Learning a Gendered Language:

L2 Acquisition and Relative Processing Costs of Spanish Grammatical Gender and Number

Alexander M. Hirsch

Yale University

Submitted in partial fulfillment for the degree of Bachelor of Arts in Linguistics

Abstract

One of the central debates in the field of language acquisition concerns the possibility of nativelike language acquisition begun after the end of the so-called "critical period." Various hypotheses have been developed to explain the drop-off of naturalistic language acquisition skills after this period, as well as the specific deficiencies shown by many, but not all, late second language (L2) learners. To be able to say something useful about such a broad subject, this study's focus is narrowed to acquisition of certain grammatical features absent from the subjects' native language but present in the L2. Using a self-paced reading task, this study measures the ability of English-L1 late Spanish learners to process Spanish gender and number mismatches across determiner-noun pairs. This study was carried out in order to determine these speakers' ability to acquire linguistic elements not present in their native language, to provide insight into the architecture underlying processing of these two similar but distinct elements, and to determine how transfer affects this ability.

Acknowledgements

Learning a Gendered Language

L2 Acquisition and Relative Processing Costs of Spanish Grammatical Gender and Number

The below research seeks to answer the question of whether second language learners can acquire aspects of language not present in their native language to the same extent as native speakers. Specifically, this study examines whether native English speakers can attain the ability to process gender mismatches, not present in English, to the same extent as natives, and how this processing ability compares to their ability to process number mismatches, which are present to a certain extent in English. I will begin this paper by providing a brief introduction to the field of child language acquisition in order to provide context for a discussion of second language acquisition research, much of which is based primarily on study of child language acquisition. From here I will discuss past theories of second language acquisition (SLA): I will give a brief picture of what I see as the landmark hypotheses of the field, in order to provide some insight into the history of SLA investigation, as well as grounding in which to root a discussion of the field's current state. I will continue by discussing some current, competing hypotheses on the nature and possible extent of L2 acquisition and their relative merits. This section will serve to narrow the overbroad field of SLA to the quadrants to which the current research belongs. I will then move on to a discussion of the previous research on gender and number acquisition, especially in Spanish, in order to provide a background upon which to place the present study. Finally, I will present my experiment and hypothesis, where they stand with reference to the present state of SLA inquiry, and why the current experiment is important to further understanding of SLA on a broad scale.

1. Early Exploration into SLA

   1A. The Critical Period Hypothesis The existence of a "critical period" of language learning has been the subject of an ongoing debate since it was proposed in a landmark paper by Penfield & Roberts (1959). The so-called "Critical Period Hypothesis" (CPH) states that after a certain discrete period, usually stated as lasting from birth until puberty or later adolescence, it is not possible for a child without a first language (L1) to learn to speak any language with the degree of fluency and command exhibited by a native speaker. Though not specifically intended by the original proposers of the hypothesis, the CPH has been extended to second language learning, with similar implications. That is, adherents to the CPH with regard to second language (L2) acquisition contend that late L2 learners (i.e. second language learning is begun after the Critical Period) can never reach the fluency of a native speaker, or will never be able to totally eliminate their native accents.

   Experimentally, late L2 learners have been shown to perform worse than natives or early L2 learners on a variety of linguistic tasks. For example, in a study by Johnson & Newport (1989), subjects performed more poorly on a grammaticality judgment task if they began learning their L2 (in this case, English) after the age of 8 than younger children, who performed at ceiling levels. Furthermore, speakers who began learning an L2 when they were 12-14 years of age performed worse than younger children, and roughly on par with older speakers. Interestingly, heritage speakers who learn a language at a young age but lose the language after infancy are generally more capable than L2 learners (Alarcon 2011). These results suggest that this critical period ends definitively between the ages of 12 and 14 after the requisite language acquisition skills have begun to taper off beginning around age 8, and they support the CPH in the sense that after the Critical Period, a second

language becomes much more difficult to acquire. However, these results make no predictions as to the possibility (versus likelihood) of end-state or nativelike L2 acquisition when begun after the critical period.

1A. Beyond the critical period: Failed Functional Features. Many refinements to the CPH, as well as many alternatives, have been proposed in the last 60-plus years. With respect to L2 acquisition, the majority of subsequent hypotheses push in one of two directions; they either refute the claim that a discrete Critical Period blocking late language development exists, or they affirm Penfield & Roberts's hypothesis and explain the precise effects of a Critical Period on the language faculty that would lead to such a loss of learning potential.[1]

The Failed Functional Features Hypothesis (FFFH), proposed by Hawkins & Chan (1997), refines the CPH with respect to its position, or lack thereof, on L2 acquisition. It posits that L1 acquisition entails the fixing of certain "functional features," which underlie various aspects of the L1. From this hypothesis, two predictions follow about the behavior of L2 learners. First, beginning late L2 learners are expected to directly overlay L2 morphology and lexicon onto L1 syntax. From this starting point, L2 learners will progressively refine the interlanguage syntax (the underdeveloped L2 before L2 completion) until it approximates the L2 syntax as closely as possible. However, Hawkins and Chan propose that certain functional features from the L1 will not be revised to fit the L2; they will exist instead as failed functional features that deviate from those of native speakers.

---

[1] As an early hypothesis of the age cutoff phenomenon, the CPH does not give a detailed explanation of what mental or neurological processes underlie the Critical Period.

Though attractive in certain respects, the FFFH has not held up under empirically-motivated scrutiny. Numerous examples have surfaced over the intervening years of late L2 learners whose syntax is essentially nativelike and does not contain flaws along the lines of those predicted by the FFFH. For example, White and Genesee (1996) find no difference in performance between late near-native and native speakers on tasks designed to test access to Universal Grammar, which is thought to be subject to Critical Period effects, and White, et al. (2004) finds that native English Spanish-L2 learners can acquire Spanish grammatical gender to the same extent as French natives, findings that directly contradict the FFFH. Luckily, other hypotheses explaining Critical Period-like effects have emerged subsequent to the FFFH.

1B. Transfer and the FTFA Hypothesis. The Full Transfer Full Access (FTFA) Hypothesis, proposed by Schwartz & Sprouse (1996), is one such hypothesis, both drawing on the CPH and FFFH and differing from these predecessors in important ways. The authors argue that adults acquiring a second language proceed with acquisition beginning with the principles gained from acquiring their first language, like the FFFH. Furthermore, all the principles from the L1 are transferred to the initial state of the L2. Indeed, central to the FTFA argument is the proposal that the end state of L1 acquisition and the beginning state of L2 acquisition are identical within a given learner. L2 development therefore must be a process of refining or changing the initial grammar imported from the L1, so that the resulting interlanguage grammar reflects some mixture of the end state of the L1 and the end state of the L2. FTFA posits, then, that learners of an L2 do not begin anew with a blank slate but rather carry any grammatical principles from the L1 into their initial conception of the L2. Moreover, these L2 learners can and do use knowledge of their L1 grammar to

decipher and produce the L2. Because the L2 is acquired by modifying the transferred grammar of the L1 until it becomes the final grammar of the L2, the L2 learner is never without a grammar but rather possesses a different (but not degenerate) grammar from the target syntax. As a result, errors in morphological inflection, for example, are ascribed not to deficits in the syntax but rather to difficulty in the realization of that syntax so that it conforms to the target construction. Herschensohn, Stevenson & Waltmunson (2005) provide evidence for the FTFA in this respect, finding that L2 learners' syntax develops more rapidly and accurately than their morphology.

        In contrast with the CPH, the FTFA Hypothesis does not argue for a decrease in language-learning ability over time, and in some senses argues against such degeneration: if an adult begins to learn a second language with many similarities to their first language—a Spanish speaker learning Portuguese or an English speaker learning Dutch, for example—FTFA predicts that transfer from the first language should trivialize the amount of information needed to achieve fluency in the second language, as there will be substantial overlap between the L1 and the L2. Furthermore, learning novel structures will not be any more difficult during L2 than L1 acquisition, since these structures will simply be added atop previous ones. White, et al. (2004), support this account, finding that native English speakers learning Spanish are able to acquire gender to the same extent as native Spanish speakers. However, flaws in the study's methodology, such as inadequate categorization of subjects with respect to proficiency and lack of online testing, limit the applicability of the its conclusions.

        1C. How much is transferred? The extent of transfer between a first and second language is a subject of some contention. Evidence following Schwartz & Sprouse's work

suggests that correspondence between the L1 and L2 is the most important factor in L2 acquisition, implying that transfer is highly important to L2 learning (Bialystok 1997). However, previous hypotheses, such as Vainikka & Young-Scholten (1994)'s Minimal Trees hypothesis and Eubank (1993/1994)'s Weak Transfer Hypothesis do not agree with Schwartz & Sprouse's idea of full transfer: Vainikka & Young-Scholten suggest that only lexical categories and the linear orientation of these categories transfers from the L1 grammar, while Eubank proposes that functional categories (and their linear orientations) transfer as well. Importantly, neither Vainikka & Young-Scholten nor Eubank allow that the strength of inflection associated with functional categories transfers between L1 and L2, and it is here that these authors differ from Schwartz & Sprouse. This is to say that no matter the specific version of the weak transfer hypothesis in question, proponents of its various varieties unite in their view that the bare syntactic structure—e.g. functional projections and categories—is left intact in the grammatical transfer from the L1 to the nascent L2, but morphologically-driven features like agreement (gender, number, etc.) do not transfer to the initial L2 grammar.

      1D. Mechanisms underlying weak transfer effects. These weak transfer hypotheses necessarily reject a unified morphosyntactic conception of language learning, since they posit that only a certain parts of language—the syntax—may transfer, while other parts are left behind. At the same time, this kind of hypothesis implies that a specific dissociation must exist between syntax and morphology (at least as they are defined within such a framework; White, et al. 2004). Perhaps what is transferred are the structural aspects of language—linguistic components that are primarily organizational and connectional in nature. This would of course include syntax, and possibly also semantics,

which has been described as a system of signs and associated meanings (e.g. de Saussure 1916).

