

# Gender and number agreement in nonnative Spanish

LYDIA WHITE, ELENA VALENZUELA, MARTYNA  
KOZLOWSKA–MACGREGOR, and YAN-KIT INGRID LEUNG  
*McGill University*

## ADDRESS FOR CORRESPONDENCE

Lydia White, McGill University, Department of Linguistics, 1085 Dr. Penfield Avenue, Montreal,  
Quebec H3A 1A7, Canada. E-mail: lydia.white@mcgill.ca

## ABSTRACT

This paper reports on an experiment investigating the acquisition of Spanish, a language that has a gender feature for nouns and gender agreement for determiners and adjectives, by speakers of a first language (L1) that also has gender (French), as well as an L1 that does not (English). Number (present in all three languages) is also investigated. Subjects were adult learners of Spanish, at three levels of proficiency, as well as a control group of native speakers. Oral production data were elicited. Subjects were also tested on an interpretation task, in which the selection of pictures corresponding to particular sentences depends on number and gender contrasts. The results from both tasks show significant effects for proficiency; low proficiency groups differ significantly from native speakers, but advanced and intermediate groups do not. There were no significant effects for L1 or for prior exposure to another second language with gender. The findings are discussed in the context of two different theories as to the possibility of parameter resetting in nonnative acquisition, namely, the failed functional features hypothesis and the full transfer full access hypothesis. The results are consistent with the latter hypothesis.

The status of universal grammar (UG) in second language (L2) acquisition has been the subject of extensive debate, much of it centering on the question of whether L2 learners can reset the parameters of UG (see White, 1989, 2003, for overview). Researchers have proposed the no parameter resetting hypothesis (Clahsen & Muysken, 1989), the idea being that learners have access to principles of UG via the first language (L1) but are unable to acquire new parameter values. According to this kind of account, the L2 learner has recourse only to those parameter settings exemplified in the L1; no subsequent parameter resetting is possible in response to L2 input. This position contrasts with so-called full access theories, according to which learners are able to set parameters to L2 values in response to L2 input (e.g., Duffield & White, 1999; Epstein, Flynn, & Martohardjono, 1996; Schwartz & Sprouse, 1994, 1996; White, 1989, 2003). Under current proposals, parametric differences between grammars are associated with functional categories (complementizer, agreement, tense, determiner [Det], number [Num], etc.), and their features (such as tense, number, person, gender, and case; Borer, 1984; Chomsky, 1995; Ouhalla, 1991; Pollock, 1989). In this paper, we examine the issue of parameter resetting in the context of gender and number features and

their agreement properties. We investigate the acquisition of Spanish, a language that has a gender feature for nouns and gender agreement for determiners and adjectives, by speakers of an L1 that also has gender (French), as well as an L1 that does not (English). We argue that gender features are acquirable even when they are absent in the L1, supporting parameter resetting.

## TWO HYPOTHESES ON PARAMETER RESETTING

Recent versions of the no parameter resetting hypothesis are couched in terms of functional categories, their features, and feature strength. Smith and Tsimpli (1995, p. 24), for example, make the following claim:

Parameterization is . . . defined in terms of a finite set of alternative values that a functional category can be associated with. Cross-linguistic variation is thus restricted to differences in the parametric values of functional categories. . . . Moreover, if we assume that the critical period hypothesis is correct, maturational constraints on the functional module can be interpreted as entailing its complete inaccessibility after the end of this period. . . . UG may still be available but parameter-setting can not [*sic*] be.

Following this line of reasoning, Hawkins and Chan (1997) propose the failed functional features hypothesis (FFFH), according to which postpuberty learners are unable to acquire abstract grammatical features that differ from those found in the L1. A kind of representational deficit is implicated. In other words, it is impossible, according to this account, for the interlanguage grammar to represent functional features that are not instantiated in the speaker's mother tongue grammar. Hence, a nativelike mental representation will necessarily be unattainable in those cases where the L1 and L2 differ in parameter values. Furthermore, the implication is that any nonnative language acquisition by adults, not just L2 acquisition, would be affected in this way; that is, when learning a third or fourth language, the crucial predictor of success is whether the features in question are represented in the L1.

When proposing the FFFH, Hawkins and Chan (1997) originally looked at syntactic consequences of the purported absence of a ( $\pm wh$ ) feature in the interlanguage grammar of Chinese-speaking learners of English. Subsequently, Hawkins and colleagues extended the FFFH to abstract features like tense [ $\pm past$ ] and gender, claiming that the failure of adult learners to consistently supply overt tense morphology or gender agreement is a consequence of the absence of the corresponding features in the L1 grammar (Franceschina, 2001; Hawkins, 1998, 2001; Hawkins & Franceschina, in press; Hawkins & Liszka, in press).

The FFFH contrasts with the full transfer full access (FTFA) hypothesis (Schwartz & Sprouse, 1994, 1996; see also White, 1989), which shares the assumption of the FFFH that grammatical features in the initial interlanguage representation will be drawn from the set realized in the L1. However, FTFA differs from FFFH in postulating the acquirability, regardless of age of acquisition, of new features that are not present in the L1 but are required to adequately represent the L2. On this view, functional features do not fail in nonnative acquisition; rather, L2 learners still have access to the full set of abstract features made available by UG. Consequently, interlanguage grammars are not restricted to L1 parameter settings and nativelike mental representations are in principle acquirable.

One question of relevance here relates to the issue of classroom versus naturalistic L2 acquisition. Gender is, typically, explicitly taught in Spanish classes, and it is relatively transparent, given the phonological forms most commonly associated with gender, namely, *-o* and *-a*. Thus, in the event of successful performance by instructed learners on measures of L2 gender, it might be claimed that some kind of explicit or prescriptive knowledge is involved rather than an abstract underlying representation, in which case an examination of gender assignment in instructed nonnative Spanish would not provide a suitable means to test the claims of FTFA and FFFH.

In fact, both theories assume that, even though input may differ in classroom and naturalistic contexts, the underlying mechanisms are similar (see Hawkins, 2001, pp. 18–22, for discussion). The FFFH appears to predict L1-based deficits even in cases involving classroom instruction (Franceschina, 2001; Hawkins & Franceschina, in press). In consequence, we assume, without further discussion, that performance on Spanish gender by instructed learners provides an appropriate means of investigating the claims of FTFA and FFFH. (Furthermore, resorting to explanations involving other kinds of learning renders such theories unfalsifiable, in the absence of a precise specification of what the alternatives might be.)

#### SYNTACTIC BACKGROUND: GENDER, NUMBER, AND N-DROP

In Romance languages, including Spanish and French, all nouns are classified in terms of grammatical gender, which is arbitrary and distinct from natural gender (although the two may coincide). Nouns are divided into two classes, masculine and feminine. Gender is an inherent lexical feature on nouns (e.g., Carroll, 1989; Corbett, 1991; Cressey, 1978; Stockwell, Bowen, & Martin, 1965), whereas adjectives and determiners show gender agreement with the head noun. Many nouns in Spanish show distinct endings, with *-o* usually indicating masculine nouns and *-a* usually indicating feminine. However, as is well known, this is not a one to one correspondence: there are a significant number of masculine nouns ending in *-a* and a very small number of feminine nouns ending in *-o*. There are also nouns with no overt gender marking at all. Similarly, many adjectives show gender agreement by means of the same endings as those found on nouns, namely, *-o* and *-a*, but adjectives lacking overt gender agreement are also common. According to Harris (1991), *-o* and *-a* are word markers, rather than gender markers, because they are not confined to lexical items that have gender, being found also on adverbs. However, for the purposes of the present discussion, we will treat them as gender markers, recognizing that this is an oversimplification.

These properties are illustrated in Example 1. In Examples 1a and 1b, the head nouns *sombrero* and *chaqueta* are masculine (M) and feminine (F), respectively, taking the standard *-o* and *-a* endings, whereas the forms of the determiner and adjective vary according to the gender of the head. In Example 1c the head noun *clase* is feminine but does not end in the usual *-a* and the form of the adjective *difícil* is invariant. In this case, the only indication that the noun is feminine is the occurrence of the feminine form of the article. In the following, *s* indicates singular.

1. a. el        sombrero negro  
      the-MS hat-MS    black-MS  
      “the black hat”  
   b. la        chaqueta negra  
      the-FS jacket-FS black-FS  
      “the black jacket”  
   c. la        clase difícil  
      the-FS class difficult  
      “the difficult class”

English has natural gender, which shows up typically in choice of pronouns (*he* vs. *she*, etc.), but it lacks noun classes based on grammatical gender; hence, determiners and adjectives do not show gender agreement with the head noun within a determiner phrase (DP).

Whereas Romance languages and English differ as to the status of gender, they both have number as a grammatical feature within the DP. In the case of English, nouns are marked for plural, typically by means of the morpheme *-s/*; there is limited number agreement within the DP in the case of certain determiners, as shown in Example 2. Adjectives in English do not agree in number.

2. a. this black hat  
   b. these black hats

Spanish also indicates plural (P) number by means of *-s/*. Adjectives and determiners show number agreement with the head noun, as can be seen in Example 3.

