

LEXICAL AND MORPHOLOGICAL ASPECTS OF GENDER AND THEIR EFFECT
ON THE ACQUISITION OF GENDER AGREEMENT IN SECOND LANGUAGE
LEARNERS

By ALENA KIROVA

A dissertation submitted to the
Graduate School-New Brunswick
Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Spanish

Written under the direction of Liliana Sánchez

And approved by

New Brunswick, New Jersey

January 2016

ABSTRACT OF THE DISSERTATION

Lexical and Morphological Aspects of Gender and their Effect on the Acquisition of
Gender Agreement in Second Language Learners

By ALENA KIROVA

Dissertation Director:

Liliana Sánchez

The main purpose of this dissertation is to investigate whether L2 learners can acquire and perform target-like on gender agreement. Gender has been studied by SLA researchers in order to address the central question in generative SLA – acquirability of syntactic features in the L2. With respect to this question, some studies have shown that L2 learners can exhibit target-like gender agreement, which has been considered as evidence in favor of acquirability of syntactic features, while other studies seem to indicate that gender agreement is permanently impaired in the L2, which has been proposed as evidence against acquirability of syntactic features. In this dissertation, I propose that gender is not an ideal candidate to address this question, because gender is not a purely syntactic feature but rather a complex lexical-morphological and syntactic feature whose acquisition hinges upon a number of factors, and not exclusively on the availability of syntactic features. Following the Separation Hypothesis (Beard, 1995; Lardiere, 2000) and the Missing Surface Inflection Hypothesis (Haznedar & Schwartz,

1997; Prévost & White, 1999), I make a distinction between abstract syntactic features and their concrete lexical-morphological manifestation in a given language, and propose that non-target like performance on the latter may not be indicative of a lack of the former in the interlanguage grammars. Furthermore, I propose that having a first language with gender marking does not necessarily facilitate the acquisition of morphological gender in the L2. In order to test these proposals, I conducted a study with L1 English-L2 Spanish learners, L1 Russian-L2 Spanish learners and native speakers of Spanish using gender comprehension tasks (a grammaticality judgment task and a picture-matching task). The findings show that both high proficiency L1 English and L1 Russian learners of Spanish can behave target-like on gender agreement, and that low proficiency learners are affected by lexical and morphological characteristics of Spanish gender. Thus, syntactic features are in fact acquirable in the L2 irrespective of whether the L1 has gender, but the lexical-morphological component of gender has to be acquired as well in order for the surface gender agreement to appear target-like in a given language.

ACKNOWLEDGMENTS

I would like to say a big and sincere thank you to the following people.

To my all time adviser Liliana Sánchez, for guiding me, for bombarding me with questions, for being nice and sweet in her emails and in person, for establishing deadlines, and for always being willing to help me. Te estaré eternamente agradecida.

To my committee members Jane Grimshaw, Jennifer Austin, and José Camacho, for their valuable feedback and comments that helped me immensely to develop and improve my work.

To my FSU advisers Robert Romanchuk and Lisa Wakamiya, for encouraging me to pursue my graduate career, for being my true friends and for keeping my cyclamen.

To my friends at the Department of Spanish and Portuguese at Rutgers University and Florida State University for participating in my research experiments. It would have been more difficult without you (not that it wasn't difficult!).

To my great friend Lake Mathison, for helping me with the manuscript revision and for letting me sleep on her sofa when I needed it.

To my sweet friend Mayumi, for the numerous cups of tea that we shared while I was working on the dissertation at Panera and she was grading and preparing for exams.

To my dear friend Galiya who I spent hours talking to, complaining and sharing happy moments in the process of writing.

To my sister Tuyaara, my father Alexander, my little nephew Dima and all my friends in Russia who always supported me along the way and who were so happy to learn that I successfully defended the dissertation.

THANK YOU.