Rule-based components, on the other hand, would be left behind. According to Anderson (1992) and others, morphology is one example of a "rule-based component": it is a system of rules governing related words, not a system of roots and concatenated morphemes as some have characterized it. Anderson's system places morphology firmly in the rule-based category, diverging from previous theories that described it in a structural light. Perhaps a more obvious member of this class is phonology, a system of rules governing the sounds of related words.

At least intuitively, this kind of categorical division makes sense. Structure building or association, which combine existing parts to form a greater whole, should involve a fundamentally different process from rule application, which applies a rule or function to an input and transforms it into a new but related output. In terms of cognitive organization, there seems to be a general division in the brain between rule-based and structure-based aspects of language.

In fact, Michael Ullman made a similar distinction over a decade ago: Ullman (2001) separates "words," associative memories that are subserved by the declarative memory system based in the temporal lobe, and "rules," combinatorial systems subserved by procedural memory and based in the frontal lobe. If this hypothesis is valid, such a division could be a motivation for dissociated transfer of rule application versus structure building, and such a division could serve as a neurological correlate of weak transfer hypotheses like those proposed by Vainikka & Young-Scholten and Eubank.

Empirical data is also available to support this hypothesis. In an fMRI study of syntactic structure building, Brennan, et al. (2012) found that structure building was strongly correlated with left anterior temporal lobe activity, while no correlation was found between structure building and left frontal lobe activation. Peele, Troiani & Grossman (2009), on the other hand, conducted a study of rule-based versus similarity-based judgments of a concept's semantic features. The investigators found that the rule-based task provoked increased frontal lobe activity, while the similarity-based task resulted in higher temporal and parietal activation. The results of these two studies suggest that there may be a spatial division in the brain between rule application and structure building processes. Rüschemeyer, et al. (2005), add more support to the idea of physically separate language processing nodes with their fMRI study of L1 and L2 processing of specific (spoken) linguistic structures. The authors find that nonnative speakers' activation patterns associated with processing semantic violations is more similar to natives than that of syntactic violations, which implies a cerebrally-based separation between these two linguistic aspects. In addition, the nonnative speakers showed activation in different cortical areas than natives, possibly implying a different overall processing routine altogether.

The "structure" versus "rule" distinction discussed above has implications on language processing both broadly—across the major linguistic divisions—and more specifically as well. In section 5 below, this distinction will be extended to the possible separation between processing of grammatical gender and number in Spanish, both as it relates to native speakers and L2 learners. In the next section, discussion of language

acquisitions hypotheses continues with descriptions and analyses of the more modern

approaches to the subject.

2. Cognitive Contributions to Language Acquisition

The Critical Period Hypothesis and its immediate successors seek primarily to

determine the initial state of second language acquisition. These hypotheses take a

theoretical approach, but they stop short of examining second language use and cognitive

processing by proficient L2 speakers. Others in the field of language acquisition have gone

further, examining second language processing—rather than simple behavior—and

interactions with the native language. One of the major questions that these researchers

attempt to answer is how native language processing differs from, and how it is similar to,

processing of a second language. With the rise of detailed brain imaging and mapping

techniques like PET and fMRI, this segment of the language acquisition field has grown

immensely in the last two decades.

2A. The Shallow Structure Hypothesis. A leading theory in the field is the Shallow

Structure Hypothesis, first proposed by Clahsen & Felser (2006). Although anecdotal

evidence suggests that L2 speakers often suffer deficits in their language production and

comprehension compared to native speakers, there was not much evidence of the specific

character of this deficiency prior to Clahsen & Felser's work. Clahsen & Felser's paper

begins by rejecting four main observations based on experimental evidence in studies

conducted by various investigators and upon which previous theories of language

acquisition were based.

First, language learners perceive and understand linguistic input in a fundamentally

different way than native speakers: studies of parsing strategies for temporarily ambiguous

sentences (e.g. "garden path" sentences) have found that native speakers utilize structural

(e.g. syntactic) information much more than lexical-semantic and contextual cues, while L2

learners use the reverse tactics. Clahsen & Felser mention that this difference could be

attributed either to differences in the processing system itself or to general cognitive

limitations like lower working memory resources allocated to L2 comprehension. Second,

time course-sensitive imaging methods such as ERP (event-related potential, a form of EEG

monitoring) have revealed that L2 learners process language more slowly than native

speakers, "possibly reflecting a lack of automaticity" (Clahsen & Felser 2006:4) compared

to natives. Third, transfer from the native language may affect production and

comprehension of the L2, as explored above; obviously, native speakers are not limited by

previous linguistic knowledge, though they also do not receive any possible beneficial

effects of transfer. Fourth, L2 learners may have wide-ranging access to lexical retrieval

faculties but limited access to the procedural memory system associated with the rules

necessary to parse natural language, whereas native speakers have comprehensive access

to both systems. Clahsen & Felser describe each of these observations in detail and attempt

to reconcile each with the contemporary acquisition and processing hypotheses. They

ultimately offer a reasoned dismissal of each observation, with the rationale that none of

the four observations adequately explains the gulf between native speakers and language

learners.

    Instead, the authors offer their Shallow Structure Hypothesis (SSH). In essence, this

hypothesis is a modification of the fourth observation mentioned above. Clahsen & Felser

begin with the notion, from the Weak Transfer hypotheses, among others, that L2 speakers

are primarily deficient in areas involving sentence processing, compared to performance in

morphology-based tasks, on which L2 speakers perform roughly on par with native

speakers. The differing success in sentence processing between L2 speakers and natives

can be related to L2 speakers' inability to utilize syntactic cues in parsing—or, at the very

least, their overreliance on lexical cues at the expense of syntactic ones. Instead of building

a series of associations based on the structure of a sentence as a whole to support effective

and quick parsing, L2 speakers rely instead on short-distance lexical-semantic cues,

pragmatic information, and relevant extralinguistic signals. In effect, Clahsen & Felser

claim, L2 speakers initially parse sentences based on their knowledge of individual words

and phrases, as well as the world around them, but not based on the relationships between

phrases.

  2B. Empirical details of the SSH. Consider the following sentence: "Someone saw

the servant of the actress who was on the balcony" (Clahsen & Felser 2006:17). This

sentence may be interpreted in one of two ways, based on whether the relative clause (RC)

attaches to the matrix clause or the prepositional phrase: either the servant is on the

balcony, or the actress is. RC attachment of this sort varies cross-linguistically and is

generally systematic within a given language. English and Spanish differ in their preferred

RC attachment, yet even highly proficient L2-learners fail to acquire the specific RC

attachment. Rather than simply adhering to their native language's RC attachment

preference, however, these speakers tend instead not to prefer either attachment strategy

and essentially choose randomly between the two available NPs. When a thematic

preposition is present—that is, a preposition that explicitly assigns a thematic role to its

complement—any randomness vanishes, and a clear preference for a specific RC

attachment emerges (Felser, Roberts, et al. 2003; Papadopoulou & Clahsen 2003). These

and similar results advocate generally for a departure from a transfer-based theory of L2

acquisition and specifically for a view of L2 processing based on word- and morpheme-

level relations.

  Though the abovementioned studies suggest a clear and persuasive distinction

between natives and L2 learners, it is important to note that their results only surface in

on-line tasks. When highly-proficient L2 learners are asked to parse the same sentences in

an off-line setting, their judgments conform to those of native speakers (Clahsen & Felser

2006), indicating that the abovementioned effect cannot be linked strictly to grammatical

deficits per se. A later literature review by the same authors, Felser & Clahsen (2009),

similarly concludes that "even advanced L2 learners seem to rely more on lexical storage

and semantic information than on grammatically-driven processing mechanisms" (316).

Other investigators have observed similar on-line deficits alongside accurate off-line

performance as well (Montrul 2011). One explanation of this phenomenon could be that

L1-L2 transfer only occurs as a result of automatic processing and not during deliberate,

conscious (i.e. off-line) parsing. Perhaps L2 learners fall back on explicit language

instruction or remembered knowledge of specific linguistic input while performing an off-

line task, but are not able to use such a strategy on-line because of time constraints.

Although this explanation is intuitively reasonable, empirical evidence for or against it is

not yet available.

  In essence, Clahsen & Felser's Shallow Structure Hypothesis amounts to the above

assertion that any L2-speaker deficits are related to a lack—or a weaker

structuralization—of syntactic relations above the word level. Rather than building

relationships between words and phrases to construct a complex sentence structure, L2

speakers instead parse sentences heuristically, using their knowledge of words' thematic

structures (if not a priori then, at the very least, via predicate-argument relations),

semantic and realistic meanings, and pragmatic relations to the discourse as a whole.

Rather than build a syntactic tree as the mental representation of a sentence, as it is

assumed that native speakers do, Clahsen and Felser propose that L2 speakers construct a

series of non-hierarchical relations from which a coherent whole emerges.

      2C. The SSH and its relation to competing hypotheses. The SSH differs from

other contemporary hypotheses in that it proposes that L2 speakers process language in a

profoundly different way than natives (either mature speakers or young L1 acquirers). In

this respect, the SSH directly supports the conclusions of the CPH for L2—namely, that L2

learners who begin a language after the critical period do not acquire that language to the

same level of completion as native speakers. Furthermore, the SSH refines Penfield &

Roberts and their adherents' views of early language acquisition by proposing a precise

modality of post-critical L2 deficits, one of the main holes in the original hypothesis.

      In understanding Clahsen & Felser's contribution to the CPH, it is important not to

conflate their SSH with the weak transfer hypotheses of Vainikka & Young-Scholten and

Eubank. Though both the SSH and the weak transfer hypotheses support the conclusions of

the CPH, SSH puts aside transfer effects entirely in favor of a wholly organic view of L2

acquisition. Unlike weak transfer, the SSH implies that all L2 deficits are related to the act

itself of learning a second language at an advanced age, not to interference from the L1.