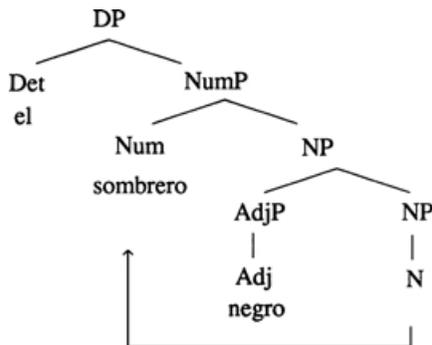
3. a. los        sombreros negros  
      the-MP hat-MP    black-MP  
      “the black hats”  
   b. las        chaquetas negras  
      the-FP jacket-FP black-FP  
      “the black jackets”

According to current analyses (e.g., Chomsky, 1995, 2001), gender and number are  $\phi$  features (agreement features) that are found on the head noun and enter into a checking relationship with corresponding features elsewhere in the structure. In fact, as Carstens (2000) and others (e.g., Bosque & Picallo, 1996; Mallen, 1997) have shown, a number of technical modifications have to be made to Chomsky's (1995) checking theory in order to account for concord within the DP. The  $\phi$  features of N are said to be “interpretable”; that is, they include information that is required for semantic interpretation. The corresponding features of determiners and adjectives are “uninterpretable” and have to be deleted (by means of feature checking) in the course of the derivation. However, it seems somewhat counter-intuitive to consider grammatical gender on nouns as interpretable in this sense, because gender is arbitrary and does not affect interpretation at all. See Carstens (2000; n. 12) for relevant remarks.

Gender is parameterized; it is realized in some languages but not in others. According to the FFFH, it is uninterpretable features that are affected in non-native acquisition: the claim is that new uninterpretable features cannot be acquired, although new interpretable features can be (Hawkins, 1998; Hawkins, & Franceschina, in press).<sup>1</sup> Thus, the breakdown in L2 acquisition will manifest itself in agreement or checking relations: in the absence of uninterpretable features, checking cannot take place. As a consequence, gender on nouns can be acquired in isolation but gender agreement cannot.

We adopt the analysis of DPs shown in Example 4, which includes a functional category Num located between Det and NP, where number features are found (Bernstein, 1993; Ritter, 1993; Valois, 1991).<sup>2</sup> Nouns raise from N to Num overtly in Romance languages (due to strong features in Num) and covertly in English (due to weak features); this yields the well-known differences in Romance languages between positions of the adjective (Adj) with respect to the head N, prenominal in English, postnominal in Romance (compare Examples 1 and 2). Following Carstens (2000), we assume that concord within the DP proceeds as follows. As the N raises, it enters into a “close enough” relationship (head/head or specifier/head) with the adjective and the determiner for its interpretable features to check the corresponding uninterpretable features.<sup>3</sup>

4.



An additional property of Spanish is important for our study, namely, the fact that the head noun in a Spanish DP does not have to be overtly realized, provided that its content can be in some way recovered from the context. Such DPs are referred to as null nominals; the phenomenon is also known as noun-drop (N-drop). This is illustrated in Example 5. In Example 5a the N, *muñeca* (doll), is overt. In Example 5b *muñeca* has been omitted. N-drop is licensed by gender and number features on determiners and adjectives within the DP (Bernstein, 1993; Licerias, Diaz, & Mongeon, 1999; Snyder, 1995), which allow the gender and number of the null noun to be recovered. In Example 5b the missing noun is identified as feminine and singular by means of the feminine singular forms of the indefinite article (*una*) and the adjective (*pequeña*).

5. a. Una muñeca pequeña está encima de la cama.  
a-FS doll-FS small-FS is on-top of the bed  
“There is a small doll on the bed.”

- b. Una pequeña está encima de la cama.  
a-FS small-FS is on-top of the bed
- c. “There is a small on the bed.”
- d. “There is a small one on the bed.”

If there is a breakdown in agreement in the Spanish interlanguage grammar due to failure to acquire uninterpretable gender features in Det and Adj, as FFFH claims, it should be impossible for nonnative speakers to accurately interpret structures involving N-drop.

N-drop is highly productive in Spanish; it is also found in French, although to a more limited extent. N-drop occurs in English but is highly restricted (expressions like *the rich*, *the poor*, as well as with color terms, e.g., *I prefer the blue*), but it is not productive (see Example 5c). Rather, the proform *one* is required, as can be seen in Example 5d.

## PREVIOUS RESEARCH ON L2 GENDER

There have been a number of recent studies on Romance L2 gender, some conducted within the generative paradigm and some within other frameworks. Carroll (1989) argues that features not instantiated in the L1 “atrophy,” a position that can be seen as a precursor of the FFFH. However, Carroll’s proposal differs from the FFFH in that, for her, it is the possibility of representing a gender feature of nouns that is determined by presence or absence of gender in the L1. In other words, the problem is seen as a problem in acquiring the inherent, interpretable feature of the noun, which will have consequences for agreement as well.

With specific reference to the issue of parameter resetting and the status of gender features in the interlanguage grammar, Hawkins (1998; as cited in Hawkins, 2001) examined production data elicited from 20 advanced learners of French whose mother tongue was English. While his subjects proved relatively accurate on gender agreement between determiners and nouns overall,<sup>4</sup> they did exhibit persistent problems: (a) showing greater accuracy with gender agreement on definite determiners than on indefinite and (b) adopting a “default” gender on determiners, leading to overuse of one or the other gender; some subjects used masculine as their default whereas others used feminine. In accordance with the FFFH, Hawkins attributes these problems to the lack of uninterpretable gender features in the L1, English, and consequently in the interlanguage grammar. As a result, feature checking cannot take place in their L2 French (there being no uninterpretable features to check) and agreement does not surface.

In the same vein, Franceschina (2001) examined spontaneous production data from one adult L2 Spanish speaker (L1 English), Martin, who has had extensive exposure to Spanish, living in South America for many years and having been married to a Spanish native speaker. Gender errors are found on 7% of all adjectives that are produced and on 8.3% of all articles that are produced. This contrasts with number errors on the same categories, which were 2 and 0.5%, respectively. The majority of gender errors involve using masculine forms where feminine forms would be required. Franceschina maintains that a gender feature must be present on nouns in Martin’s grammar, because he invariably produces nouns with correct

-o and -a endings; at the same time he shows a breakdown in agreement because of a breakdown in feature checking, due to the absence of uninterpretable gender features on adjectives and determiners in the L1 English. Hawkins (1998) makes a similar claim with respect to data reported by Andersen (1984) from Anthony, an English-speaking 12-year-old acquiring L2 Spanish. Anthony makes few errors on the form of nouns (i.e., they end correctly in -o or -a) but many gender agreement errors as far as determiners are concerned, because he basically adopts one form of the determiner as the default and uses this invariantly.

The argument that the inherent gender feature on nouns must be present in such cases is misconceived. Accuracy on noun forms in isolation is meaningless. Independent of the agreement facts, there is no certain way to tell whether learners have acquired gender on the noun itself. Franceschina (2001) and Hawkins (1998) seem to think that if nonnative speakers of Spanish use the same noun endings (-o or -a) for particular lexical items as native speakers do, this must mean that they have the appropriate inherent gender feature on the N. However, of course, this is not necessarily the case: it simply shows that they have acquired the appropriate phonological shape of the words in question (and possibly that they have also acquired word markers, in the sense of Harris, 1991). The fact that a noun in the interlanguage appropriately ends in -o or -a does not entail that the noun is represented in the mental lexicon with inherent and interpretable gender features.

Hawkins (2001) and Franceschina (2001) do not compare learners of L1s with and without gender. Bruhn de Garavito and White (2002) show that problems similar to those reported by Hawkins (1998, 2001) for English speakers also occur in the acquisition of Spanish by French speakers, both languages with gender, which suggests that the absence of gender in the L1 is not the only factor affecting L2 gender acquisition. Researchers working in different frameworks have reported similar results from learners of a variety of L1s with and without gender, suggesting that whatever is going on is not simply an L1 effect (e.g., Bartning, 2000; Dewaele & Véronique, 2001; Fernández-García, 1999).

This is not to deny L1 effects altogether: these clearly are found in gender acquisition (see, for example, Sabourin, 2001). However, the central claim of the FFFH is that, in the absence of uninterpretable gender features in the L1, it will be impossible to represent gender agreement in the interlanguage grammar. Relevant to this hypothesis is a study by Gess and Herschensohn (2001) who look at gender, among other properties of DPs. Subjects were English-speaking learners of French, at several levels of proficiency. Learners at more advanced levels achieved a high degree of accuracy on gender and number agreement in a written sentence-completion task, suggesting that learners can acquire features that are absent in the L1. Learners at lower proficiency levels were quite inaccurate on agreement, suggesting initial L1 effects. However, as Gess and Herschensohn do not separate gender and number agreement in their presentation of results, it is impossible to determine whether the problem at lower levels of proficiency is restricted to gender. Furthermore, their task was quite metalinguistic, so that the results may not in fact reflect underlying and unconscious knowledge of gender.

Difficulties in acquisition of L2 gender have largely been reported in the case of spontaneous production data. Carroll (1989, p. 576) points out that a theory that assumes atrophy or unavailability of gender features predicts poor performance

across a variety of tasks. FFFH would appear to make the same prediction: if gender agreement is disrupted because of the unavailability of uninterpretable features, then it should be impossible regardless of task. In other words, a representational deficit implies across the board effects.

In conclusion, previous research on gender features in nonnative grammars does not adequately address the contrasting claims of FTFA and FFFH. In our study, we examine and compare production and comprehension data from French and English speakers learning Spanish, as an L2 or third language (L3), at various levels of proficiency, in order to determine whether it is indeed the case that English speakers are unable to represent gender features, with their associated agreement properties, in the interlanguage grammar. We contrast gender features with number features, which are present in all three languages under consideration. In addition, we investigate gender in comprehension as well as production. For production, the issue is whether, when learners produce a noun of a particular gender, they also produce appropriate agreement on the corresponding determiner and adjective. In the case of interpretation, the issue is whether, given a determiner and adjective of a particular gender, learners select a noun from their interlanguage lexicon that corresponds in gender.