TABLE OF CONTENTS

ABSTRACT.....	ii-vi
ACKNOWLEDGMENTS.....	iv-vi
CHAPTER ONE: SYNTACTICALLY CHALLENGED OR LEXICALLY COMPROMISED?.....	1
CHAPTER TWO: GENDER AS AN INTERFACE FEATURE.....	26
2.1 Gender in Minimalism	27
2.2 Gender in Theories of Lexical Retrieval.....	35
2.3 Gender in Monolingual Studies of Lexical Organization.....	40
2.4 Studies of Gender in the Bilingual Lexicon.....	45
2.5 The Study of Gender in Language Acquisition.....	48
2.5.1 L1 gender acquisition.....	51
2.5.2 L2 gender acquisition.....	55
2.5.3 Lexical/syntactic hypothesis.....	62
CHAPTER THREE: METHODOLOGY.....	77
3.1 Experiment 1.....	79
3.1.1 Participants	79
3.1.2 Research Tasks.....	82
3.1.2.1 Grammaticality judgment task.....	82
3.1.2.2 Picture-matching task.....	88
3.2. Experiment 2.....	90
3.2.1 Participants.....	91
3.2.2 Research tasks.....	93

CHAPTER FOUR: RESULTS.....	96
4.1 Experiment 1.....	96
4.1.1 Accuracy.....	96
4.1.1.1 Age of Exposure.....	97
4.1.1.1.1 GJT.....	97
4.1.1.1.2 PMT.....	100
4.1.1.2 Proficiency.....	103
4.1.1.2.1 GJT.....	103
4.1.1.2.2 PMT.....	108
4.1.2 Reaction Times.....	111
4.1.2.1 Age of Exposure.....	111
4.1.2.1.1 GJT.....	111
4.1.2.1.2 PMT.....	115
4.1.2.2 Proficiency.....	118
4.1.2.2.1 GJT.....	118
4.1.2.2.2 PMT.....	124
4.2 Experiment 2.....	128
4.2.1 Accuracy.....	128
4.2.2 Reaction times.....	131
CHAPTER FIVE: DISCUSSION.....	138
5.1 Conclusion	158
REFERENCES.....	162
APPENDICES.....	174

CHAPTER 1: Syntactically Challenged or Lexically Compromised?

Two central areas of study in the field of Second Language Acquisition (SLA) are the acquisition of syntax and the acquisition of the lexicon. While syntax is defined as a computational mechanism that derives syntactic structures (Hauser, Chomsky, & Fitch, 2002), the lexicon can be defined as the storage of memorized lexical items that can be smaller than words (affixes) or larger than words (idioms) (Jackendoff, 2002). In the field of SLA, while acquisition of these two fundamental components of language has been studied very extensively, the interface between them has not been explored in depth (Austin, Blume & Sanchez, 2015). SLA research on acquisition of the lexicon has mainly focused on the acquisition of separate lexical items – mostly nouns as semantic meanings mapped onto a phonological form (Costa et al., 2003; Colomé and Miozzo, 2010; Dijkstra, 2005; Dijkstra & Van Heuven, 2002; Green, 1998; Kroll & Stewart, 1994; Kroll & Tokowitz, 2005; Hermans et al., 1998; Hermans et al., 1998; Lee & Williams, 2001; La Heji et al., 1996; Potter et al., 1984), and syntactic SLA research has predominantly pursued the question of whether Universal Grammar is available after the critical period (Bruhn de Garavito & White, 2002; Carroll, 1989; Dewaele & Véronique, 2001; Gess & Herschensohn, 2001; Granfeldt, 2000; Guillelmon & Grosjean, 2001; Franceschina, 2005; Hawkins & Franceschina, 2004; Montrul et al., 2008; White et al., 2004).

In this dissertation, I attempt to fill this gap by exploring the acquisition of gender - a phenomenon at the interface between lexicon and syntax – by second language learners of Spanish. Following Carstens (2000, 2010), I assume that grammatical gender feature is an inherent lexical feature on noun roots that triggers syntactic operations

within the nominal domain. Gender is an ideal linguistic feature to study the relationship between the lexical and syntactic representations, because it is a complex lexical-syntactic feature represented in multiple linguistic domains such as syntax, lexicon and morphology. In this connection, Corbett (1991) notes, “Gender is the most puzzling of the grammatical categories. (...) One of its attractions for linguists is that there are interesting aspects of the study of gender in each of the core areas of linguistics.” (p.1). Before I attempt to understand how such a complex linguistic phenomenon is acquired, I would like to clearly define and provide a comprehensive conceptual analysis of all aspects of the phenomenon. For this reason, in this dissertation I will make a selected review of previous works in several areas of research on language representation, access and acquisition – formal linguistics, psycholinguistics, neurolinguistics, SLA, and I will use some of the previous findings to inform my proposal on how to approach the second language (L2) acquisition of gender - a phenomenon at the lexicon/syntax interface.