Clahsen & Felser's observations jibe with an account of the Critical Period based on the loss

or decline of neurological plasticity and reorganization, as advanced by Lenneberg as early

as 1967. Furthermore, the SSH avoids an arcane discussion of missing linguistic constructs,

relying instead on a more holistic approach that associates linguistic inequities with differing processing methods between natives and late L2 learners.

2D. Limitations of the SSH. Though Clahsen & Felser's Shallow Structure Hypothesis does provide the CPH with much greater explanatory power, it is not without its limitations. First, their claim that L1 transfer plays no part in L2 acquisition may be too strong. Sorace (2006) notes that the limited processing abilities that Clahsen & Felser cite as the impetus for shallow structure-building may actually encourage L1 transfer over shallow L2 structures. That is, an L2 learner may revert to their L1 syntax when such a practice would be more "economical." Furthermore, Sorace argues that it is likely that shallow processing is available to all speakers as a mechanism for parsing complicated structures, but that L2 speakers simply utilize this mode of processing more often. To this author's knowledge, these questions have not been adequately investigated, but they do raise legitimate issues with Clahsen & Felser's proposal.

Second, as Clahsen & Felser themselves note, the studies they cite focus almost exclusively on advanced language learners. There is not only a paucity of data on less-proficient learners (and, likewise, extremely proficient, "near-native" learners), but also on the changes in a speaker's grammatical processing methods that take place as his or her proficiency increases. Such documentation is crucial to the understanding of L2 acquisition, since subsequent studies have shown that proficiency level is an important influence of processing ability (Foote 2009). Proficiency level will be explored in detail in the next section, especially with regard to the question of whether and to what extent nativelike proficiency in late L2 learners is possible.

3. L2 Acquisition and Ultimate Attainment

3A. Advanced proficiency vs. near-nativeness. The relative dearth of data on

near-native L2 speakers has become more significant in recent years as experimentation

has revealed important differences between advanced L2 speakers and these near-natives,

who are sometimes said to have reached "ultimate attainment." Furthermore, Carroll

(2006) raises the important point that the lack of consistent terminology for proficiency

levels makes interpretation of studies of L2 learners quite difficult. Standardization of

terminology is especially important because, by both behavioral and neurological

measures, near-native, or end-state, speakers pattern separately from advanced speakers

who have not reached near-native levels of proficiency. Carroll also suggests that there may

be a fundamental problem with comparing L2 learners to L1 monolinguals rather than to

bilinguals (native speakers of more than one language); Gillon Dowens & Carreiras (2006)

agree, suggesting that investigators should compare late L2 learners to early L2 learners[2]

rather than comparing adult L2 processing to child L1 processing. In any event, the

separation between advanced and near-native speakers is visible even when comparing

only late L2 learners to one another.

This separation manifests itself in various ways. In a picture verification study by

Sorace & Filiaci (2006) focusing on interpretation of intrasentential anaphora, near-native

L2 Italian speakers employed many of the same parsing strategies as the native controls, as

evidenced by similar patterns of performance on null-subject sentences. However, in

contrast with the native controls, the experimental subjects tended to prefer to assign the

antecedent of overt pronouns to the matrix clause rather than the relative clause. Sorace &

---

[2] That is, L2 learning begun after infancy but before the critical period. Such a situation often occurs in children of linguistically isolated immigrants who speak one language until beginning pre- or elementary school, at which point they begin to learn the L2.

Filiaci argue that while the near-native subjects have successfully acquired the underlying

syntactic rules for such situations, some processing limitation prevents these speakers

from accurately parsing exceptional sentences such as those containing overt pronominals.

Furthermore, Sorace & Filiaci contend that transfer does in fact play some role in speakers'

performance on such tasks—in this instance, certain aspects of pronominal interpretation

strategies in English may exacerbate the existing processing difficulties that near-native

Italian speakers were shown to have with the particular construction under examination.

Jegerski, Van Patten & Keating (2011) continue the line of inquiry begun by Sorace &

Filiaci with a similar study of anaphora resolution in pro-drop and non-pro-drop languages,

specifically Spanish and English. Jegerski, et al. performed four experiments, two of which

established baseline strategies of anaphora resolution by English and Spanish natives, and

two of which tested the same strategies in L1-English Spanish L2 learners. The L2 learners

were divided into two groups for the two experiments, categorized as intermediate and

advanced. It is important to note that the advanced L2 speakers were of relatively high

proficiency but were crucially not near-native or end-state learners. From the results of the

first two experiments, the researchers found that Spanish natives tend to resolve questions

of anaphor antecedents using strategies predicted by Carminati's (2002) Position of

Antecedent (PA) hypothesis, the details of which are not relevant to this paper. English

natives, on the other hand, tend to refer to the discourse structure of a two-clause sentence

in order to decide the proper antecedent for a given anaphor. From these conclusions, the

researchers predicted that the actions of the intermediate and advanced learners would be

correlated with one of the established strategies of the natives.

Of the second-language learners, those at the intermediate level performed much like the native English speakers when parsing both English and Spanish sentences. That is, even when resolving anaphoric dependencies in sentences of their second language, they acted in certain respects as if these were sentences of English. Specifically, intermediate learners used the sentences' discourse structures, not the more nativelike PA strategy, to parse such sentences; in fact, no significant PA-like effects were observed within the intermediate group. In terms of anaphor resolution, the intermediate learners seemed to simply transfer their L1 strategies to the L2, even if, in other respects, their abilities were more nativelike.

In parsing the same sentences, the advanced L2 learners also utilized the sentences' discourse structure as part of their parsing routine; statistical analysis showed a borderline main effect for discourse structure. However, these speakers, in contrast to the intermediate learners, were observed to use a PA-like strategy as well. Though there was no statistically significant main effect for a general PA-like strategy, advanced learners did use this strategy in parsing sentences with coordination structures (as opposed to those with subordination structures, which were parsed randomly).

For the above results, the question arises as to how a single individual can utilize two processing strategies at once, given that they are often at odds and give conflicting results. According to Jegerski, et al., these speakers attempt to utilize both strategies in every situation. In certain situations, only one strategy is viable and produces a single, specific result. In other situations, the strategies conflict and the speaker chooses randomly between the two, as in the subordination example above. Furthermore, Jegerski, et al. propose that such dual-strategy behavior may actually reflect the transition from L1-

transfer-influenced parsing to more nativelike parsing of the kind seen in native speakers and perhaps near-native L2 speakers as well.

Alternatively, such hybrid behavior may suggest that what prevents these speakers from achieving nativelike performance is not a failure to acquire the correct strategies of the target L2 but rather a failure to *discard* the interfering strategies of the speakers' L1. Given that previous research has shown successfully nativelike performance in off-line tasks by L2 speakers (Clahsen & Felser 2006), the explanation above is certainly plausible. Jegerski, et al. say as much, citing previous research by Rothman & Iverson (2007) and others showing that the relevant syntactic construction has already been acquired by the time speakers reach a proficiency level on par with Jegerski, et al.'s intermediate group. Because the Jegerski, et al. study did not include end-state L2 learners, it is not possible to determine whether the advanced learners would have eventually shifted to or towards nativelike parsing strategies, given sufficient gains in proficiency.

3B. Attaining near-nativeness. Hopp (2006) aims to answer this question; namely, whether it is possible to attain nativelike proficiency. In a study of the parsing routines of English L1 and Dutch L1 speakers of German, compared to native German controls, Hopp seeks to separate L1 influence (i.e. transfer) from language proficiency levels of L2 speakers. To do so, Hopp separates 20 L1-English and 20 L1-Dutch speakers into two groups, advanced and near-native, based on the speakers' results on a C-test, a general language proficiency test that is designed to attach an objective rating of language proficiency to each speaker. Even though these speakers spent similar amounts of time immersed in the German language and were all residents of Germany at the time, Hopp is

able to separate these speakers into the two statistically distinct populations, the

aforementioned advanced and near-native groups.

Using two experiments designed to examine the subjects' syntactic reanalysis

abilities—specifically their ability to resolve case and verb agreement mismatches—Hopp

finds that not only do advanced and near-native speakers differ in their task success rate,

they also differ drastically in their parsing and reanalysis strategy. As might be expected,

the advanced group did not behave quite like native speakers when parsing the

experimental sentences: they were not sensitive to (syntactically-licensed) subject-object

scrambling in a self-paced reading task, and showed no difference in reaction time (RT) in a

speeded acceptability judgment task (compared to native speakers' increased RT).

Near-native speaker performance, on the other hand, was aligned with native

speakers in almost every way measured. In the self-paced reading task, near-native

speakers showed the same slowdown effects as a result of subject-object scrambling as

native speakers, and likewise experienced similar RT increases to the native speakers in

the speeded acceptability judgment task. Furthermore, near-native speakers outperformed

advanced speakers in determining grammaticality of case marking, performing in a

nativelike way even when under time pressure.

Although near-native speakers are quantitatively closer to native speakers than the

advanced speakers, it is important to note that both L2 groups are proportionally worse at

accurately judging case violations than verb agreement violations. Whether this

shortcoming represents effects of L1 transfer or computational difficulties is not entirely

clear. Because neither of the two L2-German groups—L1-English and L1-Dutch—speak

languages with noun case marking (apart from pronominal case marking), it is not possible

to determine from the results of the Hopp study alone whether the presence of robust case

marking in the L1 would contribute to better performance compared to verb agreement.

Perhaps the most revelatory conclusion that Hopp makes is his contention that the

results of the study do not support the broadest version of Clahsen & Felser's Shallow

Structure Hypothesis. While it is true that advanced L2-German speakers do not perform at

a level comparable to the native speakers, as predicted by the SSH, the near-native group

does not conform to the predictions of the SSH. Hopp's results show that these near-natives

process syntactic elements in a significantly deeper way than advanced speakers. They

reliably utilize syntactic features during phrase structure reanalysis, and they are able to

process these syntactic cues during serial parsing as well. Furthermore, the near-natives do

not use heuristic processing routines, similar to those observed in some aphasics, to

process sentences from a surface-relational perspective only (e.g. "a linear 'subject-first'

preference," Hopp 2006:391).