## HYPOTHESES AND PREDICTIONS

In this paper we challenge the claim of Hawkins and colleagues that gender features cannot be acquired in postpuberty acquisition in the absence of an L1 with gender. Instead, in accordance with FTFA, we hypothesize that the grammars of nonnative speakers are not restricted to uninterpretable formal features found in the L1. In other words, interlanguage grammars are not defective when it comes to representation of parameterized features like gender; hence, associated properties, such as agreement within the DP, are acquirable.

Although both hypotheses predict transfer of features from the L1 in initial stages of nonnative acquisition, only FTFA predicts change in the developing interlanguage grammar with respect to grammatical features that are present in the L2 but not the L1. According to the FFFH, English and French speakers should behave differently from each other at all stages of development, with respect to Spanish.<sup>5</sup> Only the French speakers should show evidence of acquiring uninterpretable gender features. According to FTFA, in contrast, English speakers and French speakers will indeed start off with different initial states for Spanish. Thus, differences are expected initially in accuracy of gender agreement as the early interlanguage grammar of French speakers will represent gender whereas the early interlanguage grammar of English speakers will fail to do so. Nevertheless, at some point, Spanish gender features (both interpretable and uninterpretable) should be represented in the interlanguage grammar, regardless of presence or absence of gender in the L1. Thus, advanced nonnative speakers will exhibit knowledge of Spanish gender, including appropriate gender agreement throughout the DP. Hence, the behavior of both L1 groups at later stages with respect to realization of gender should be similar.

Finally, when comparing the features number and gender, the FFFH necessarily predicts superior performance on number (because this is a feature found in the

grammars of English, French, and Spanish). FTFA, on the other hand, does not predict any necessary long-term differences between these features.

To summarize, our predictions for production and interpretation of Spanish DPs are as follows:

1. a. At lower proficiency levels, learners whose L1 is English will perform more accurately on number agreement than on gender agreement.
- b. At lower proficiency levels, learners whose L1 is English will have greater difficulty with gender agreement than learners whose L1 is French.

FFFH would make similar predictions: L1 effects are necessarily expected in this view. Hence, performance on number should necessarily be better than on gender in the case of English speakers, and they should necessarily perform worse on gender than French speakers.

The major difference between the two approaches is in the predictions for later stages, certainly for advanced proficiency speakers and possibly for intermediate-level proficiency as well. In the case of learners whose L1 does not have gender, we predict:

2. a. Advanced learners should perform as well on gender as on number.<sup>6</sup>
- b. Advanced learners should perform accurately on gender agreement, regardless of the status of gender features in the L1.

In contrast, FFFH predicts persistent problems for postpuberty learners whose L1 lacks gender, even at advanced levels of proficiency.

## EXPERIMENT

### *Subjects*

As previously mentioned, the nonnative speakers of Spanish tested in this study were native speakers of French ( $n = 48$ ) and English ( $n = 68$ ). At the time of testing, they were taking Spanish courses as adults in French-speaking and English-speaking university settings in Canada. Most subjects were first exposed to Spanish in their mid to late teens. In other words, they were postpuberty learners, presumably past any critical period. For some of the English speakers ( $n = 14$ ), Spanish was an L2, and for others ( $n = 54$ ) it was an L3, the L2 being French, another language with grammatical gender; this issue will be addressed in the analyses below. (Recall that the FFFH does not distinguish between L2 and L3 acquisition as far as its predictions are concerned: any case of nonnative acquisition will be problematic if the L1 lacks gender and will not be problematic if the L1 has gender.) People who reported having been brought up bilingual (where either language had gender) were excluded, as was anyone who reported early exposure to Spanish, Italian, or Portuguese. In addition, a few subjects were eliminated on the basis of their performance on the vocabulary task (failing to respond to more than 66% of test items); subjects who did not complete all tasks were also eliminated.

Table 1. *L2 subjects by L1 and proficiency level*

	L1 English	L1 French
Low	34	12
Intermediate	24	20
Advanced	10	16

Subjects were given a Spanish proficiency test, consisting of the reading/vocabulary section of the MLA Cooperative Foreign Language Test (Educational Testing Service, Princeton, NJ) and a cloze test from the Diploma de Español como Lengua Extranjera (Spanish Embassy, Washington, DC). On the basis of results from this test, they were divided into three proficiency levels: low, intermediate, and advanced. Table 1 reports the experimental groups retained in the present study, according to L1 and proficiency level. Twenty native speakers of Spanish served as a control group.

### Tasks

Four tasks were devised for this experiment: two elicited production tasks, a vocabulary test, and a picture identification task. The vocabulary task was always taken after the picture identification task in order to avoid priming of gender cues. Subjects were tested individually.

*Production.* Both production tasks were designed to elicit DPs containing adjectives. The first task involved an adaptation of the “Guess Who” game. The subject and the experimenter each had an identical set of cards showing various characters (male and female, with and without hats, beards, glasses, different colors of hair, different ages, etc.). The experimenter would choose one character and the subject had to guess which person the experimenter had chosen by asking questions involving descriptions, such as those in Example 6, to which the experimenter would only respond with *yes* or *no*. Each subject played the game three times, asking an average of seven questions each time.

6. a. ¿Es un chico muy viejo? (English L1, low proficiency)  
Is a-M guy-M very old-M  
“Is it a very old guy?”  
b. ¿Tiene barba roja? (English L1, advanced proficiency)  
Have beard-F red-F  
“Does he have a red beard?”

The second production task involved picture description. Subjects were shown three pictures (one after the other) and asked to describe what was going on in them. Both tasks were taped. The data were subsequently transcribed and checked by two native speakers of Spanish and coded for gender and number agreement, as well as adjective placement.

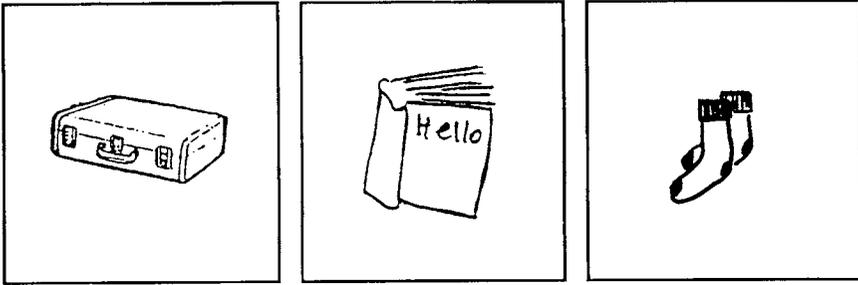


Figure 1. Sample items from the picture identification task.

*Comprehension: Picture identification.* The comprehension task, specifically picture identification, takes advantage of the existence of null nominals in Spanish. Appropriate interpretation of a null nominal crucially depends on gender and number features realized overtly on adjectives and/or determiners within the DP. In consequence, if there is a breakdown in gender agreement due to unavailability of uninterpretable features, as the FFFH claims, expressions involving null nominals identified by gender should be interpreted inconsistently by English-speaking learners of Spanish at any level of proficiency.

The picture identification task consisted of a booklet (which subjects read), containing a story about two characters preparing to go on vacation. The story included 48 sentences, 16 of which were distractors, 14 of which targeted number, and 18 of which targeted gender. Each sentence appeared with three pictures immediately below it, all of which were equally plausible in the context of the story. Each test sentence included a null nominal, in other words, a DP consisting of a determiner and an adjective but no overt noun. (The distractors did not include null nominals.) Subjects had to indicate which of the three pictures was the appropriate one for any given sentence by circling the appropriate response. One picture in each triplet would be the targeted one, showing an item whose number or gender corresponded to that of the missing nominal. One picture would show an item with the opposite number or gender (holding gender constant in the case of items testing number and number constant in the case of items testing gender). If learners do not have gender agreement, they should pick randomly between the correct picture and its opposite. The third picture in each set was a foil.

Consider Figure 1, which is an item testing gender agreement. The three pictures (all colored red in the actual test) show a suitcase, a book, and a pair of socks. Here, the stimulus sentence is *María contesta: "Sí, claro, va a hacer mucho sol. Ponlas ahí cerca de la roja."* (Maria answers: "Yes, of course, it is going to be very sunny. Put them over there by the red [one]"). The phrase *la roja* contains a null nominal, which is feminine and singular, as shown by the determiner *la* and the adjective *roja* (red). If learners have gender agreement percolating through the DP, they should pick the picture of the suitcase (*la maleta*), whose gender is feminine. The book (*el libro*) should not be selected, since it is masculine. The foil is provided by the socks (*los calcetines*) which are masculine and plural, in this case differing on both features.<sup>7</sup>

Table 2. *Production data on number of DPs produced*

L1	Proficiency	Det N	(Det) N Adj	*(Det) Adj N
Native speakers		1951	706	0
L1 English	Advanced	770	211	0
	Intermediate	1714	409	0
	Low	1989	514	6
L1 French	Advanced	1644	429	0
	Intermediate	1703	398	31
	Low	744	123	39

It is important to note that the missing vocabulary items were not mentioned anywhere in the picture identification task, occurring only in the form of pictures. Because the gender or number of each missing noun could only be established on the basis of the gender or number of the determiner and adjective in the stimulus sentence, this task provides a means of determining, via comprehension rather than production, whether abstract features are present in learner grammars, in particular, the uninterpretable gender and number features responsible for concord within the DP.

The picture identification task crucially depends on learners (and controls) identifying the same vocabulary item for any particular picture, as was intended by the researchers in devising the test, as well as knowing the inherent gender of the nouns in isolation. It is, after all, possible that there is more than one lexical item consistent with a particular picture, which could undermine the objective of the task.<sup>8</sup> For this reason, an additional vocabulary test was included (see the following section) to determine whether the learners in fact had the relevant vocabulary and appropriate gender in their interlanguage lexicons. Where they did not, their results were excluded from analysis (see the picture identification task section).