The enterprise to explore gender and the interface between syntax and the lexicon in this dissertation was motivated by a large body of research on L2 acquisition of gender that has yielded evidence against acquirability of this feature after puberty (Carroll, 1989; Franceschina, 2005; Hawkins, 2009; Hawkins & Franceschina, 2004, *inter alia*). More specifically, gender was used to test availability of syntactic features, and some of these studies showed that L2 learners were unable to acquire gender. The difficulties that adult language learners have with gender agreement have caused some SLA researchers to conclude that the capacity to acquire syntax through access to Universal Grammar is unavailable in adults, unlike in children (Carroll, 1989; Franceschina, 2005; Hawkins, 2009; Hawkins & Franceschina, 2004). However, I will argue in this dissertation that

gender acquisition cannot be used as empirical evidence for such an L2 syntactic impairment theory.

It does not seem appropriate to argue that UG is or is not available based on data from gender acquisition. UG is a universal computational mechanism that derives potentially infinite *syntactic* structures, but it cannot ensure acquisition of all language-related (e.g., morphological or lexical) phenomena in any given language. Gender is not a purely syntactic feature, but rather an interface feature that is represented morphologically, syntactically and lexically; hence, its acquisition should hinge upon a number of factors, not exclusively the (un)availability of the UG. It may be pertinent to recall that UG was initially posited by Chomsky as a counterargument for the behaviorist notion that all learning takes place through stimulus (Chomsky, 1959). Chomsky showed that children produced linguistic structures as well as made errors that they could not have heard in the input, and that they acquired language in its entirety based on limited input. Simply put, not all language can be acquired from the input, but all of it is acquired; therefore, there must be *some* preexisting knowledge available from birth. In some previous SLA research, however, difficulties in gender acquisition has been taken as evidence of the unavailability of UG (Franceschina, 2005; Hawkins, 2009; Hawkins & Franceschina, 2004, *inter alia*).

Therefore, while I do not undertake to prove that UG is or is not available in the L2, I will suggest that the observed difficulties in the acquisition of gender agreement may be due to factors other than a specific critical period for acquisition of abstract syntactic features. In particular, I will argue that a lack of comprehension of a *surface* linguistic structure in a phrase involving gender such as (1) in a specific language

(Spanish in this example) cannot be interpreted as a lack of a syntactic representation, because understanding such a structure involves not only having abstract syntactic representations, but also having language-specific lexical and morphological knowledge.

(1) La casa blanca

DEF.F house-F white-F

“The white house.”

Moreover, I will show that a learner’s performance in tasks involving processing of syntactic structures in the second language can be hindered both by a lack of lexical knowledge as well as by transfer of the lexical knowledge from their L1. That is, even L2 learners whose L1 has gender may experience difficulties in L2 gender acquisition, and in their case the difficulties stem from transfer of L1 lexical knowledge. Recently, some SLA studies have also tied difficulties with performance on gender agreement to a lack of lexical knowledge (Grüter et al., 2013; Hopp, 2012; Sabourin & Stowe, 2008). These studies are highly informative and novel in their approach to gender acquisition, but they do not make explicit a specific theoretical foundation for the acquisition of lexical features. I will further explore this area and, crucially, I will attempt to provide a unified theoretical basis for the analysis of phenomena that involve lexical and syntactic features.