All of this is to say that, first, according to Hopp's findings, the precise proficiency

level of subjects can be very important when making conclusions on the nature of end-state

acquisition or the possibility of true near-nativeness. Once proficiency is accurately

quantified, it becomes clear that achieving native-level language proficiency in certain

aspects of language is indeed possible for late L2 learners. However, this is not to say that

all domains of language are fully attainable to late learners. According to some research,

late L2 learners may not be able to fully acquire all facets of language. For instance, Hahne

(2001) observes that even near-native speakers do not exhibit the same degree of

automaticity as natives do in integrating syntactic category information into phrase

structure. In this particular study, which utilized electroencephalogram (EEG) brain-

monitoring techniques, phrase-structure violations elicited a pronounced ERP (event-

related potential) effect in native speakers but not near-natives, suggesting that the L2

speakers were not sensitive to these violations on the same level as the native speakers.

Although Hahne did not use the same kind of rigorous language-proficiency testing as

Hopp, her results suggest that certain domains of language—specifically morphosyntactic

features—cannot be acquired to the same extent as syntactic features like case marking.

Keating (2009) comes to a similar conclusion. In an eye-tracking study of beginning,

intermediate, and advanced late L2-Spanish speakers[3], the investigator found that,

although the advanced speakers are sensitive to noun-adjective gender agreement

mismatches when such mismatches occur within a DP, their sensitivity drops off as the

distance between the agreeing constituents increases. Native speakers, on the other hand,

suffer no such drop-off in sensitivity. Keating concludes that, while gender agreement is in

principle acquirable even after the critical period, late L2 learners do not seem to be able to

achieve nativelike sensitivity to gender agreement even at endstate proficiency. However,

he goes on to say that the broad trend of sentence-type-related reading times for advanced

L2 learners tilts in the direction of nativelike performance, even on long-distance noun-

adjective dependencies, so it is possible that with further progress, the advanced speakers

could exhibit more nativelike performance in this domain. To be sure, Hopp's conclusions

are at odds with those of Hahne and Keating in certain important ways, namely with regard

to their observation of true nativelike processing in their subjects. Even so, all three

researchers decide that, given the trajectories of their near-native subjects' linguistic

---

[3] Again, these L2 speakers were described as beginning, intermediate, or advanced because of their exposure to the language, not because of quantitative proficiency scores. In justifying the "advanced" label, Keating notes that all members of this group had at least a B.A. in Spanish and had experienced at least one period of complete Spanish immersion.

abilities, nativelike processing may be achievable at higher levels than those observed in

their subjects, even if such proficiency cannot be attained on a broad scale. In the following

sections, the possibility of nativelike processing by late L2 learners will be explored within

the framework of grammatical gender and number processing.

4. Descriptive Accounts of Grammatical Gender and Number Processing

  4A. Behavioral data. In the section above, it was noted that Keating's initial results on the acquisition of grammatical gender by L1-English speakers suggest that late learners can fully acquire a structure not present in their native language. As mentioned in section 1, above, the results of the behavioral study by Alarcon (2011) suggest that L2 learners can indeed acquire abstract linguistic features such as Spanish gender, perhaps just as capably as heritage speakers. Sagarra & Herschensohn (2010) continue this line of investigation by examining L2 learners' processing of gender and number agreement via various tasks involving Spanish nouns and their adjectival complements, which must agree in gender and number in Spanish. After testing native speakers and English-L1 Spanish-L2 beginning and intermediate learners, the investigators found that, as predicted, beginning L2 learners performed poorly on on-line and off-line discrimination tasks of gender and number agreement, native speakers performed at ceiling levels, and intermediate L2 learners performed somewhere in between these two groups. Furthermore, all three groups' off-line performances indicated equal performance on gender mismatches and number mismatches. While the lack of a significant performance difference between gender and number concord/discord in the beginning learners and the native speakers can be explained by floor and ceiling effects, respectively, the performance of the intermediate learners cannot be explained as such. Indeed, the intermediate L2 learners identified number mismatches more accurately than gender mismatches in the off-line grammatical judgment task (natives again performed at ceiling levels). Furthermore, working memory capacity correlated to reaction times to gender disagreement but not to number agreement, suggesting that gender parsing is more cognitively taxing than number parsing.

The investigators reconcile the differing results of the on-line and off-line tasks by stating

that their (on-line) self-paced reading task is not sensitive enough to discriminate between

processing performance based on such fine distinctions. This explanation is not altogether

satisfying. A better interpretation might invoke the relatively low proficiency of these

intermediate L2 learners, a group which has been shown, as mentioned above, to perform

significantly differently from more advanced learners and from near-natives, possibly to an

exponential extent.

Regardless of the explanation of the lack of demarcation between gender and

number in the on-line task, the fact that Sagarra & Herschensohn do find any distinction

among beginning and intermediate learners between gender and number concordance

suggests that such a processing distinction may in fact exist. A subsequent study by the

same authors (Sagarra & Herschensohn 2011) on gender agreement processing of

inanimate versus animate nouns showed that intermediate learners tend to behave more

like native speakers than like beginners. That is, they exhibit longer reaction times (and

therefore more processing difficulty) for animate than inanimate nouns and for gender

discord than concord. If the results of this second study can be extended to those of the

first, they suggest that the test itself may not have been the source of the lack of

differentiation between gender and number, especially considering that the same

methodology was used in both studies. Rather, it seems that some other issue was at play.

Perhaps the low proficiency of the intermediate learners allows for a distinction to be made

between animacy and inanimacy but not between gender and number and such a

distinction will emerge in higher-proficiency speakers. Regardless of the specific

deficiencies of the 2010 study, the question of whether gender and number agreement

carry different processing burdens is one of the major inquiries that the study at hand seeks to address and one that this investigator feels has not been adequately engaged in the past.

   4B. Neurological data. Electrophysiological data supports the behavioral findings that highly advanced or near-native L2 speakers can achieve similar proficiency to that of native speakers. Gillon Dowens, et al. (2010) conducted an ERP study of L1-English, near-native L2-Spanish speakers in which subjects (and native Spanish controls) were exposed to sentences containing determiner-noun and noun-adjective number and gender mismatches either within a phrase or across a phrase boundary while their brain activity was monitored via EEG. Compared to the native controls, the L2 speakers exhibited similar patterns of brain activity. Intra-phrase violations elicited the same pattern of an early left anterior negativity (LAN) followed by a P600 response from the L2 speakers as from the natives. When a violation occurred that spanned across two phrases, however, no LAN effect was elicited. The investigators ascribe this difference to the increased processing demands that come with retaining information of long-distance dependencies, which would presumably reduce a speaker's ability to react to these kinds of violations. Returning to the intra-phrase responses, while the native speakers showed identical responses to gender as to number violations, the L2 speakers showed qualitative differences between the response curves for intra-phrase gender versus number violations, suggesting that these constructions are not being processed in a completely nativelike manner.

   Although the observations that this study makes on the nature of second language processing are clear, the question remains as to whether L2 transfer substantially contributes to L2 speakers' division between gender and number violations. A similar

study by Gillon Dowens, et al. (2011) seeks to answer this question. Using the same

methodology as Gillon Dowens, et al. (2010)'s study, the experimenters tested L2 Spanish

learners whose native language was Chinese, rather than English, against native Spanish

speakers, again using ERP techniques. Unlike English and Spanish, Chinese contains neither

gender nor number agreement,[4] so there is no chance that a difference in processing

between gender and number agreement could be conditioned by L1 transfer. ERP

recordings showed a P600 effect in response to both gender and number violations, with

no reliable difference between the two conditions. Unlike the English natives described

above, the Chinese natives showed no LAN effect even when a violation occurred within a

phrase. The prevailing explanation of the LAN is that it corresponds to more automatic

stages of morphosyntactic processing, while the P600 is thought to be related to more

controlled processing of higher syntactic features (Friederici, Steinhauer, & Pfeifer 2002;

Friederici 2001). Therefore, Gillon Dowens, et al. (2011), suggest that the lack of any LAN

response from the Chinese natives could be due to their somewhat lower proficiency in

Spanish—they were living in a Chinese-speaking environment at the time of the study and

were not immersed in the L2—compared to the English natives, who were living in a

Spanish-speaking country and were speaking in Spanish on a daily basis. Unfortunately,

such a complication could be a major confounding factor for this type of research and it is

therefore difficult to take any solid conclusions from the lack of a LAN. Even so, the fact that

the Chinese natives did not show a difference in their P600 responses to gender versus

number violations suggests that L1 transfer may indeed have played a role in the English

natives' differential responses observed in the earlier study.

---

[4] In fact, Chinese completely lacks any inflection for gender, number, person, or case. (Gillon Dowens, et al. 2011)

Although the abovementioned work by Sagarra & Herschensohn and Gillon Dowens, et al., is a good first step to describing the behavior of advanced and near-native L2 speakers in response to gender and number violations, these authors do not attempt to provide explanations for this behavior. The following section continues in this direction, exploring the underlying cognitive architecture supporting nativelike parsing of L2 gender and number processing.