*Vocabulary.* The vocabulary test was administered immediately after the picture identification task. This task consisted of a set of 47 pictures identical to the pictures used in the picture identification task and targeting the same lexical items, namely 24 masculine nouns and 23 feminine.<sup>9</sup> Below each picture was a blank space preceded by a choice of the masculine or feminine form of the article (*el/la*). Subjects had to insert a lexical item corresponding to the picture and circle the appropriate article.

## Results

*Production data.* The production tasks were extremely successful in eliciting DPs. Data from the two tasks were combined, because there were no essential differences in performance across tasks. The breakdown of DPs produced by each group is shown in Table 2. A total of 10,515 DPs in which there was a determiner

but no adjective were elicited, as in Example 7a, as well as 2,867 DPs containing adjectives, as in Example 7b. Sometimes, DPs included an adjective without a determiner, as in Example 7c:

7. a. una camisa (English L1, intermediate proficiency)  
a-F shirt-F  
“a shirt”  
b. una camiseta roja (French L1, intermediate proficiency)  
a-F t-shirt-F red-F  
“a red t-shirt”  
c. Hay pantalones cortos. (English L1, advanced proficiency)  
Are pants-MP short-MP  
“There are short pants.”

Before turning to a detailed analysis of number and gender, we consider to what extent there were word order errors in DPs containing adjectives, that is, DPs of the form \*Det Adj N or \*Adj N instead of (Det) N Adj. An example of such an error is provided in Example 8.

8. a. un viejo hombre (English L1, low proficiency)  
a-M old-M man-M  
“an old man”

As can be seen in Table 2, the incidence of word order errors within the DP is very low (3.5% of all DPs containing adjectives produced by the nonnative speakers), replicating the results from previous studies (Bruhn de Garavito & White, 2002; Gess & Herschensohn, 2001; Hawkins, 1998; Parodi, Schwartz, & Clahsen, 1997). Furthermore, and somewhat surprisingly, the errors were mostly produced by francophone subjects. French, like Spanish has predominantly Det N Adj order in the DP. (In fact, the majority of errors were produced by just two subjects, one in the French-speaking intermediate group and one in the French-speaking low proficiency group.) We interpret these data as evidence that almost all the nonnative speakers have acquired the strong feature value of Num. In other words, the English speakers can reset feature values from weak to strong. This would appear to be counter to the claims of the FFFH, although for some reason, Hawkins (1998) excludes changes in feature strength from his predictions.

Turning now to number and gender agreement, all DPs were coded twice, once for number agreement and once for gender agreement. In the case of DPs containing adjectives, invariant adjectives (such as *difícil*) were excluded from the gender agreement analysis, whereas all adjectives were retained for the number agreement analysis. Similarly, possessives (which do not show gender agreement) were only included in the number agreement analysis. Due to the nature of the tasks, there was considerable variation in the number of DPs produced by each group and by individuals within the groups. Consequently, in order to allow comparisons, results are presented in percentages (proportion of accurate responses over total responses for any particular category).

Table 3. *Production data (Det N) on gender versus number accuracy (%)*

	Spanish Controls	Advanced Proficiency		Intermediate Proficiency		Low Proficiency	
		L1 English	L1 French	L1 English	L1 French	L1 English	L1 French
Gender							
Mean	99.872	96.744	99.154	92.025	95.154	87.884	83.05
SD	0.572	6.413	0.931	7.987	6.211	11.107	10.957
Number							
Mean	99.965	99.259	99.087	99.669	98.275	98.566	94.357
SD	0.159	0.868	2.951	0.808	2.861	2.077	6.097

Table 4. *Production data (Det N Adj) on gender versus number accuracy (%)*

	Spanish Controls	Advanced Proficiency		Intermediate Proficiency		Low Proficiency	
		L1 English	L1 French	L1 English	L1 French	L1 English	L1 French
Gender							
Mean	99.625	98.277	97.594	90.608	90.277	76.44	71.842
SD	0.965	2.803	3.316	12.085	11.464	25.581	17.229
Number							
Mean	99.931	100	96.199	98.514	92.037	92.728	83.485
SD	0.311	0	6.645	3.374	11.888	15.067	12.028

Table 3 compares accuracy on gender and number agreement in DPs consisting only of a determiner and a noun, whereas Table 4 compares gender and number agreement in DPs containing adjectives. In both cases, number is relatively unproblematic. As Table 3 shows, accurate production of appropriate gender in DPs without adjectives is also in general high (ranging from 83% for the low proficiency L1 French group to over 99% for the advanced L1 French group); nevertheless, performance on number is more accurate than on gender for the intermediate and low proficiency groups. A repeated measures analysis of variance (ANOVA) shows no effect for L1, a highly significant effect for proficiency,  $F(2, 110) = 20.08, p < .0001$ , a highly significant effect for feature (gender vs. number,  $F(1, 110) = 52.08, p < .0001$ , a highly significant interaction between proficiency and feature,  $F(2, 110) = 11.2, p < .0001$ , and no interaction between L1 and feature. These results are only in partial accordance with Prediction 1a as the greater problems with gender are not confined to L1 English. The advanced groups are as accurate on gender as on number, in accordance with Prediction 2a. The lack of L1 effects is consistent with Prediction 2b and contrary to the predictions of the FFFH, as well as Prediction 1b.

An example of correct gender agreement is given in Example 7a; incorrect gender is illustrated in Example 9. Examples with correct and incorrect number agreement are provided in Example 10.

- |                           |  |
|---------------------------|--|
| 9. a. un estrella         | (French L1, intermediate proficiency)  |
| a-M star-F                |  |
| “a star”                  |  |
| 10. a. el chico           | (English L1, intermediate proficiency) |
| the-S boy/guy-S           |  |
| b. los sofá               | (French L1, intermediate proficiency)  |
| the-P sofa-S              |  |
| “the sofas” <sup>10</sup> |  |

Table 4 shows similar trends in the case of DPs containing adjectives. A repeated measures ANOVA again shows no effect for L1 although it borders on significance,  $F(1, 109) = 3.6, p < .06$ , a highly significant effect for proficiency,  $F(2, 109) = 19.71, p < .0001$ , a highly significant effect for feature (gender vs. number),  $F(1, 109) = 12.78, p < .001$ , a significant interaction between proficiency and feature,  $F(2, 109) = 4.98, p < .01$ , and no interaction between L1 and feature. For both Det N and Det N Adj, post hoc Scheffé tests show that the advanced and intermediate groups do not differ from the native speakers, whereas both low proficiency groups differ from the controls on gender only.

An example of correct gender in DPs that include adjectives is provided in Example 7b; incorrect gender agreement in such DPs is illustrated in Example 11. Correct and incorrect number agreement are illustrated in Example 12.

- |                        |  |
|------------------------|--|
| 11. a. la barba rojo   | (English L1, intermediate proficiency) |
| the-F beard-F red-M    |  |
| “the red beard”        |  |
| 12. a. los ojos azules | (English L1, low proficiency)          |
| the-P eyes-P blue-P    |  |
| “the blue eyes”        |  |
| b. un pantalones rojo  | (French L1, low proficiency)           |
| a-S pants-P red-S      |  |
| “red pants”            |  |

Comparing Tables 3 and 4, it can also be seen that accuracy on gender is generally lower when an adjective is present, a finding reported by other researchers (e.g., Bruhn de Garavito & White, 2002; Dewaele & Véronique, 2001). Comparing performance on DPs with and without adjectives, once again, there is no effect for L1, a highly significant effect for proficiency,  $F(2, 109) = 29.61, p < .0001$ , a significant effect for DP type (Det N vs. Det N Adj),  $F(1, 109) = 6.66, p < .01$ , a marginally significant interaction between proficiency and DP type,  $F(2, 109) = 3.13, p < .05$ , and no interaction between L1 and DP type.

Because many of the anglophones reported having learned French as an L2 before learning Spanish as L3, we also looked for potential L2 to L3 effects (cf. Leung, 2002), given that the L2 and L3 share a gender feature that is absent in the

Table 5. *Production data (Det N) on gender accuracy (%)*

	Spanish Controls	Advanced Proficiency		Intermediate Proficiency		Low Proficiency	
		L1 English	L1 French	L1 English	L1 French	L1 English	L1 French
<b>Masculine</b>							
Mean	99.722	98.541	99.378	97.012	95.386	94.357	90.362
SD	1.243	2.618	1.339	4.348	5.398	10.174	9.538
<b>Feminine</b>							
Mean	100	95.531	98.896	87.69	95.138	81.994	79.298
SD	0	8.95	1.813	13.644	8.756	17.367	19.13

Table 6. *Production data (Det N Adj) on gender accuracy (%)*

	Spanish Controls	Advanced Proficiency		Intermediate Proficiency		Low Proficiency	
		L1 English	L1 French	L1 English	L1 French	L1 English	L1 French
<b>Masculine</b>							
Mean	100	97.167	99.236	97.917	94.425	87.305	78.174
SD	0	6.575	2.098	7.058	10.489	22.408	25.795
<b>Feminine</b>							
Mean	99.108	98.889	92.687	69.773	82.101	48.839	58.961
SD	2.207	3.333	13.705	37.958	28.217	35.994	32.453

L1. English speakers with prior exposure to French might be at an advantage when compared to learners with no prior exposure to a language with gender. (However, according to the FFFH, there should presumably be no such advantage [at least in postpuberty L2 and L3 acquisition], because it is absence of the relevant feature in the L1 that is crucial. If gender is absent in the L1, it cannot be acquired in any language, regardless of whether this is an L2, an L3, or *L<sub>n</sub>*.)