First of all, I will turn to linguistic theories of language structure (Chomsky, 1995, 2000, 2002; inter alia) to gain insights into how *syntactic* features are represented in the mind. These theories “have at their root the desire to account for linguistic facts of great sophistication, ..., and the relation of syntax to semantics, morphology, and phonology” (Jackendoff, 2013, p.149). They hold a fundamental assumption - the idea of a ternary distinction in all human languages in the sense that there are semantic and syntactic

concepts, a computational mechanism that derives syntactic structures (syntax) and some memorized phonological component that maps onto these syntactic structures. Most importantly, and this is one of the crucial ideas of this dissertation, recent versions of the syntactic theory (Hauser, Chomsky, & Fitch, 2002; Fitch et al., 2005) assume that the computational mechanism is the only innate and modular linguistic component, and that other components of language such as the lexicon may be acquired through general cognitive mechanisms, and not through access to the innate linguistic knowledge.

Second, I will turn to models of lexical retrieval in order to gain insights into how *lexical* features are represented in the mind (Bock and Levelt, 1994; Caramazza, 1997; Garrett, 1976; Levelt et al., 1999; Roelofs, 1992). These models do not make predictions about syntax as an innate computational mechanism, but rather mostly focus on the lexical representations, because they make hypotheses about the production of isolated words. According to these theories, word production involves an activation of lexical-semantic, lexical-syntactic and formal (lexical-morphological and lexical-phonological) representations of the word. For example, the lexical representations for the word *queso* ‘cheese’ are very briefly sketched below:

- a) Semantic features: INANIMATE, FOOD, MADE OF MILK,
YELLOW/WHITE, UNCOUNTABLE, etc.
- b) Syntactic features: CATEGORY: NOUN, GENDER: MASC,
UNCOUNTABLE
- c) Formal features:
 - Morphological form: <kes> <o>
 - Phonological form: [‘ke-so]

A prolific line of research on gender in psycholinguistics and neurolinguistics (Badeker et al., 1995; Biran & Friedman, 2012; Caramazza & Miozzo, 1997; Costa et al., 2003; Kulke & Blanken, 2001; Vigliocco et al., 1997, *inter alia*) has shown that these three types of lexical representation are stored separately from each other in relevant lexicons – the semantic lexicon, the syntactic lexicon and the phonological lexicon, because they can be impaired selectively in aphasic patients as well as retrieved discretely in monolinguals. For example, in studies with aphasic patients some of these patients were able to provide the gender of a word correctly in 95% of the cases, but they were unable to name or write the word (Badeker et al., 1995; Biran & Friedman, 2012; Caramazza & Miozzo, 1997, Vigliocco et al., 1997). This led the authors to conclude that the lexical-syntactic information is in fact stored separately from the lexical-phonological representation. In addition, these studies have shown that the lexical-semantic and the lexical-syntactic information is also represented in the mind as different types of knowledge, because some patients failed to perform gender agreement on nouns with lexical-semantic gender, but they performed 100% correct on nouns with lexical-semantic gender (Biran & Friedman, 2012).

These empirical findings led Levelt et al. (1999) to develop a computational theory of lexical access in word production called **WEAVER ++** (Word form Encoding by Activation and VERification), which was an extension to the **WEAVER** (Roelofs, 1997). The main assumptions of this theory are very similar to the assumptions about the organization of the lexicon sketched above. The three types of lexical features (semantic, syntactic and formal features) are stored in their respective strata – semantic feature stratum, syntactic feature stratum and formal feature stratum. According to Levelt et al.

(1999), the stratum that stores syntactic features (category, gender, number) is referred to as ‘lemma’ – the abstract syntactic representation of each noun of the language, whereas the stratum that stores formal features is referred to as ‘lexeme. Each noun of the language has one lemma and one lexeme. Furthermore, all of the syntactic features in the lemma of a given noun are connected to syntactic nodes in the lexicon. The syntactic feature ‘category’ of all nouns will be connected to specific category nodes of this language. There are as many category nodes as categories exist in this language (e.g., noun node, verb node, adjectives node, adverb node, preposition node, etc.). Similarly, the syntactic category ‘gender’ of all nouns of the language will be connected to gender nodes of that language, and there will be as many gender nodes as genders in the language. For example, there will be two gender nodes in Spanish (masculine and feminine gender nodes), whereas in Russian there will be three gender nodes (masculine, feminine and neuter). Lemma of each noun in a given language will be connected to one of the gender nodes existing in that language; in a language that lacks gender (e.g., English), on the other hand, there will be no gender features in the lemmas and no gender nodes in general.