5. Foundations of Gender and Number Processing

The studies above begin to provide a clearer picture of the possible differences or similarities in processing gender versus number inflection in Spanish. However, these studies tend towards description of behavior and cerebral activity, while leaving out any deeper syntactic and cognitive basis for these observations. The studies discussed below, on the other hand, delve deeper into the cognitive underpinnings of observed distinctions between gender and number, as well as between native and nonnative processing of these features. Although the data from the two studies headed by Gillon Dowens, discussed above, seems fairly clear-cut and cogent at first blush, other studies have come to different conclusions about processing of gender and number marking. For syntactic reasons that are largely beyond the scope of this paper, it has been argued at least as early as 1993 (by Ritter) that gender and number may be fundamentally different morphosyntactic constructs. This claim can be compared to the hypothesis described in section 1D, above. Perhaps gender and number fall on different sides of the division between "words" and "rules," to use Ullman's terminology. Gender, an intrinsic lexical feature, would belong to the "words" subset, while number, a feature layered onto the existing lexical item, could be thought of as a "rule" applied after lexical retrieval. If this kind of categorization is valid, it

is easy to extend this difference to processing as well and to hypothesize that these two features (for lack of a better word) may be processed differently from one another, even in native speakers. Preliminary behavioral studies seem to bear this out: in one study of gender and number agreement errors by native speakers (Vigliocco, Butterworth & Garrett 1996), number errors far outnumbered gender errors. If the lower instance of gender agreement errors can be taken to be suggestive of a deeper integration of gender than number, these results imply that gender may be an intrinsic part of a lexically-stored word or lemma, while number is not (and perhaps is instead derived by the application of a rule or some other mechanism).

The claim that grammatical gender is inherent to the noun while number is rule-derived has interesting implications. In a study of reaction times to gender versus number incongruencies, Faussart, Jakubowitz, & Costes (1999) found that gender agreement violations elicited a longer reaction time than number violations. The authors explain this finding by positing that the reanalysis of a morphosyntactic feature that is processed earlier in the lexical retrieval routine requires that more of the routine be re-run, while a later feature like number only requires that the final step of the routine be repeated. Therefore, the reanalysis process, measured in this study via reaction time, is shorter for number agreement violations than those for gender, even though gender is processed at an earlier and possibly more innate or instinctive stage in the parsing process than number. Without physiological evidence like that obtained through the use of ERP techniques, however, such hypotheses are mostly limited to a broad psycholinguistic conception of the phenomenon at hand. Luckily, many ERP studies on the subject have been conducted in recent years.

Certain ERP studies in the past have sought to answer exactly the question of whether gender or number is comparatively easier to process than the other. One such study, conducted by Barber & Carreiras (2005), examined gender and number agreement in Spanish both between determiners and nouns and between nouns and adjectives. Although many studies of ERP responses to either gender or number agreement individually have been conducted in the past, no study had previously looked at gender and number agreement within the same study while also using a language like Spanish that contains both markers in a grammatical framework (compared to a language like English in which gender is determined only by real-world semantic value and there is no robust gender agreement). Therefore, the Barber & Carreiras study was effectively the first to directly compare processing of gender versus number mismatches via ERP.

The study consisted of two separate but related experiments. In the first, native Spanish-speaking participants were given two-word pairs, either an article and a noun or a noun and an adjective. In the second, a separate group of native Spanish-speakers were shown a sentence with an agreement violation either at the beginning of the sentence—between the article and the subject noun—or in the middle of the sentence—between the subject noun and the adjective. Filler sentences without agreement errors were also included. In both experiments, participants were then asked specifically to judge the grammaticality of the word pair or sentence, though the subjects were not instructed to complete the task within any kind of time limit. The subjects' cognitive activity was monitored throughout via ERP recording.

In accordance with the research described above, Barber & Carreiras's results suggest that processing of gender violations is indeed more onerous than processing of

number violations. Specifically, ERP results showed that both types of agreement violation

elicited a P3 component (a positivity between about 200 and 500 ms) in response to the

word pairs, and a P600 component in response to the sentences. The two types of

violations differed qualitatively, however: gender violations elicited a later P3 and P600

than the equivalent number violations. These differences in number versus gender

processing occurred in late time windows, during which parsing reanalysis is generally

thought to occur. This finding can be explained if it is assumed that gender is a lexical

feature as well as a syntactic feature (because syntactically connected words must agree in

gender), while number is solely syntactic, as it is layered on top of preexisting lexical items

after retrieval. Therefore, reanalysis of a gender agreement violation will have to check not

only syntactic structure-building but also lexical access, while detection of a number

violation would only require syntactic structure reanalysis.

The above results are unlike those of previous ERP-based studies which found no

dissociation between processing of gender and which seemed to contradict behavioral

studies showing robust reaction-time differences. Instead, Barber & Carreiras's findings are

consistent with previous behavioral results, showing delayed late-stage processing of

gender compared to number violations.

Interpretation of these findings with respect to L2 processing of gender versus

number could go one of two ways. If it is assumed that L2 learners assign gender to the

lexical entries for each word in their lexicon, then L2 processing should follow the same

pattern as native speakers, with gender violations eliciting a longer response than number

violations. However, if it is instead assumed that L2 learners without native syntactic

gender cannot internalize gender to the same extent as native speakers, which perhaps

could explain Gillon Dowens, et al. (2011)'s L1-Chinese results, discussed above, it is

feasible that L2 learners treat gender like number in that it is applied after lexical retrieval,

perhaps through some other retrieval method. Even so, the argument could be made that

whatever lexically-external means is used to retrieve gender information would, upon

reanalysis, exact a similar processing cost to lexical re-retrieval in native speakers.

In the final introductory section below, the current study is discussed both with

reference to the research examined above and with a mind to the detailed methods, results,

and conclusions that follow.

6. The Current Study

The current study approaches the question of gender vs. number processing

differentiation from a behavioral perspective. Using the self-paced reading methodology,

participants' reaction times to gender and number violations between nouns and adjectives

within a sentential framework were used to evaluate Spanish-L2 learners' facility for

processing the two markers compared to native controls. Based on the research presented

above, we predict that native speakers will exhibit longer reaction times to gender errors

than to those for number, both of which will be longer than the reaction times to the error-

free control sentences. L2 learners with English as their native language, on the other hand,

are predicted to react more slowly to number than gender errors. Because the L2 learners

are assumed to already be familiar with number agreement through transfer from their L1,

they will be sensitive to number agreement in their L2 as well (according to any of the

various theories of L1 transfer). Being sensitive to number agreement, they will exhibit

longer reaction times when number agreement is violated, and delayed reactions in

response to number will surface immediately following the disagreeing word. In response

to gender errors, on the other hand, L2 learners will not be sensitive (or as sensitive) to

gender agreement compared to number. This attenuated sensitivity could surface as either

limited differentiation between reaction times to gender disagreement compared to

control, or it could surface as a delayed effect, in which case the slower reaction time would

still be elicited by gender errors, but instead of emerging immediately following the

disagreeing word, it would consistently emerge later in the sentence.

Evidence from the study by Gillon Dowens, et al. (2010), among others, suggests that L2

learners process gender qualitatively differently from number. Furthermore, evidence from

Keating (2009) and others suggests that while advanced and near-native L2 speakers are

often able to process syntactic constructions without an L1 analogue with nearly the same

facility as natives when no additional processing cost exists (e.g. long distance

dependencies for gender, number, case, etc.), their performance falls off with the addition

of these types of external costs. This result suggests that near-natives may indeed possess

similar processing capabilities to natives but that their processing gets delayed and bogged

down when extra requirements are introduced. In a situation such as the one under study

here, where L2 learners' reactions to errors are judged in an on-line scenario, while

processing of a feature present in the L1, such as number agreement, may be processed in a

nativelike manner, processing of a feature such as gender agreement that does not benefit

from transfer effects may be delayed compared to number. Therefore, it is predicted that,

like native speakers, L2 learners will still exhibit slower reaction times to gender errors

than to control sentences. However, these reaction times will surface later on, perhaps after

the word following the error rather than after the error itself. On the other hand, L2

learners will process number errors immediately rather than in the delayed fashion that

they will process gender errors, resulting in reaction time delays immediately after the

error. Finally, as general trend, L2 learners may exhibit slower reading times overall than

native speakers, as observed by Sagarra & Herschensohn (2010). The precise methodology

utilized here will be described in detail in the next section.

Methods

Participants

Participants were drawn from the student population of two Spanish summer school

classes at a northeastern United States university, as well as from students in other

departments at the university. All participants were native speakers of either American

English or Spanish. Participants were divided into two groups based on length of Spanish

exposure, either in a formal or naturalistic setting: advanced (greater than three semesters

of Spanish, some Spanish immersion), and native (Spanish immersion from before age 10,

continued Spanish use). Division of participants was similar to that of Sagarra &

Herschensohn (2010). Data on which the division of participants was based was obtained

through a short language background questionnaire administered prior to the experiment.

Participants were compensated $5 in the form of an Amazon gift card.

Stimuli

Language Background Questionnaire

Prior to the experimental task, subjects were asked to complete a questionnaire detailing

their linguistic background in Spanish. Subjects were asked about language nativity, age of

acquisition, speaking frequency (i.e. hours per week, approximately), living experience in a

Spanish-speaking country, and experience in formal Spanish classes. Subject responses

were anonymous and were tied to experimental responses using subject codes that were

not connected to any personally identifiable data.

Self-Paced Reading Task

The present study utilized a self-paced reading task in which participants were asked to

read sentences and respond to the prompt, "Does this sentence accurately describe the

picture?" given a sample picture. The experimental sentences were either grammatical or

contained errors in determiner-noun agreement in either gender (masculine/feminine) or

number (singular/plural). Sentences that were semantically anomalous given the picture

were also included as filler sentences. For example:

1. *Hay un hombre a la izquierda que lleva un camisa con estrellas blancas.
    "There is a man on the left wearing a$_{MASC.SG}$ shirt$_{FEM.SG}$ with white stars."

2. *Dos soldado vestidos de verde detienen un hombre delante del coche.
    "Two soldier$_{SG}$ dressed in green are arresting a man in front of the car."

3. #El vendedor de cuadras a la izquierda lleva una camisa azul.
    "The picture vendor on the left is wearing a blue shirt." – anomalous given that the
    vendor pictured is wearing a pink shirt

4. Hay cinco hombres sentados bajo del toldo de una tienda.
    "There are five men seated below the awning of a store."

Of the 80 sentences presented, 40 were semantically relevant to the picture and 40 were

semantically anomalous. In addition, 40 of the 80 were grammatical and 40 were

ungrammatical. Of the ungrammatical sentences, 20 contained gender mismatches and 20

contained number mismatches. Ungrammatical sentences were distributed throughout the

semantically-relevant and semantically-anomalous groups. Participants were instructed to

read the sentences as quickly as possible without sacrificing comprehension. Participants

were also informed that the sentences might contain errors but that they should only

answer the question of whether the sentence corresponded to the picture based on the

information contained in the sentences.