The English speakers were divided into three groups: early exposure to French (*n* = 36; age 9 and under), late exposure to French (*n* = 18; age 10 and over), and no prior exposure to French (*n* = 14). It turned out that there was no significant difference between these groups, those with no exposure to French performing as accurately on Spanish gender as those with early or late exposure to French. This is true for gender in DPs with and without adjectives.

Accuracy in gender production is examined in more detail in Tables 5 (Det N) and 6 (Det N Adj), which are broken down in terms of masculine versus feminine agreement. As can be seen in Table 5, when the DP consisted only of a determiner and a noun, accuracy was generally high. However, the low proficiency groups

were more accurate on agreement with masculine nouns than with feminine nouns. In other words, they sometimes produced feminine nouns with masculine articles, as in Example 13a, whereas they rarely produced masculine nouns with feminine articles, as in Example 13b.

13. a. el comida (English L1, low proficiency)  
the-M food-F  
“the food”  
b. la pelo (English L1, low proficiency)  
the-F hair-M  
“the hair”

In the case of DPs consisting of Det N, a repeated measures ANOVA shows no effect for L1, a highly significant effect for proficiency,  $F(2, 110) = 15.96$ ,  $p < .0001$ , a highly significant effect for feature (masculine vs. feminine),  $F(1, 110) = 17.51$ ,  $p < .0001$ , a significant interaction between proficiency and feature,  $F(2, 110) = 13.98$ ,  $p < .02$ , and no interaction between L1 and feature. Scheffé tests show no differences between the advanced and intermediate groups and the native speakers, whereas both low proficiency groups differ from them on feminine items.

As Table 6 shows, the inclusion of an adjective within the DP led to considerably reduced accuracy on feminine gender for the intermediate and low proficiency groups. A repeated measures ANOVA shows no effect for L1, a highly significant effect for proficiency,  $F(2, 97) = 18.33$ ,  $p < .0001$ , a highly significant effect for feature (masculine vs. feminine),  $F(1, 97) = 23.15$ ,  $p < .0001$ , a significant interaction between proficiency and feature,  $F(2, 97) = 4.0$ ,  $p < .02$ , and no interaction between L1 and feature. Here Scheffé tests show that the only group differing significantly from the controls is the low proficiency English L1 group on feminine items only.

What we found was that masculine forms of the adjective were produced with feminine nouns (see Example 14a); feminine forms of the adjective were much less likely to be produced with masculine nouns (see Example 14b). Although the most common error type was as in Example 14a, where only the gender of the adjective was inappropriate, there were also cases where the adjective was of the appropriate gender but the determiner was not, as shown in Examples 15a and 15b, as well as errors where both the determiner and the adjective were of inappropriate gender, particularly when the noun was feminine, as shown in Example 16a, rather than masculine, as in Example 16b. In these latter cases, one cannot exclude the possibility that the nouns in question had been assigned the wrong gender in the interlanguage lexicon, because the agreement, although inappropriate, is consistent.

14. a. la barba rojo (English L1, intermediate proficiency)  
the-F beard-F red-M  
“the red beard”  
b. el hombre sentada (French L1, low proficiency)  
the-M man-M sitting-F  
“the seated man”

Table 7. *Vocabulary test gender accuracy (%)*

	Spanish Controls	Advanced Proficiency		Intermediate Proficiency		Low Proficiency	
		L1 English	L1 French	L1 English	L1 French	L1 English	L1 French
<b>Masculine</b>							
Mean	99.772	96.511	97.469	93.037	92.878	88.187	94.228
SD	1.017	6.65	2.662	10.64	6.529	15.101	5.539
<b>Feminine</b>							
Mean	100	95.529	97.521	96.157	97.73	94.699	97.248
SD	0	4.533	3.567	5.448	3.277	7.86	4.48

- 15. a. la pelo blanco (English L1, low proficiency)  
 the-F hair-M white-M  
 “the white hair”
- b. un nariz larga (French L1, advanced proficiency)  
 a-M nose-F long-F  
 “a long nose”
- 16. a. un camiseta rojo (French L1, low proficiency)  
 a-M t-shirt-F red-M  
 “a red t-shirt”
- b. una color roja (English L1, low proficiency)  
 a-F color-M red-F  
 “a red color”

*Vocabulary task.* Where subjects supplied a lexical item in the vocabulary task, they usually got its gender right. Suppliance of the targeted lexical item with its correct gender ranged from just over 50% in the case of the low proficiency francophone group to over 90% in the case of the native speakers. There were three other response types that deserve comment. First, there was a fairly high incidence of failure to respond on the part of the low and intermediate proficiency learners (range = 20–38%; see Table 9 and discussion). Second, in some cases, subjects provided a lexical item that was not the one intended yet its gender was appropriate; the native speakers also fell in this response category. Such responses, which are clearly not errors, constituted less than 10% of each group’s responses. The only real error was suppliance of a lexical item with incorrect gender. For example, under the picture of a bracelet, the subject would write the word *pulsera* (actually feminine in Spanish) and circle the article *el* (the, which is masculine). Incidence of such errors was extremely low (4% or less per experimental group). This suggests that postpuberty learners, regardless of L1 or proficiency level, can successfully acquire the inherent gender on Spanish nouns.

Concentrating now on cases where subjects supplied a noun with its gender correctly identified, we consider whether correct suppliance varied according to L1, proficiency level or gender (masculine vs. feminine), as shown in Table 7.

Table 8. *Vocabulary test gender accuracy (%) by Spanish or French differences*

	Spanish Controls	Advanced Proficiency		Intermediate Proficiency		Low Proficiency	
		L1 English	L1 French	L1 English	L1 French	L1 English	L1 French
Opposite gender							
Mean	99.667	93.091	90.877	90.86	73.209	91.709	75.687
SD	1.491	15.653	11.755	14.507	26.528	13.094	28.552
Same gender							
Mean	100	96.418	99.173	95.523	98.817	91.816	99.097
SD	0	3.57	1.481	5.57	2.581	9.278	2.234

A repeated measures ANOVA shows no significant effects for L1, although this approaches significance,  $F(2, 110) = 3.13, p < .08$ , no effect for proficiency, a significant effect for gender (responses to feminine being more accurate than to masculine),  $F(1, 110) = 5.66, p < .02$ , and no interactions. The only group to differ significantly from the controls is the low proficiency English L1 group on masculine items. Considering, once again, potential effects of prior exposure to French among the English speakers, a repeated measures ANOVA shows no significant effect for prior exposure, a significant effect for gender,  $F(1, 65) = 5.69, p < .02$ , and no interaction between amount of exposure to L2 French and accuracy on gender.

Given the slightly lower accuracy on masculine items, we also checked whether there is an effect depending on whether the noun ends in *-o*, the usual marker for masculine (11 items on the task), or some invariant form (13 items on the task).<sup>11</sup> All experimental groups do show significantly greater accuracy on forms ending in *-o*,  $F(2, 110) = 16.43, p < .0001$ ; there are no effects for L1 or proficiency level. The native speakers show no such effect.

Because French, like Spanish, is a language with grammatical gender, we also consider the potential influence of French as an L1 (for the francophones) or as an L2 (for many of the anglophones). On the vocabulary task, the majority of lexical items share the same gender in French and Spanish (14 masculine and 18 feminine items out of 47); however, there were items where the gender differs (10 masculine and 5 feminine items out of 47). Table 8 presents mean percentage accuracy on forms that realize gender in the same way or the opposite way in the two languages. Similarities and differences between French and Spanish in realization of gender did indeed affect accuracy, but only in the case of the native speakers of French. The francophones are noticeably less accurate on lexical items where the gender differs in the two languages. There is a near significant effect for L1,  $F(1, 109) = 3.74, p < .06$ , and for proficiency,  $F(2, 109) = 2.64, p < .08$ , a highly significant effect for gender realization,  $F(1, 109) = 33.5, p < .0001$ , and a highly significant interaction between L1 and gender realization,  $F(1, 109) = 18.2, p < .0001$ . Scheffé tests show that only the French L1 intermediate and low

Table 9. *Vocabulary test failures to respond (% of total responses)*

	Spanish Controls	Advanced Proficiency		Intermediate Proficiency		Low Proficiency	
		L1 English	L1 French	L1 English	L1 French	L1 English	L1 French
Overall							
Mean	0	8.723	11.171	20.567	29.575	34.356	38.122
SD	0	7.862	10.084	16.807	16.407	14.863	23.718
Opposite gender							
Mean	0	16.666	24.167	32.222	49	49.608	52.223
SD	0	5.466	4.94	23.15	22.712	19.639	28.545
Same gender							
Mean	0	5.003	5.079	15.107	20.471	27.208	31.513
SD	0	4.219	6.544	14.907	14.066	14.587	22.474

proficiency groups differ significantly from the native speakers on items whose gender is different in French and Spanish.

Considering the English L1 groups by amount of prior exposure to French, there are no significant effects for opposite gender in French and Spanish and no interactions. In other words, even those with early exposure to French did not show an advantage on items where gender in French and Spanish is the same.

There are even clearer effects of French gender on the incidence of failures to respond. These results are presented in Table 9. Francophones are less likely to supply an answer in the case of items which differ in gender between the two languages; the behavior of the anglophones is also affected by this distinction. There is a near significant effect for L1,  $F(1, 110) = 3.15, p < .08$ , a highly significant effect for proficiency,  $F(2, 110) = 18.82, p < .0001$ , a highly significant effect for whether items have the same or opposite gender in Spanish and French,  $F(1, 110) = 198.41, p < .0001$ , and a significant interaction between L1 and gender realization,  $F(1, 110) = 4.09, p < .05$ .

