French interlanguage: A cross-sectional study. *Bilingualism: Language and Cognition, 4*(3), 275–297.

Domangue, T. J., Mathews, R. C., Sun, R., Roussel, L. G., & Guidry, C. E. (2004). Effects of model-based and memory-based processing on speed and accuracy of grammar string generation. *Journal of Experimental Psychology: Learning Memory and Cognition, 30*(5), 1002–1011.

Ellis, N. C. (Ed.). (1994). *Implicit and explicit learning of languages*. London: Academic Press.

Ellis, R., Loewen, S., Elder, C., Erlam, R., Philp, J., & Reinders, H. (2009). *Implicit and explicit knowledge in second language learning, testing, and teaching*. Bristol, UK: Multilingual Matters.

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

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences, 6*(2), 78–84.

Friederici, A. D., Hahne, A., & Mecklinger, A. (1996). The temporal structure of syntactic parsing: Early and late effects elicited by syntactic anomalies. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(5), 1219–1248.

Friederici, A. D., & Kotz, S. A. (2003). The brain basis of syntactic processes: Functional imaging and lesion studies. *NeuroImage, 20*(Suppl. 1), S8–S17.

Friederici, A. D., Kotz, S. A., Werheid, K., Hein, G., & von Cramon, D. Y. (2003). Syntactic comprehension in Parkinson's disease: Investigating early automatic and late integrational processes using event-related brain potentials. *Neuropsychology, 17*(1), 133–142.

Friederici, A. D., Pfeifer, E., & Hahne, A. (1993). Event-related brain potentials during natural speech processing: Effects of semantic, morphological and syntactic violations. *Cognitive Brain Research, 1*(3), 183–192.

Friederici, A. D., Steinhauer, K., & Frisch, S. (1999). Lexical integration: Sequential effects of syntactic and semantic information. *Memory and Cognition, 27*(3), 438–453.

Friederici, A. D., Steinhauer, K., & Pfeifer, E. (2002). Brain signatures of artificial language processing: Evidence challenging the critical period hypothesis. *Proceedings of the National Academy of Sciences of the United States of America, 99*(1), 529–534.

Friederici, A. D., von Cramon, D. Y., & Kotz, S. A. (1999). Language related brain potentials in patients with cortical and subcortical left hemisphere lesions. *Brain, 122*(Pt. 6), 1033–1047.

Gass, S. M., & Alvarez Torres, M. J. (2005). Attention when? An investigation of the ordering effect of input and interaction. *Studies in Second Language Acquisition, 27*, 1–31.

Goad, H., & White, L. (2006). Ultimate attainment in interlanguage grammars: A prosodic approach. *Second Language Research, 22*(3), 243–268.

Gunter, T. C., Friederici, A. D., & Schriefers, H. (2000). Syntactic gender and semantic expectancy: ERPs reveal early autonomy and late interaction. *Journal of Cognitive Neuroscience, 12*, 556–568.

Hagoort, P. (2003). Interplay between syntax and semantics during sentence comprehension: ERP effects of combining syntactic and semantic violations. *Journal of Cognitive Neuroscience, 15*(6), 883–899.

Hagoort, P., & Brown, C. M. (1999). Gender electrified: ERP evidence on the syntactic nature of gender processing. *Journal of Psycholinguist Research, 28*(6), 715–728.

Hahne, A., & Friederici, A. D. (1999). Electrophysiological evidence for two steps in syntactic analysis: Early automatic and late controlled processes. *Journal of Cognitive Neuroscience, 11*(2), 194–205.

Hahne, A., & Friederici, A. D. (2001). Processing a second language: Late learners' comprehension strategies as revealed by event-related brain potentials. *Bilingualism: Language and Cognition, 4*, 123–141.

Hahne, A., Mueller, J. L., & Clahsen, H. (2006). Morphological processing in a second language: Behavioral and event-related brain potential evidence for storage and decomposition. *Journal of Cognitive Neuroscience, 18*(1), 121–134.

Hawkins, R., & Chan, C. Y.-h. (1997). The partial availability of universal grammar in second language acquisition: The "failed functional features hypothesis." *Second Language Research, 13*(3), 187–226.

Hernandez, A., Li, P., & MacWhinney, B. (2005). The emergence of competing modules in bilingualism. *Trends in Cognitive Science, 9*(5), 220–225.

Hernandez, A. E., & Li, P. (2007). Age of acquisition: Its neural and computational mechanisms. *Psychological Bulletin, 133*(4), 638–650.

Indefrey, P. (2006). A meta-analysis of hemodynamic studies on first and second language processing: Which suggested differences can we trust and what do they mean? *Language Learning, 56*(Suppl. 1), 279–304.

Kaan, E. (2007). Event-related potentials and language processing: A brief overview. *Language and Linguistics Compass,* 571–591.

Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes, 15*(2), 159–201.

Keating, G. D. (2009). Sensitivity to violations of gender agreement in native and nonnative Spanish: An eye-movement investigation. *Language Learning*, *59*(3), 503–535.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207*(1), 203–205.