After each sentence, a picture was displayed on the screen and, as mentioned above,

participants were asked to identify whether the information contained in the preceding

sentence matched the picture. Question response accuracy rate was used solely to

determine whether the participants were in fact attentive to the stimulus at hand.

The experimental paradigm for self-paced reading was a non-cumulative Moving

Windows task, as in Just, et al. (1982), and was run using the Paradigm software

(Perception Research, 2012). Sentences were displayed onscreen as a series of dotted lines,

each representing a single word; the end of the sentence was signaled by a period.

Participants were asked to press a button on the computer keyboard, which would show

the first word of the sentence and then each subsequent word, hiding any previous words.

After the last word of the sentence was displayed, a picture and then the question "Does the

sentence accurately describe the picture?" were displayed, followed by a prompt for yes or

no. Data was collected for time between button presses as well as for question response

accuracy.

Results

Self-Paced Reading Task

The study followed a mixed effect model, with one between-subjects factor, group

(L1 versus L2 speakers), and one within-subjects factor, condition (gender violation,

number violation, or gender/number agreement).

Table 1. Mean RTs at the word immediately preceding the noun, the noun itself, and the word immediately following the noun, in milliseconds.

	Gender Violation		Number Violation		Gender/Number Agreement	
	Mean	SD	Mean	SD	Mean	SD
Mean RTs at the word immediately preceding the noun						
L1 speakers	496.3	170.2	540.3	264.4	776	3316
L2 speakers	403.3	119.4	448.1	234.0	404.8	248.0
Mean RTs at the noun						
L1 speakers	843.1	689.0	839.0	651.6	678.8	649.4
L2 speakers	577.3	332.7	685.2	490.3	553.0	444.4
Mean RTs at the word immediately following the noun						
L1 speakers	790.4	571.8	915	834	641.8	630.3
L2 speakers	635.3	418.1	707.6	534.1	519.7	354.0

Note: n = 3 for L1 speakers and n = 4 for L2 speakers; K = 20 for gender violations, K = 20 for number violations, and K = 40 for agreement.

Results were analyzed via three two-factor ANOVAs of reaction time (RT) with a 3 × 2 factorial design (determiner agreement × group): one for the word immediately preceding the noun, one for the noun, and one for the word following the noun. Descriptive statistics for the self-paced reading trials are presented in Table 1, above. RT preceding the noun was analyzed to ensure baseline stimulation preceding the experimental conditions. RT to the word following the noun was analyzed to determine whether subjects experienced delayed or prolonged processing of the experimental conditions. Late processing effects are important because differences between gender and number processing often have a delayed onset (Barber & Carreiras 2005), and because L2 processing is generally delayed compared to L1 (Weber-Fox & Neville 1996). In addition, performance on comprehension questions was analyzed to ensure higher-than-chance accuracy, which would indicate that subjects were paying adequate attention to the sentences at hand.

Reactions to the Word Preceding the Noun The ANOVA performed on the word immediately preceding the noun showed no main effect for either group (F(1,554) = 1.77, *p*

> 0.05) or condition (F(1,554) = 0.652, p > 0.05), and no interaction was found between group and condition (F(2,554) = 0.575, p > 0.05). A post-hoc T-test between L1 and L2 speakers did not find a significant difference in reading time of the kind found by Sagarra & Herschensohn (2010) and others.

Reactions to the Experimental Noun The ANOVA performed on the noun itself showed main effects for both group (F(1,554) = 13.82, p << 0.05) and condition (F(1,554) = 3.75, p < 0.05) but no interaction between group and condition (F(2,554) = 0.80, p > 0.05). Post-hoc one-way ANOVAs performed within each group but across conditions showed no main effect for the L1 group (F(2,237) = 1.86, p = 0.158), while the L2 group's results approached a significant main effect (F(2,317) = 2.56, p = 0.079). In addition, post-hoc T-tests found a significant difference for number violation RT compared to agreement for L2 speakers (p = 0.044), but not for gender (p = 0.636). Results approached significance for number (p = 0.118) and gender (p = 0.123) violation RTs compared to agreement for L1 speakers, suggesting that a larger sample size might have revealed a significant result in this domain. Furthermore, a T-test comparing gender violation RTs to number violation RTs showed no significant result with an extremely high p value for L1 speakers (p = 0.973) and results approaching significance for L2 speakers (p = 0.106).

Reactions to the Word Following the Noun The ANOVA performed on the word immediately following the noun showed a main effect for group (F(1,540) = 10.72, p << 0.05) and condition (F(1,540) = 8.51, p << 0.05) but no interaction between group and condition (F(540) = 0.27, p >> 0.05). Post-hoc T-tests showed significant differences between number violation and agreement RTs for L1 speakers (p = 0.028) and results approaching significance for gender violations versus agreement for the same L1 speakers

(p = 0.118). Significant differences were observed in the results of T-tests between both

gender (p = 0.037) and number (p = 0.005) violations and agreement for L2 speakers.

Gender vs. number comparisons of RT to the word following the noun did not yield

significant results for either L1 or L2 speakers (L1: p = 0.342; L2: p = 0.316).

Comprehension Questions

Table 2. Subject performance on comprehension questions.

	Gender Violation		Number Violation		Gender/Number Agreement	
	Mean	SD	Mean	SD	Mean	SD
L1 speakers	11.33	0.5774	15.33	0.5774	31	2.646
L2 speakers	16.75	1.258	12	1.826	31.25	1.5

Note: n = 3 for L1 speakers, n = 4 for L2 speakers; K = 20 for gender and number violation and K = 40 for agreement.

As for the comprehension questions, all participants performed significantly above

chance, as shown by individual binomial tests of each subject's data (p << 0.05 for all

subjects). Data broken down by sentence type and speaker group is presented in Table 2,

above. The average score was 73.75% correct, or 59 out of 80 questions correct. However,

the fact that six of the 80 questions were answered incorrectly by every subject and one

more was answered incorrectly by all but one subject likely points to bad sentence

construction for the given picture associated with those questions than to overall poor

subject performance. A T-test comparing the results of the comprehension questions by the

L1 speakers versus the L2 speakers did not show a significant difference between the two

populations (p = 0.223)

Background Questionnaire

Questionnaire results indicated that three of the seven participants were native

speakers, while four were advanced L2 learners. L2 speakers had studied Spanish for

between two and 20 years, with a mean length of 10 years.

Discussion

The results of the self-paced reading task have various implications. I will begin by

discussing the data from each sentence position independently of the others. Then, I will

discuss the results generally with regard to their implications for gender acquisition by L2

speakers specifically and second language acquisition as a whole.

1. Experimental Data and Results

1A. Reactions to the Word Preceding the Noun First, the results for the word

immediately preceding the noun did not show any meaningful differences between

conditions or groups, nor did they show any interactions between the two. These results

largely follow the predicted outcomes: there should be no significant effect produced by the

word preceding the noun since this word was not selected specifically for any reason.

Indeed, this word was often a determiner or a number word, which are very common and

should not elicit any specific response. It was predicted, however, that L1 speakers would

have lower reaction times overall than L2 speakers given that L1 speakers are known to be

able to read faster and more automatically than L2 speakers (Sagarra & Herschensohn

2010). This type of effect was not observed in these results. It is unclear whether this was

due to the actual nonexistence of this effect in the studied populations or to the small

sample size of this study that may not have allowed such an effect to emerge. As a result, it

is possible that with more subjects, there would be a clear difference overall in mean RTs between the L1 and L2 groups to the word preceding the noun (as well as the noun and the following word).

1B. Reactions to the Experimental Noun As predicted, the data collected on RT to the experimental noun differed from those on RT to the preceding word. Namely, significant results were observed for the L2 population and results approaching significance were observed for the L1 population. Main effects for both group and condition suggest that both variables play a role in subjects' comprehension of these types of sentences. For Spanish natives (the L1 group), although individual T-tests did not show significant differences between gender violations and the agreement control sentences or number violations and agreement controls, the fact that the results approached significance suggest that a larger sample size might have yielded reaction times to gender and number violations that were statistically significant compared to those to agreement. The L2 group did show this kind of differentiation with respect to number violations compared to agreement, though not to gender violations. This result can be explained by the fact that, because English has number but not gender agreement between determiners and nouns, the L2 speakers were sensitive to number violations but not to gender violations in the narrow window immediately after the experimental noun. This possibility will be explored further in the General Discussion below. Native speakers, on the other hand, only reacted at time scales approaching significance. Although it is possible that a larger sample population would have produced more robust RT data, it can only be conjectured that natives would react significantly slower to gender as well as number mismatches. That native speakers did not show different reaction times for gender compared to number—an extremely high

p value was calculated for the reaction times to gender compared to number mismatches—

suggests that, in this early time window, native speakers react quite the same to gender

and number errors. If this is the case, native speakers would seem to lack the separation in

their responses to gender and number mismatches that has been proposed by Faussart,

Jakubowitz, & Costes (1999); again, this possibility will be examined further below.

However, because significance was not achieved, further experimentation is necessary to

confirm or deny this possibility.

      1C. Reactions to the Word Following the Noun Looking at the results for RT to the

word following the noun, a main effect for condition suggests that speakers remained

sensitive to agreement beyond the window immediately after the noun, while a main effect

for group suggests that the sustained responses of the two groups were still divergent from

one another. Post hoc, the results of the various T-tests give a somewhat different picture

from the noun-RT data discussed above. In response to the word following the noun, L1

speakers responded significantly slower only to number violations but not gender

violations (although gender RTs approached significance). This result suggests that

number reanalysis by the native speakers occurs more quickly that gender reanalysis.