Considering the incidence of *no responses* in the English L1 subjects grouped by amount of prior exposure to French, there was a significant effect for lexical items that differed in gender between the two languages,  $F(2, 65) = 92.37, p < .0001$ , no effect for prior exposure, and no interaction. In other words, even those with no prior exposure to French were more likely to provide an answer when the picture illustrated a word whose gender was the same in French and Spanish, a curious result for which we have no explanation.

**Picture identification task.** As far as the picture identification task results are concerned, certain test items were eliminated from the analyses because of inconsistent responses by the native speaker control group.<sup>12</sup> Because gender agreement in this task can only be investigated in cases where the pictures unambiguously target a particular lexical item, it was important to exclude any potentially problematic cases. This left 10 items testing number (5 singular, 5 plural) and 12 items testing

Table 10. *Picture identification task gender versus number accuracy (%)*

	Spanish Controls	Advanced Proficiency		Intermediate Proficiency		Low Proficiency	
		L1 English	L1 French	L1 English	L1 French	L1 English	L1 French
Gender							
Mean	96.387	94.349	96.511	82.868	84.736	56.086	66.554
SD	5.596	6.793	7.098	16.413	17.731	4.689	30.03
Number							
Mean	97.444	99	95.868	90.214	90.467	83.174	78.243
SD	5.586	3.162	8.063	14.965	10.483	17.086	25.867

Table 11. *Picture identification task gender accuracy (%)*

	Spanish Controls	Advanced Proficiency		Intermediate Proficiency		Low Proficiency	
		L1 English	L1 French	L1 English	L1 French	L1 English	L1 French
Masculine							
Mean	96.666	90.999	95.312	77.727	83.148	51.944	54.167
SD	6.841	13.431	13.598	19.304	23.583	34.985	49.81
Feminine							
Mean	96.667	98.333	95.833	86.945	87.416	61.03	74.86
SD	8.719	5.272	12.91	17.519	16.837	35.616	21.99

gender (6 masculine, 6 feminine), as well as 16 distractors. In addition, where subjects had failed to provide a response on the vocabulary task (see vocabulary test section), provided an incorrect response, or provided a lexical item whose gender did not correspond to the targeted form, the corresponding item was excluded from analysis of their results on the picture identification task. This was done to ensure that the analysis is devoted to items for which learners can be deemed independently (i.e., via the vocabulary test) to know a lexical item and its inherent gender. Only then is it reasonable to examine whether they in fact have gender agreement, that is, whether they can interpret gender and number agreement consistently, thus allowing them to identify the picture corresponding to the relevant null nominal. (Recall that the FFFH claims that even when learners have the appropriate inherent gender on nouns, gender agreement will fail.) In consequence, all results will be presented in terms of mean percentage accuracy, because the actual numbers involved are not comparable.

Results from the picture identification task are presented in Tables 10 and 11. All groups were highly accurate on the distractor sentences, performing almost at ceiling, suggesting that they had no difficulty in principle with this task. Table 10 compares mean accuracy on items testing the features number and gender.

Once again, number causes relatively few problems. As can be seen in Table 10, the low and intermediate groups are less accurate on gender than on number but this is regardless of L1 (contra FFFH and not in conformity with Prediction 1a either). A repeated measures ANOVA shows no effect for L1, a highly significant effect for proficiency,  $F(2, 108) = 24.37, p < .0001$ , a highly significant effect for feature (gender vs. number),  $F(1, 108) = 16.21, p < .0001$ , a significant interaction between proficiency and feature,  $F(2, 108) = 4.96, p < 0.01$ , and no interaction between L1 and feature. Scheffé tests show that both low proficiency groups differ significantly from the controls on gender. When the English groups are considered in terms of their prior exposure to French, there is no effect for exposure, a significant effect for feature,  $F(1, 63) = 29.04, p < .0001$ , and no interaction. The results are consistent with those of the production tasks, suggesting some difficulty with gender at lower proficiency levels but no influence from the L1 (contrary to Prediction 1b) or prior exposure to an L2 with gender. The advanced groups, regardless of L1, have no problems in interpreting number or gender agreement, in accordance with Predictions 2a and 2b.

Results comparing masculine versus feminine gender assignment are presented in Table 11. All learner groups show greater accuracy on feminine items. A repeated measures ANOVA shows no effect for L1, a highly significant effect for proficiency,  $F(1, 102) = 20.54, p < .0001$ , a significant effect for masculine versus feminine items  $F(1, 102) = 8.9, p < .01$ , and no interactions. According to Scheffé tests, both low proficiency groups differ significantly from the controls on masculine items only. When the English speakers alone are considered, there is a significant effect for masculine versus feminine gender,  $F(2, 59) = 4.24, p < .05$ , no effect for prior exposure to French, and no interaction.

In effect, given a sentence with a null nominal introduced by the feminine form of the determiner and adjective (as in Example 17a), subjects are very likely to pick a picture corresponding to a feminine noun (in this case, a suitcase; see Figure 1). Given a null nominal introduced by the masculine form of the determiner and adjective (as in Example 17b), subjects at the advanced and intermediate levels are mostly very accurate, picking the appropriate picture of a masculine noun (a sweater in this example). Nevertheless, some subjects sometimes select a picture (a blouse) corresponding to a feminine noun.

17. a. Ponlas ahí cerca                    de la roja  
      Put            them over-there by the-F red-F  
      “Put them over there by the red one.”  
      b. ¿Dónde puse el            nuevo que compré?  
      Where put-1S the-M new-M that I-bought  
      “Where did I put the new one that I bought?”

In other words, masculine forms of determiners or adjectives are sometimes taken to agree with feminine nouns, whereas feminine forms of determiners or adjectives are significantly less likely to agree with masculine nouns. In the case of the low proficiency groups, masculine determiners and adjectives “agree” with feminine nouns about 50% of the time, that is, they are performing at chance when the agreeing forms are masculine but not when they are feminine.

## DISCUSSION

To summarize, the results are highly consistent across the production and picture identification tasks. Number proved comparatively unproblematic for all learners. In both tasks, lower proficiency subjects were more accurate on number than gender and more accurate when the noun was masculine than when it was feminine; the advanced and intermediate groups did not differ significantly from native speakers; and there were significant effects for proficiency but not for L1 or for prior exposure to an L2 with gender. Nevertheless, there are some L1 or prior L2 effects in the vocabulary test, wherever there are gender differences on the same nouns in French and Spanish. In general, exposure to an L2 with gender did not help or hinder the performance on gender in the L3.

In accordance with our predictions in our second hypothesis and contrary to the FFFH, advanced learners of Spanish whose L1 is English revealed no difficulties with gender agreement in either production or comprehension, performing as well on gender as on number just like the francophones and the native speakers. Learners of intermediate proficiency performed similarly. These results suggest the acquirability of not only interpretable gender features on nouns but also uninterpretable features on determiners and adjectives. This is particularly evident in the results from the picture identification task, where the stimuli *ONLY* contained uninterpretable features, because the nouns were nonovert; thus, reconstruction of the appropriate gender on the noun was entirely dependent on intact agreement mechanisms. In other words, results strongly support the claim that postpuberty learners are able to acquire gender agreement regardless of the status of gender features in the L1, suggesting, contra the FFFH, that there is no impairment in access to “new,” uninterpretable formal features. Furthermore, the proficiency effects that we found in the case of the French speakers are somewhat unexpected according to the FFFH account: when the L1 has gender, agreement should not be subject to proficiency effects, because there is nothing to learn other than the gender of the nouns themselves: once the gender of the noun is known, uninterpretable features on adjectives and determiners are checked by mechanisms already in place because of the L1 grammar.

An anonymous reviewer speculates that the success of the intermediate and advanced groups might be attributable to ceiling effects, with the experimental tasks being insufficiently demanding. Even if this were the case, it is unclear what ceiling effects would reflect other than appropriate, and unconscious, knowledge of L2 gender. One alternative is that explicit prescriptive knowledge of gender is somehow brought to bear on performance in such circumstances. In this context, it is important to recall that neither production nor comprehension tasks involved an explicit focus on gender. Thus, it seems unlikely that the L2 learners somehow understood that gender was being tested. Furthermore, as far as we can see, the FFFH predicts problems in acquisition of gender agreement regardless of the ease or difficulty of the tasks involved. On the other hand, as discussed below, it is conceivable that task difficulty comes into play in the sense that certain kinds of tasks or communication pressures may lead L2 learners to resort to the use of default forms, even when they have the appropriate underlying representation of gender.

Although our results favor some kind of full access account, it is not clear to what extent they support FTFA, which would predict L1 effects at the low

proficiency level. The absence of L1 effects in both the production and picture identification tasks is inconsistent with both FFFH and FTFA. Our Prediction 1b was that low proficiency anglophones would perform less well on gender than low proficiency francophones, a prediction which was not borne out. Our Prediction 1a was that low proficiency anglophones would perform more accurately on number agreement than gender, a prediction that was borne out but also proved true of the low proficiency francophone group, suggesting that this result did not reflect L1 effects. One might speculate that the uninterpretable gender features had been acquired relatively early, at some stage prior to the level reached by our English L1 low proficiency group. This requires further investigation with groups who have been identified as true beginners (see Gess & Herschensohn, 2001).