Kutas, M., & Schmitt, B. (2003). Language in microvolts. In M. T. Banich & M. Mack (Eds.), *Mind, brain, and language* (pp. 171–209). Mahwah, NJ: Erlbaum.

Luck, S. J. (2005). *An introduction to the event-related potential technique*. Cambridge, MA: MIT Press.

Lyster, R., & Mori, H. (2006). Interactional feedback and instructional counterbalance. *SSLA, 28*, 269–300.

Lyster, R., & Ranta, L. (1997). Corrective feedback and learner uptake: Negotiation of form in communicative classrooms. *Studies in Second Language Acquisition, 19*, 37–66.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah, NJ: Erlbaum.

MacWhinney, B. (2002). Extending the competition model. In R. R. Heredia & J. Altarriba (Eds.), *Bilingual sentence processing* (pp. 31–58). New York: North Holland/Elsevier.

MacWhinney, B. (2005). New directions in the competition model. In M. Tomasello & D. I. Slobin (Eds.), *Beyond nature-nurture: Essays in honor of Elizabeth Bates*. Mahwah, NJ: Erlbaum.

McCarthy, G., Nobre, A. C., Bentin, S., & Spencer, D. D. (1995). Language-related field potentials in the anterior-medial temporal lobe: I. Intracranial distribution and neural generators. *Journal of Neuroscience, 15*(2), 1080–1089.

Molinaro, N., Vespignani, F., & Job, R. (2008). Deeper reanalysis of a superficial feature: An ERP study on agreement violations. *Brain Research, 1228*, 161–176.

Montrul, S., Foote, R., Perpinan, S., Thornhill, D., & Vidal, S. (2008). Gender agreement in adult second language learners and spanish heritage speakers: The effects of age and context of acquisition. *Language Learning, 58*(3), 503–553.

Montrul, S. A. (2004). *The acquisition of Spanish: Morphosyntactic development in monolingual and bilingual L1 acquisition and adult L2 acquisition*. Amsterdam: Benjamins.

Mueller, J. L. (2005). Electrophysiological correlates of second language processing. *Second Language Research, 21*(2), 152–174.

Mueller, J. L., Hahne, A., Fujii, Y., & Friederici, A. D. (2005). Native and nonnative speakers' processing of a miniature version of Japanese as revealed by ERPs. *Journal of Cognitive Neuroscience, 17*(8), 1229–1244.

Neville, H. J., Nicol, J. L., Barss, A., Forster, K. I., & Garrett, M. F. (1991). Syntactically based sentence processing classes: Evidence from event-related brain potentials. *Journal of Cognitive Neuroscience, 3*(2), 151–165.

Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning, 50*(3), 417–528.

Norris, J. M., & Ortega, L. (2001). Does type of instruction make a difference? Substantive findings from a meta-analytic review. *Language Learning, 51*(1), 157–213.
Ojima, S., Nakata, H., & Kakigi, R. (2005). An ERP study on second language learning after childhood: Effects of proficiency. *Journal of Cognitive Neuroscience, 17*(8), 1212–1228.
Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9*(1), 97–113.
Opitz, B., & Friederici, A. D. (2003). Interactions of the hippocampal system and the prefrontal cortex in learning language-like rules. *NeuroImage, 19*(4), 1730–1737.
Opitz, B., & Friederici, A. D. (2004). Brain correlates of language learning: The neuronal dissociation of rule-based versus similarity-based learning. *Journal of Neuroscience, 24*(39), 8436–8440.
Osterhout, L., Bersick, M., & McLaughlin, J. (1997). Brain potentials reflect violations of gender stereotypes. *Memory and Cognition, 25*(3), 273–285.
Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language, 31*, 785–806.
Osterhout, L., McLaughlin, J., Pitkänen, I., Frenck-Mestre, C., & Molinaro, N. (2006). Novice learners, longitudinal designs, and event-related potentials: A means for exploring the neurocognition of second language processing. *Language Learning, 56*(Suppl. 1), 199–230.
Osterhout, L., Poliakov, A., Inoue, K., McLaughlin, J., Valentine, G., Pitkanen, I., et al. (2008). Second-language learning and changes in the brain. *Journal of Neurolinguistics, 21*(6), 509–521.
Paradis, M. (2004). *A neurolinguistic theory of bilingualism*. Amsterdam: Benjamins.
Radford, A., Atkinson, M., Britain, D., Clahsen, H., & Spencer, A. (1999). *Linguistics: An introduction*. Cambridge: Cambridge University Press.
Rosa, E. M., & O'Neill, M. D. (1999). Explicitness, intake, and the issue of awareness. *Studies in Second Language Acquisition, 21*(4), 511–556.
Rossi, S., Gugler, M. F., Friederici, A. D., & Hahne, A. (2006). The impact of proficiency on syntactic second-language processing of German and Italian: Evidence from event-related potentials. *Journal of Cognitive Neuroscience, 18*(12), 2030–2048.
Sabourin, L., & Haverkort, M. (2003). Neural substrates of representation and processing of a second language. In R. van Hout, A. Hulk, F. Kuiken, & R. Towell (Eds.), *The lexicon-syntax interface in second language acquisition* (pp. 175–195). Amsterdam: Benjamins.
Sabourin, L., & Stowe, L. A. (2008). Second language processing: When are first and second languages processed similarly? *Second Language Research, 24*(3), 397–430.
Sanders, L. D., Neville, H., & Woldorff, M. G. (2002). Speech segmentation by native and non-native speakers: The use of lexical, syntactic, and stress-pattern cues. *Journal of Speech, Language, and Hearing Research, 45*(3), 519–530.