Furthermore, it accords with work by Barber & Carreiras (2005) and Faussart, Jakubowitz,

& Costes (1999) (described in the literature review, above): if gender is both a lexical and

syntactic feature, it would be built into the syntax at the lexical retrieval stage, while

number, a syntactic but not lexical feature, is attached to words already built into the

syntactic structure. Therefore, number reanalysis would be a shorter process than gender

reanalysis, leading to a less-delayed reaction time for number than gender. This hypothesis

will be explored in greater detail below. Although a significant difference in RT could not be

obtained when comparing gender and number directly, the available results suggest that such a possibility should be explored in greater detail, as a statistically significant difference between the two would further highlight the delayed processing of gender errors compared to number.

L2 speakers, on the other hand, showed significant differences in reaction time for both gender and number as compared to agreement. Compared to their reactions to the noun itself, this result suggests that L2 speakers indeed exhibit delayed processing for gender alone, even when processing of number violations is not delayed (significant RT differences due to mismatch reanalysis were observed between number violations and agreement at the noun). If it were true that L2 speakers could not acquire a nativelike gender processing ability, these speakers would not show slowed reaction times in response to gender errors but rather would only exhibit such slowing in response to number errors. Instead, the L2 speakers reacted more slowly to both mismatch conditions than to agreement, suggesting that both gender and number errors provoked a nativelike response from the L2 speakers at the word following the noun. The fact that these speakers were observed to react to gender mismatches in such a nativelike way suggests that gender may be acquirable by L2 speakers past the superficial level that the various Critical Period hypotheses suggest.

2. General Discussion

As discussed in the introduction to this paper, although the behavioral data on native speaker processing of gender and number in Spanish does not always show a clear distinction between the two features, neurophysiological data from Barber & Carreiras (2005) and others show a concrete qualitative difference in processing of number versus gender mismatches. Combined with the behavioral data that do yield a significant result,

gender agreement production has been shown to be more automatic than number, while

gender mismatch comprehension and reanalysis has been shown to be delayed compared

to number. As mentioned above, data collected on the RTs of L1 speakers to gender and

number mismatches was aligned with the conclusions from the literature, though the

results obtained were not adequately significant to confirm or deny any difference between

responses to gender versus number mismatches. Luckily, the results for the experimental

group – the L2 learners – were more conclusive.

Because the L2 Spanish speakers were all native speakers of English, it was

predicted that number processing would be more automatic (faster) than gender, which

would be delayed. As mentioned in the introduction, English has gender marking on

neither common nouns nor determiners, and there is no syntactic gender agreement

whatsoever. Members of these word classes often do carry number marking, however, and

number must agree between nouns and certain types of determiners (e.g. this/these,

that/those, a/some, etc.). Therefore, English natives should be familiar with number

marking and agreement as a functional concept to a greater extent than gender. The

expected consequences of this observation vary depending on adherence to Clahsen &

Felser's Shallow Structure Hypothesis. According to the SSH, transfer does not play a role in

L2 acquisition, although as described in the introduction, subsequent work by Sorace

(2006) and others suggests that this position is too strong and that lower L2 proficiency

may actually encourage parsing via transfer up to a certain point. The results of the present

experiment suggest that L2 learners employ at least some degree of transfer in their

parsing of number mismatches. L2 learners exhibited a significantly slower reaction time

to number but not gender mismatches compared to agreement immediately following the

noun, while native speakers showed approximately similar reactions to gender and

number mismatches following the noun. As a result, it is likely that L2 learners did utilize

transfer to a certain extent. If L2 learners were not using their English-derived facility with

number agreement, they would be expected to react equivalently to gender and number

mismatches. Instead, a statistical comparison between these speakers' responses to gender

and number mismatches was nearly strong enough to justify the statement that L2 learners

process number errors more natively than gender, suggesting that a larger sample size may

have elicited significant separation between their reactions to the two features. The fact

that L2 learners reacted to both gender and number mismatches with significantly

different RTs from agreement at the position of the word after the noun suggests that at

this later time window, gender mismatch processing has caught up with number. The

insignificant difference between the two reinforces the idea that gender mismatch

processing is present but delayed in L2 learners.

Although the SSH explains many situations involving L2 processing, it is possible

that performance differences between the two features could be attributed to the features'

differing concreteness and relation to reality rather than pure syntax. In both English and

Spanish, number marking is directly related to the reality that the sentence describes: a

plural noun describes a group of objects or people, while a singular noun describes only

one object (or collection of objects, in the case of words like "set," "family," etc.). Gender, on

the other hand, is completely abstract except when describing animate objects (people or

animals). Furthermore, nouns describing animals often do not correspond in any real way

to an animal's physical sex (e.g. *el perro*, the-M dog-M vs. *la rana*, the-F frog-F), and even

some words describing people do not correspond to sex or gender (e.g. *el testigo*, the-M

witness-M, which can refer to both male and female witnesses, or *la victima,* the-F victim-F,

which can refer to both male and female victims).[5] Therefore, conceptually speaking, a

concrete feature like number is easier to reconcile with the subject matter at hand than

gender, a feature that is largely abstract. Furthermore, Corbett (1991) notes that, while

gender is abstract, it is also invariable for a given noun, except in cases where the noun's

gender reflects the gender of the person described by the noun. Number, on the other hand,

is variable and may take one of two forms in both English and Spanish. Therefore, L1-

English Spanish L2 speakers might have greater ease with the processing of number than

gender because of the processing cost demanded by featural (gender) retrieval. Native

speakers' lexicons, on the other hand, would not treat gender as a feature on top of the

noun but instead would contain lexical items in which the noun and its gender are tightly

integrated and largely inseparable in terms of recall. Therefore, native speakers' lexical

retrieval processes would recall both gender and noun together and automatically, while

the decision of whether to mark the noun as singular or plural (or retrieve the singular or

plural form) would require an extra processing step not required by gender recall, allowing

faster production of gender than number. Error reanalysis would then proceed by checking

agreement of the disagreeing feature. Number mismatches would provoke a check of

number, a feature *applied to* the noun with the same feature on the determiner, which

would not require a full lexical retrieval process. Gender mismatches, on the other hand,

would require the *noun itself* to be checked with the determiner's gender, instantiating

lexical retrieval and ultimately taking longer than number mismatch reanalysis. Therefore,

---

[5] See Appendix A for more

though gender would be produced more automatically than number, gender mismatch

reanalysis would take longer than the equivalent process for number.

        Although the L2 learner data does not match up exactly with the SSH, it supports the

above hypothesis; namely, gender is retrieved and built onto the syntax before number and

therefore gender error reanalysis educes a longer reaction time than the equivalent

process for number. However, unlike native speakers, L2 learners exhibit signs of transfer,

as results of studies by Gillon Dowens, et al. (2010 and 2011) suggest. Even when

processing of mismatches in both gender and number is quite nativelike, L2 learners still

exhibit a delay in the onset of slowed reaction time to gender but not number errors,

possibly due to the influence of their previous experience with number agreement in

English. One major limitation of this study was the failure to distinguish between L2

speakers based on proficiency. Though all participants were of relatively high proficiency,

likely equivalent to the "advanced" level described by many other investigators of second

language acquisition, this study was not able to separate L2 learners into discrete,

statistically robust groups based on proficiency because proficiency level was based on the

results of a questionnaire rather than a more standardized method. Therefore, as noted by

Jegerski, Van Patten, & Keating (2011), it is quite possible that L2 learners at this level

could attain a more nativelike level of gender mismatch processing that would bring their

performance more into line with the native speakers. In addition, although the task at hand

relied on online performance and did not directly test the L2 learners' grammars as others

have done, such a consideration should not affect the reliability of the results, as Jegerski, et

al., found that shortcomings in subjects' performance on online tasks does not entail

reduced grammatical abilities. Although this observation does not ensure participants were

adequately advanced that deficiencies in their grammars would not have an effect on their performance, the responses to the background questionnaire showed that participants were sufficiently proficient for the requirements of the supplied task.

Although the results of this study do point to the specific conclusions discussed above, the lack of statistical significance in certain areas suggests that the inquiry into the subject of gender and number processing by second language learners is far from over. Future directions should include running a similar experiment with a much larger sample size, especially to determine the true nature of both groups' reactions to mismatches in the window following the noun. Furthermore, these participants should be more rigorously quantified with regard to their proficiency levels, and a similar procedure should be conducted with near-native speakers as well as the advanced group seen here.

This experiment dealt only with participants' reactions to mismatches between determiners and nouns and did not deal with reactions to mismatches between nouns and adjectives. Although studies have examined noun-adjective mismatches in the past, future research should be conducted comparing the two relationships directly to determine whether speakers behave differently depending on what type of constituent disagrees with the noun. In addition, this experiment did not attempt to separate nouns with regular, i.e. canonical and transparent, gender from those with irregular gender inflection. It is possible that speakers react differently to regular and irregular gender, which could have skewed the results of the study. Future studies should consider this aspect of Spanish gender carefully, especially those that rely on grammatical judgment tasks or the like.

Conclusion

Based on the results of this experiment, I conclude that advanced late L2 learners who have not achieved ultimate attainment are still able to process determiner-noun gender mismatches in nearly the same way as native speakers. The delayed reaction time representing nativelike processing indicates, however, that L2 speakers are slower to process these mismatches. Therefore, while able to correctly analyze gender mismatches in an online task, these speakers cannot do so as quickly as native speakers, indicating that they have not reached the same proficiency levels that are exhibited by near-native L2 speakers. Furthermore, their delayed performance on gender mismatch reanalysis contrasts with their immediate performance on number mismatch. This disjunction between L2 speakers' reanalysis of these two types of syntactic mismatches is explained by transfer effects in accordance to psycholinguistic accounts by Barber & Carreiras (2005) and contra Clahsen & Felser (2006)'s Shallow Structure Hypothesis.

References

Alarcón, I. V. (2011). Spanish gender agreement under complete and incomplete

   acquisition: Early and late bilinguals' linguistic behavior within the noun phrase.

   Bilingualism, 14(3), 332-350.