Indeed, the fact that the results from the vocabulary task did show L1 effects whereas results from the production data and picture identification task did not suggests a modified version of Carroll's (1989) claim that it is learning the inherent gender on nouns that is problematic, rather than gender agreement as such. Where gender in French and Spanish differed, the francophones were affected to a significant extent in terms of the gender they assigned to particular lexical items. When failures to respond are taken into consideration, the anglophones were affected as well. All the problematic items were removed from the analysis of each subject's performance on the picture identification task, allowing us to be reasonably confident that performance on that task reflects gender agreement not gender assignment. Having acquired inherent gender on nouns, agreement (via feature checking) appears to come free, so to speak.

One final issue remains to be addressed, namely, the fact that the two genders are not equally problematic. Accuracy in production is lower on feminine nouns than on masculine nouns. In other words, masculine forms of determiners and adjectives are found with feminine nouns rather than vice versa. On the face of things, it might appear that performance on picture identification is the opposite, because subjects were more accurate in picking out feminine items. In fact, these results are mutually consistent. Given a sentence with a feminine determiner or adjective, subjects picked the picture of the corresponding feminine noun, whereas given a masculine determiner or adjective, they did not necessarily pick the picture corresponding to a masculine noun. In both production and picture identification, then, the predominant errors are of the form  $\text{Det}_M N_F$  or  $\text{Det}_M N_F \text{Adj}_M$ .

These findings are consistent with other results in the literature, where learners of French and Spanish have been reported as using one or other gender (usually masculine) as a default (e.g., Bruhn de Garavito & White, 2002; Dewaele & Véronique, 2001; Fernández-García, 1999; Franceschina, 2001; Hawkins, 1998).<sup>13</sup> This preference for one gender over another clearly requires explanation. It does not follow from FTFA, because this phenomenon is observed among learners whose L1 has gender, nor from FFFH, where there is no reason to expect unidirectionality of errors. If there is a breakdown in feature checking but no loss of gender on nouns, one would anticipate problems in both directions.

There is a tendency in the literature to assume that the grammars of adult native speakers do not show variability in gender assignment or gender agreement (e.g., Schriefers & Jescheniak, 1999) and that an error rate of even 10% on the part of nonnative speakers is significant and requires explanation, often in terms of failure

of acquisition, as we have seen. In fact, variability in gender assignment is also found in native speaker grammars. For instance, in an analysis of gender usage in a corpus of 19th century Quebec French, Klapka (2002) reports the incidence of variable gender assignment to be 7%; normally feminine nouns, for example, are sometimes found with masculine forms of articles and/or adjectives, as in *un affaire écrit* (a-M matter-F written-M).<sup>14</sup> This means that variability in general has to be accounted for; the same kind of account may prove appropriate for both native and nonnative speakers.

One such account is provided by distributed morphology (Halle & Marantz, 1993; Harley & Noyer, 1999), according to which lexical items undergo late insertion into the structure. In other words, feature checking in the syntax, as described in the second section, involves movement of abstract bundles of features, not actual lexical items. Whereas features in the syntax are fully specified, those on lexical items can be underspecified (Lumsden, 1992). When it comes to lexical insertion, the features on the lexical item must be consistent with those on the relevant syntactic node. They do not need to be identical, but they must form a proper subset; otherwise, a feature clash will result.

Under current assumptions, features are not in fact represented with binary plus or minus values; rather, they are present or absent (e.g., Harley, 1994). Furthermore, uninterpretable features are not specified as to value but have their values determined by the mechanism responsible for agreement, namely Agree (Chomsky, 2001). It has been proposed that the default gender is masculine in Spanish (Harris, 1991), as well as in French (Hulk & Tellier, 1999) and in Italian (Riente, 2003). According to Harris (1991, p. 44), masculine is literally the unmarked default, “the absence of any information about gender in lexical entries”; instead, the only value to be entered in the lexical entry of Spanish nouns is feminine.

Consider, now, the implications of such assumptions. If a noun marked [+fem] is inserted into a DP with its head specified [+fem], the determiner and adjective positions will become feminine by agreement (or feature checking), so feminine forms can be inserted. However, masculine forms of determiners or adjectives (not specified, hence defaults) are not precluded because they do not result in a feature clash: feminine in the syntax, no specification on the lexical items. On the other hand, if the head of the DP is marked [+masc], only default forms unspecified for gender can be inserted, because insertion of [+fem] items would result in a clash of features. This accounts for the fact that the errors that we found by and large take the forms  $\text{Det}_M \text{N}_F$ ,  $\text{Det}_M \text{N}_F \text{Adj}_F$  or  $\text{Det}_F \text{N}_F \text{Adj}_M$  rather than  $\text{Det}_F \text{N}_M$ ,  $\text{Det}_F \text{N}_M \text{Adj}_M$  or  $\text{Det}_M \text{N}_M \text{Adj}_F$ .<sup>15</sup> In the normal course of events, the most specified forms “win” as far as lexical insertion is concerned, which is why failure of gender agreement is unusual in native speaker grammars and why the learners show correct gender agreement in the majority of cases.

Why do learners resort to defaults at all? It is conceivable that such difficulties reflect performance issues of some kind. That is, learners, even at low levels of proficiency, might represent abstract gender features but fail to implement agreement on occasion, so that there is a discrepancy between abstract properties and their surface manifestation, perhaps because of communication pressures, that make it impossible to retrieve the appropriate item from the lexicon (see Lardiere, 2000, and Prévost & White, 2000, for relevant discussion). Such proposals were

based on examination of production data. In the present study, the surfacing of masculine as a default was found, not only in the production data but also on the picture identification task. Indeed, although showing the same response patterns as found in production, performance on gender on the picture identification task was significantly lower,  $F(1, 109) = 5.61, p < .02$  (compare Tables 3 and 10).<sup>16</sup> What these results suggest is that resorting to default forms is not confined to production, contrary to earlier assumptions (Prévost & White, 2000).

In conclusion, we have demonstrated, contra the FFFH, that adult learners of Spanish are able to acquire uninterpretable gender features on determiners and adjectives and to show gender concord within Spanish DPs, even given an L1 (English) that lacks grammatical gender. Although there were problems with gender at the lowest level of proficiency, they occurred regardless of the status of gender in the L1. Accuracy was high, even at the intermediate level of proficiency, whereas the performance of advanced learners was indistinguishable from native speakers. Nevertheless, there was some variability in gender agreement both in production and in picture identification, mostly occurring as overuse of masculine determiners and adjectives with feminine nouns. We suggest that this reflects the insertion of masculine default items into the structure, in accordance with mechanisms proposed for natural language in general.

#### ACKNOWLEDGMENTS

We thank Hela Ben Ayed and Pablo Ruiz for their assistance, Joyce Bruhn de Garavito and Philippe Prévost for arranging the recruiting and testing of many of the subjects, and Julia Herschensohn for comments. We are also most grateful to Joyce Bruhn de Garavito for suggesting the picture identification task. This research was conducted with the support of research grants from the Social Sciences Research Council of Canada (SSHRCC) and the Government of Québec Fonds pour la formation de Chercheurs et l'Aide à la Recherche (FCAR).

#### NOTES

1. It is not clear why FFFH assumes a distinction between acquirability of interpretable versus uninterpretable features, especially in the case of grammatical gender, where the “interpretable” feature, in fact, has no consequences for semantic interpretation.
2. According to Ritter (1993), in Romance languages gender is also located in Num. However, we will follow others in assuming that gender is a feature on the noun itself (e.g., Carstens, 2000), an assumption that Ritter makes for languages like Hebrew.
3. There are other ways of accounting for concord, for example, by feature copying from N to Det and Adj (Halle & Marantz, P. 1993, p. 115) or by feature percolation from the head N to other categories within the DP (Carstens, 1991, 1993). For the sake of the argument, we adopt the feature-checking account here because this is what is assumed by the FFFH.
4. Adjectives were not examined.
5. In recent versions of FFFH, Hawkins (2001) proposed modulated structure building, whereby the initial state of L2 acquisition involves minimal trees (lacking functional

structure), along the lines proposed by Vainikka and Young–Scholten (1994). He also proposes (in contrast to Vainikka and Young–Scholten) that once functional categories and features emerge, they take on L1 properties. Because the representation of gender and number features crucially depends on functional categories (Det, Num) and features (gender, number), the proposal for such a prefunctional stage, even if correct, is irrelevant to our present concerns.

6. In fact, this prediction does not strictly follow if it can be shown that there are independent reasons to consider gender as more difficult than number.
7. All DPs in Spanish encode both gender and number, which raised the question of what to do with the gender and number features of the third picture (the foil) in each set. In some triplets, number was held constant across all three pictures when gender was targeted and gender was held constant when number was targeted. In other triplets, the foil differed from the target item in both gender and number (as in Figure 1). These differences in the nature of the foils did not affect the overall results.
8. An anonymous reviewer points out that, even so, one cannot be completely sure what lexical item subjects had in mind when responding to the pictures in the picture identification task. We can see no way around this problem and acknowledge it as a weakness.
9. One feminine item was inadvertently omitted from this task.
10. It is clear from the context provided by the picture that plural was intended here. See Bruhn de Garavito (1994) for effects of phonological transfer between French and Spanish such that the L2 learner's knowledge of number can be obscured.
11. A corresponding analysis for feminine items was not possible, because all feminine nouns in this task ended in *-a*.
12. Piloting of the test on native speakers had led to the elimination of other items. As it turned out, items retained in the task after piloting nevertheless resulted in some variability among native speaker controls.
13. Use of masculine as the default agreement within the DP is also reported in L1 acquisition (see Fernández–García, 1999).
14. See Hulk and Tellier (1999) for discussion of variable gender assignment by native speakers of Romance languages in some rather subtle cases involving “conflictual agreement.” (However, in these cases, the problematic adjectives are predicative, involving agreement external to the DP, rather than attributive, involving internal agreement.)
15. We omit from consideration here cases where both Det and Adj have “incorrect” agreement, because one cannot rule out the possibility that the noun has been assigned the wrong gender, in which case there is no agreement problem.
16. The comparison here is between DPs containing adjectives in the production task, because all the null nominals in the picture identification task contained adjectives.