Sanz, C., & Morgan-Short, K. (2004). Positive evidence vs. explicit rule presentation and explicit negative feedback: A computer-assisted study. *Language Learning, 54*(1), 35–78.
Sanz, C., & Morgan-Short, K. (2005). Explicitness in pedagogical interventions: Input, practice, and feedback. In C. Sanz (Ed.), *Mind and context in adult second language acquisition: Methods, theory, and practice* (pp. 234–263). Washington, DC: Georgetown University Press.
Simos, P. G., Basile, L. F. H., & Papanicolaou, A. C. (1997). Source localization of the N400 response in a sentence-reading paradigm using evoked magnetic fields and magnetic resonance imaging. *Brain Research, 762*(1–2), 29–39.
Steinhauer, K., & Connolly, J. F. (2008). Event-related potentials in the study of language. In B. Stemmer & H. A. Whitaker (Eds.), *Handbook of the neuroscience of language* (pp. 57–67). Oxford: Elsevier.
Steinhauer, K., White, E. J., & Drury, J. E. (2009). Temporal dynamics of late second language acquisition: Evidence from event-related brain potentials. *Second Language Research, 25*(1), 13–41.
Tokowicz, N., & MacWhinney, B. (2005). Implicit and explicit measures of sensitivity to violations in second language grammar: An event-related potential investigation. *Studies in Second Language Acquisition, 27*(2), 173–204.
Ullman, M. T. (2001). The neural basis of lexicon and grammar in first and second language: The declarative/procedural model. *Bilingualism: Language and Cognition, 4*(1), 105–122.
Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition, 92*(1–2), 231–270.
Ullman, M. T. (2005). A cognitive neuroscience perspective on second language acquisition: The declarative/procedural model. In C. Sanz (Ed.), *Mind and context in adult second language acquisition: Methods, theory and practice* (pp. 141–178). Washington, DC: Georgetown University Press.
Ullman, M. T. (2006). Is Broca's area part of a basal ganglia thalamocortical circuit? *Cortex, 42*(4), 480–485.
VanPatten, B., & Oikkenon, S. (1996). Explanation versus structured input in processing instruction. *Studies in Second Language Acquisition, 18*(4), 495–510.
Weber-Fox, C. M., & Neville, H. J. (1996). Maturational constraints on functional specializations for language processing: ERP and behavioral evidence in bilingual speakers. *Journal of Cognitive Neuroscience, 8*(3), 231–256.
White, L. (2003). Fossilization in steady state L2 grammars: Persistent problems with inflectional morphology. *Bilingualism: Language and Cognition, 6*, 129–141.
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., Moreno, E. M., & Kutas, M. (2004). Anticipating words and their gender: An event-related brain potential study of semantic integration, gender

expectancy, and gender agreement in Spanish sentence reading. *Journal of Cognitive Neuroscience, 16*(7), 1272–1288.

Appendix A

Example section from the explicit language training condition, in which the use of articles is described. Note that the text below was actually presented aurally. During this aural presentation, corresponding game constellations of examples, which are represented in bold here, were presented visually on the computer screen.

<u>Articles</u>

- Articles. Articles are used with nouns and specify that you are referring to one object in particular. There is only one article in BROCANTO2, which has two forms: a masculine form and a feminine form.
  - **Li** – is the masculine form
  - **Lu –** is the feminine form
  - You should remember two points about articles in BROCANTO2. First, articles always come after nouns. Second, articles must 'agree' with the gender of the nouns. In other words, if a noun is masculine, the masculine form of the article must be used. Likewise, if a noun is feminine, the feminine form of the article must be used. Thus, whenever a noun is masculine, you will need to use the masculine article – *Li*, as in the following example:
    - **pleck li**

      In this example, *Pleck* is a masculine noun, so you must use the masculine form of the article: *li*.

    On the other hand, if a noun is feminine, you will need to use the feminine article – *Lu*, as in the following example:
    - **blom lu**

      In this example, *Blom* is a feminine noun, so you must use the feminine form of the article: *lu*.

    Here is another example. Think about whether the noun is masculine or feminine. Notice that the article is placed after the noun.
    - **vode lu**

Appendix B

Example section from the implicit language training condition, in which examples of article use are presented. Note that the BROCANTO2 phrases were aurally presented together with visually presented corresponding game constellations.

pleck li
pleck troise li
pleck li
pleck neime li
neep li
neep neime li
neep li
neep troise li
pleck troise li
neep troise li
neep li
pleck li
neep neime li
pleck neime li
blom lu
blom troiso lu
blom lu
blom neimo lu
vode lu
vode neimo lu
vode troiso lu
vode lu
blom troiso lu
vode troiso lu
vode lu
blom lu
vode neimo lu
blom neimo lu