Anderson, S. R. (1992). *A-morphous morphology*. Cambridge, England; New York:

   Cambridge Univ Press.

Barber, H., & Carreiras, M. (2005). Grammatical gender and number agreement in Spanish:

   An ERP comparison. *Journal of Cognitive Neuroscience, 17*(1), 137-153.

Bialystok, E. (1997). The structure of age: In search of barriers to second language

   acquisition. *Second Language Research, 13*(2), 116-137. doi: 10.1191/

   026765897677670241

Brennan, J., Nir, Y., Hasson, U., Malach, R., Heeger, D. J., & Pylkkanen, L. (2012). Syntactic

   structure building in the anterior temporal lobe during natural story listening. *Brain

   and Language, 120*(2), 163-173. Retrieved from

   http://search.proquest.com/docview/1018384030?accountid=14512

Carminati, M. N. *The processing of Italian subject pronouns. Dissertation Abstracts

   International, A: The Humanities and Social Sciences,* Retrieved from

   http://search.proquest.com/docview/85570011?accountid=14512. (85570011;

   200304533).

Carroll, S. E. (2006). Shallow processing: A consequence of bilingualism or second language

   learning? *Applied Psycholinguistics, 27*(01), 53. doi: 10.1017/S0142716406060061

Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied

   Psycholinguistics, 27*(1), 3-42.

Corbett, G. G. (1991). *Gender*. Cambridge: Cambridge Univ Press.

de Saussure, F. (1916). In C. Bally and A. Sechehaye, with the collaboration of A. Riedlinger

(Ed.), *Cours de linguistique générale* [ ]. Lausanne and Paris: Payot.

Dowens, M. G., Vergara, M., Barber, H. A., & Carreiras, M. (2010). Morphosyntactic

processing in late second-language learners. *Journal of Cognitive Neuroscience,*

22(8), 1870-1887.

Dowens, M. G., & Carreiras, M. (2006). The shallow structure hypothesis of second language

sentence processing: What is restricted and why? *Applied Psycholinguistics, 27*(01),

49. doi:10.1017/S014271640606005X

Dowens, M. G., Guo, T., Guo, J., Barber, H., & Carreiras, M. (2011). Gender and number

processing in Chinese learners of Spanish-evidence from event related potentials.

Neuropsychologia, 49(7), 1651-1659.

Eubank, L. (1993). On the transfer of parametric values in L2 development. *Language*

Acquisition, 3(3), pp. 183-208. Retrieved from http://www.jstor.org/stable/

20011400

Faussart, C., Jakubowicz, C., & Costes, M. (1999). Gender and number processing in spoken

French and Spanish. *Rivista Di Linguistica, 11*, 75-101.

Felser, C., Roberts, L., Marinis, T., & Gross, R. (2003). The processing of ambiguous

sentences by first and second language learners of English. *Applied Psycholinguistics,*

24(3), 453-489. Retrieved from http://search.proquest.com/docview/

85594953?accountid=14512

Foote, R. (2010). Age of acquisition and proficiency as factors in language production:

Agreement in bilinguals. *Bilingualism, 13*(2), 99-118.

Friederici, A. D. (2001). Syntactic, prosodic, and semantic processes in the brain: Evidence

from event-related neuroimaging. *Journal of Psycholinguistic Research, 30*(3), 237-

250.

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.

Hahne, A. (2001). What's different in second-language processing? Evidence from event-

related brain potentials. *Journal of Psycholinguistic Research,* (30), 251-66.

Hawkins, R., & Chan, C. Y. -. (1997). The partial availability of universal grammar in second

language acquisition: The 'failed functional features hypothesis'. *Second Language

Research,13*(3), 187-226.

Herschensohn, J., Stevenson, J., & Waltmunson, J. (2005). Children's acquisition of L2

Spanish morphosyntax in an immersion setting. *IRAL - International Review of

Applied Linguistics in Language Teachings, 43*(3), 193-217.

Hopp, H. (2006). Syntactic features and reanalysis in near-native processing. *Second

Language Research, 22*(3), 369-397.

Jegerski, J., VanPatten, B., & Keating, G. D. (2011). Cross-linguistic variation and the

acquisition of pronominal reference in L2 Spanish. *Second Language Research, 27*(4),

481-507. Retrieved from http://search.proquest.com/docview/

908017995?accountid=15172

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning:

The influence of maturational state on the acquisition of English as a second

language. *Cognitive Psychology, 21*(1), 60-99. Retrieved from

http://search.proquest.com/docview/58191526?accountid=14512

Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading

comprehension. *Journal of Experimental Psychology: General, 111*(2), 228-238.

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.

Lenneberg, E. (1967). *Biological foundations of language*. New York: Wiley.

Montrul, S. (2011). Morphological errors in Spanish second language learners and heritage

speakers. *Studies in Second Language Acquisition, 33*(2), 163-192.

Papadopoulou, D., & Clahsen, H. (2003). Parsing strategies in L1 and L2 sentence

processing: A study of relative clause attachment in Greek. *Studies in Second

Language Acquisition, 25*(4), 501-528. doi: http://dx.doi.org/

10.1017.S0272263103000214

Peelle, J. E., Troiani, V., & Grossman, M. (2009). Interaction between process and content in

semantic memory: An fMRI study of noun feature knowledge. *Neuropsychologia,

47*(4), 995-1003. Retrieved from http://search.proquest.com/docview/

85628801?accountid=14512

Penfield, W., & Roberts, L. (1959). *Speech and brain mechanisms*. Princeton: Princeton

University Press.

Perception Research Systems Incorporated. (2012). *Paradigm*

Ritter, E. (1993). Where is gender? *Linguistic Inquiry, 24*, 795-803.

Rothman, J., & Iverson, M. (2007). The syntax of pronominal subjects in L2 Spanish:

      Comparing two L2 populations with different exposure. *Revista Española De*

      Lingüística Aplicada, (20), 185-214.

Rüschemeyer, S. -., Fiebach, C. J., Kempe, V., & Friederici, A. D. (2005). Processing lexical

      semantic and syntactic information in first and second language: FMRI evidence

      from German and Russian. *Human Brain Mapping, 25*(2), 266-286.

Sagarra, N., & Herschensohn, J. (2010). The role of proficiency and working memory in

      gender and number agreement processing in L1 and L2 Spanish. *Lingua, 120*(8),

      2022-2039.

Sagarra, N., & Herschensohn, J. (2011). Proficiency and animacy effects on L2 gender

      agreement processes during comprehension. *Language Learning, 61*(1), 80-116.

      doi:10.1111/j.1467-9922.2010.00588.x

Schwartz, B. D., & Sprouse, R. A. (1996). L2 cognitive states and the full Transfer/Full access

      model. *Second Language Research, 12*(1), 40-72. Retrieved from

      http://search.proquest.com/docview/58331647?accountid=15172

Sorace, A. (2006). Possible manifestations of shallow processing in advanced second

      language speakers. *Applied Psycholinguistics, 27*(01), 88.

      doi:10.1017/S0142716406060164

Sorace, A., & Filiaci, F. (2006). Anaphora resolution in near-native speakers of Italian.

      Second Language Research, 22(3), 339-368. Retrieved from

      http://search.proquest.com/docview/85638037?accountid=14512

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*(2), 105-

122. Retrieved from http://search.proquest.com/docview/

85571343?accountid=14512

Vainikka, A., & Young-Scholten, M. (1994). Direct access to X'-theory: Evidence from Korean

and Turkish adults learning German. In T. Hoekstra and B.D. Schwartz (Ed.),

Language acquisition studies in generative grammar: Papers in honor of Kenneth

Wexler from the 1991 GLOW workshops (pp. 265-316). Philadelphia: John Benjamins.

Vigliocco, G., Butterworth, B., & Garrett, M. F. (1996). Subject–verb in Spanish and English:

Differences in the role of conceptual factors. *Cognition, 51*, 261-298.

Weber-Fox, C., & Neville, H. J. (1996). Maturational constraints on functional specializations

for language processing: ERP and behavioral evidence in bilingual speakers. *Journal*

of Cognitive Neuroscience, 8, 231-256.

White, L., Valenzuela, E., Kozlowska-Macgregor, M., & Leung, Y. -. I. (2004). Gender and

number agreement in nonnative Spanish. *Applied Psycholinguistics, 25*(1), 105-133.

White, L., & Genesee, F. (1996). How native is near-native? The issue of ultimate attainment

in adult second language acquisition. *Second Language Research, 12*(3), 233-265. Retrieved

from http://search.proquest.com/docview/85638794?accountid=15172

Appendix

A. Language Description

Spanish contains both grammatical gender and number, both of which are independently bivariate.[6] Both gender and number in Spanish are canonically marked as a verbal suffix on the ends of nouns and any adjectives or determiners that agree with the nouns in question. Canonically, masculine marking appears as the suffix –*o*, while feminine marking appears as the suffix –*a*. Similarly, singular nouns are unmarked (or marked with a null suffix), while plural nouns are marked with the suffix –*s* following any gender marking, as in Table A1, below. Although some nouns may not be overtly marked for gender, all nouns have grammatical gender which must agree with modifying adjectives and determiners, as in item 3, below.

Table A1. Spanish gender and number marking paradigm.

1. *El*   *tornill-o*   *L-a*   *botell-a*
   DET.M.SG screw-M.SG   DET-F.SG bottle-F.SG

2. *El*   *hombre*   *L-a*   *mujer*
   DET.M.SG man.M.SG   DET-F.SG woman.F.SG

3. *L-os*   *tornill-os*   *L-as*   *botell-as*
   DET.M.PL screw-M.PL   DET-F.PL bottle-F.PL

Note that gender may or may not correspond to actual, biological gender. For example, *perr-o*, dog-M.SG, may denote a male or female dog (if the sex of the dog is not specified), but *perr-a*, dog-F.SG, can only denote a female dog.

---

[6] Cf. German, which has masculine-singular, feminine-singular, neuter-singular, and the genderless plural.