## REFERENCES

- Andersen, R. (1984). What's gender good for, anyway? In R. Andersen (Ed.), *Second languages: A cross-linguistic perspective*. Rowley, MA: Newbury House.
- Bartning, I. (2000). Gender agreement in L2 French: Pre-advanced vs. advanced learners. *Studia Linguistica*, 54, 225–237.
- Bernstein, J. (1993). *Topics in the syntax of nominal structure across Romance*. PhD dissertation, City University of New York.

- Borer, H. (1984). *Parametric syntax*. Dordrecht: Foris.
- Bosque, I., & Picallo, C. (1996). Postnominal adjectives in Spanish DPs. *Journal of Linguistics*, 32, 349–385.
- Bruhn de Garavito, J. (1994). *Acquisition of the Spanish plural: What is taught and what must be acquired*. Paper presented at the Fourteenth Cincinnati Conference on Romance Languages and Linguistics, May 1994, Cincinnati, OH.
- Bruhn de Garavito, J., & White, L. (2002). L2 acquisition of Spanish DPs: The status of grammatical features. In A. T. Pérez-Leroux & J. Liceras (Eds.), *The acquisition of Spanish morphosyntax: The L1/L2 connection* (pp. 153–178). Dordrecht: Kluwer.
- Carstens, V. (1991). *The morphology and syntax of determiner phrases in Kiswahili*. Unpublished PhD dissertation, University of California at Los Angeles.
- Carstens, V. (1993). *Deriving agreement*. Unpublished manuscript, Cornell University.
- Carstens, V. M. (2000). Concord in minimalist theory. *Linguistic Inquiry*, 31, 319–355.
- Carroll, S. (1989). Second-language acquisition and the computational paradigm. *Language Learning*, 39, 535–594.
- Chomsky, N. (1995). *The minimalist program*. Cambridge, MA: MIT Press.
- Chomsky, N. (2001). Beyond explanatory adequacy. *MIT Occasional Papers in Linguistics*, 20.
- Clahsen, H., & Muysken, P. (1989). The UG paradox in L2 acquisition. *Second Language Research*, 5, 1–29.
- Corbett, G. (1991). *Gender*. Cambridge: Cambridge University Press.
- Cressey, W. (1978). *Spanish phonology and morphology: A generative view*. Washington, DC: Georgetown University Press.
- Dewaele, J.-M., & Véronique, D. (2001). Gender assignment and gender agreement in advanced French interlanguage: A cross-sectional study. *Bilingualism: Language and Cognition*, 4, 275–297.
- Duffield, N., & White, L. (1999). Assessing L2 knowledge of Spanish clitic placement: Converging methodologies. *Second Language Research*, 15, 133–160.
- Epstein, S., Flynn, S., & Martohardjono, G. (1996). Second language acquisition: Theoretical and experimental issues in contemporary research. *Brain and Behavioral Sciences*, 19, 677–758.
- Fernández-García, M. (1999). Patterns in gender agreement in the speech of second language learners. In J. Gutiérrez-Rexach & F. Martínez-Gill (Eds.), *Advances in Hispanic linguistics*. Somerville, MA: Cascadilla Press.
- Franceschina, F. (2001). Morphological or syntactic deficits in near-native speakers? An assessment of some current proposals. *Second Language Research*, 17, 213–247.
- Gess, R., & Herschensohn, J. (2001). Shifting the DP parameter: A study of anglophone French L2ers. In C. R. Wiltshire & J. Camps (Eds.), *Romance syntax, semantics and their L2 acquisition* (pp. 105–119). Amsterdam: Benjamins.
- Halle, M., & Marantz, A. (1993). Distributed morphology and the pieces of inflection. In K. Hale & S. J. Keyser (Eds.), *The view from building 20* (pp. 111–176). Cambridge, MA: MIT Press.
- Harley, H. (1994). Hug a tree: Deriving the morphosyntactic feature hierarchy. *MIT Working Papers in Linguistics*, 21, 289–320.
- Harley, H., & Noyer, R. (1999). State-of-the-article: Distributed morphology. *GLOT International*, 4, 3–9.
- Harris, J. (1991). The exponence of gender in Spanish. *Linguistic Inquiry*, 22, 27–62.
- Hawkins, R. (1998, March). *The inaccessibility of formal features of functional categories in second language acquisition*. Paper presented at the Pacific Second Language Research Forum (PacSlrf), Tokyo.
- Hawkins, R. (2001). *Second language syntax: A generative introduction*. Oxford: Blackwell.
- Hawkins, R., & Chan, C. Y.-H. (1997). The partial availability of Universal Grammar in second language acquisition: The “failed functional features hypothesis.” *Second Language Research*, 13, 187–226.
- Hawkins, R., & Franceschina, F. (in press). Explaining the acquisition and nonacquisition of determiner-noun gender concord in French and Spanish. In J. Paradis & P. Prévost (Eds.), *The acquisition of French in different contexts*. Amsterdam: Benjamins.
- Hawkins, R., & Liszka, S. (in press). Locating the source of defective past tense marking in advanced L2 English speakers. In R. van Hout, A. Hulk, F. Kuiken, & R. Towell (Eds.), *The interface between syntax and lexicon in second language acquisition*. Amsterdam: Benjamins.

- Hulk, A., & Tellier, C. (1999). Conflictual agreement in Romance nominals. In J. M. Authier, B. Bullock, & L. Reed (Eds.), *Formal perspectives on Romance linguistics* (pp. 179–195). Amsterdam: Benjamins.
- Klapka, L. (2002, May). *L'accord du genre en français québécois au 19ième siècle*. Paper presented at the Annual Meeting of the Canadian Linguistics Association, Toronto.
- Lardiere, D. (2000). Mapping features to forms in second language acquisition. In J. Archibald (Ed.), *Second language acquisition and linguistic theory* (pp. 102–129). Oxford: Blackwell.
- Liceras, J., Diaz, L., & Mongeon, C. (1999, October). *N-drop and determiners in native and non-native Spanish: More on the role of morphology in the acquisition of syntactic knowledge*. Paper presented at the Conference on the First and Second Language Acquisition of Spanish and Portuguese, Washington, DC.
- Leung, Y.-K. I. (2002). L2. vs. L3 initial state: Evidence from the acquisition of French DPs by Vietnamese monolinguals and Cantonese-English bilinguals. In M. Hooghiemstra (Ed.), *Proceedings of the Second International Conference on Third Language Acquisition and Trilingualism* (CD-ROM). Ljouwert/Leeuwarden, The Netherlands: Fryske Akademy.
- Lumsden, J. (1992). Underspecification in grammatical and natural gender. *Linguistic Inquiry*, 22, 469–486.
- Mallen, E. (1997). A minimalist approach to concord in noun phrases. *Theoretical Linguistics*, 23, 49–77.
- Ouhalla, J. (1991). *Functional categories and parametric variation*. London: Routledge.
- Parodi, T., Schwartz, B. D., & Clahsen, H. (1997). On the L2 acquisition of the morphosyntax of German nominals. *Essex Research Reports in Linguistics*, 15, 1–43.
- Pollock, J.-Y. (1989). Verb movement, Universal Grammar, and the structure of IP. *Linguistic Inquiry*, 20, 365–424.
- Prévost, P., & White, L. (2000). Missing surface inflection or impairment in second language acquisition? Evidence from tense and agreement. *Second Language Research*, 16, 103–133.
- Riente, L. (2003). Ladies first: The pivotal role of gender in the Italian nominal inflection system. *McGill Working Papers in Linguistics*, 17(2), 1–53.
- Ritter, E. (1993). Where's gender? *Linguistic Inquiry*, 24, 795–803.
- Sabourin, L. (2001). L1 effects on the processing of grammatical gender in L2. In S. Foster-Cohen & A. Nizgorodcew (Eds.), *Eurosla yearbook* (Vol. 1, pp. 159–169). Amsterdam: Benjamins.
- Schriefers, H., & Jescheniak, J.-D. (1999). Representation and processing of grammatical gender in language production: A review. *Journal of Psycholinguistic Research*, 28, 575–600.
- Schwartz, B. D., & Sprouse, R. (1994). Word order and nominative case in nonnative language acquisition. In T. Hoekstra & B. Schwartz (Eds.), *Language acquisition studies in generative grammar* (pp. 317–368). Amsterdam: Benjamins.
- Schwartz, B. D., & Sprouse, R. (1996). L2 cognitive states and the full transfer/full access model. *Second Language Research*, 12, 40–72.
- Smith, N., & Tsimpli, I.-M. (1995). *The mind of a savant*. Oxford: Blackwell.
- Snyder, W. (1995). *Language acquisition and language variation: The role of morphology*. PhD dissertation, MIT.
- Stockwell, R. R., Bowen, J. D., & Martin, J. W. (1965). *The grammatical structures of English and Spanish*. Chicago: University of Chicago Press.
- Vainikka, A., & Young-Scholten, M. (1994). Direct access to X'-theory: Evidence from Korean and Turkish adults learning German. In T. Hoekstra & B. D. Schwartz (Eds.), *Language acquisition studies in generative grammar* (pp. 265–316). Amsterdam: Benjamins.
- Valois, D. (1991). *The internal syntax of DP*. PhD dissertation, University of California at Los Angeles.
- White, L. (1989). *Universal grammar and second language acquisition*. Amsterdam: Benjamins.
- White, L. (2003). *Second language acquisition and universal grammar*. Cambridge: Cambridge University Press.