Native speakers of a language like Spanish or Russian remember to which gender node each of the lemmas of a language are connected, and in speech production they automatically retrieve the gender value from the lemma by accessing the gender node to which the lemma is connected. That is, they do not establish the gender of a noun based on its morphological makeup (e.g., ‘look up’ the word ending), but rather retrieve it from the abstract lexical representation of the noun.

In general, as it can be seen, although the syntactic theories mostly investigate phrase-level syntax and the lexical retrieval theories mostly focus on isolated words, they both share an essential assumption that there are three types of features/representations – semantic, syntactic and phonological features.

Based on the review and the analysis of the linguistic theories and of the models of lexical retrieval, I will make several crucial distinctions with respect to gender. First of all, I will distinguish between gender on nouns and gender on other elements of the DP. Gender on nouns is both a lexical and a syntactic feature, while gender on other elements of the DP is a syntactic, not a lexical feature. This is so because gender features on nouns come from the lexicon with a specific value, whereas adjectives and determiners only obtain a value as a result of the syntactic operation gender agreement. More specifically the noun *carro* ‘car’ comes from the lexicon with a masculine value in Spanish, but an adjective such as *rojo* ‘red’ is underspecified for a gender value and receives it from the noun to which it is related syntactically.

Second, in addition to the distinction between lexical-syntactic gender features on nouns and syntactic gender features on the other elements of the DP, I will further differentiate between two lexical representations of gender on the noun – lexical-*syntactic* gender and lexical-*semantic* gender (sex). Both are represented lexically, but while semantic gender is a feature present on animate nouns such as boy vs. girl or uncle vs. aunt and stored in the *semantic* lexicon, lexical-syntactic gender is a categorization feature that classifies nouns into classes such as F or M (or other) and stored in the *syntactic* lexicon (lemma – the syntactic representation of each lexical item that is connected to a gender node) (Levelt et al., 1999). Whereas lexical-syntactic gender

features have to be acquired in the L2, lexical-semantic gender features (he/she distinction) are shared across languages, and thus should not present acquisition difficulties.

Third, I will distinguish between lexical-syntactic gender values such as F and M, on the one hand, and morphological gender markers such as *-a* for feminine and *-o* for masculine in Spanish, as is shown in (4) and (5) below. Lexical-syntactic gender values are an abstract lexical representation of a noun gender, while morphological markers on nouns arguably are a concrete manifestation of this representation, although some (Harris, 1991) do not consider these markers as gender markers but rather as word class markers. As it was mentioned above, according to WEAVER++ (Levelt et al., 1999), native speakers of a gendered language memorize gender values for *all* nouns of the language. They do not rely on morphological gender markers for the purposes of word production, because they store and retrieve the gender value of a given noun through accessing the respective gender node from the lemma of the noun, without having to retrieve morphological gender markers.

On the other hand, during *comprehension*, native speakers do use morphological markers as a cue to retrieve the gender of nouns, because they retrieve the gender of nouns with transparent morphemes (the ones that are informative with respect to gender) faster than the gender of nouns with opaque gender morphemes (Andonova et al., 2004, for Bulgarian; Bates et al., 1995, for Italian; Cacciari et al., 1997, for Italian; Leinbach et al., 1989, for German; MacWhinney et al., 1977, for French; Sekerina et al., 2005, for Russian). Even more robust results were obtained for unbalanced bilinguals (Bordag, 2004, 2006, for L1 Czech L2 German; Oliphant, 1998, for L1 English L2 Italian; Taraban

& Kempe, 1999, for L1 English L2 Russian; Taraban & Roark, 1996, for L1 English L2 French). Therefore, during comprehension (and possibly during production), *L2 learners* at the beginning stages of L2 acquisition should predominantly rely on morphological markers on the nouns, because they have not incorporated all of the gender values of all nouns in the lexicon and hence they cannot activate gender values through the lemma. If that is the case, morphemes that are more transparent with respect to gender should serve as a better cue to acquire the gender value of the noun.

The three above-mentioned distinctions lead to the fourth and the most crucial distinction in this dissertation - the one between the “abstract syntactic operation gender agreement” and the “surface gender agreement”. I use the phrase ‘abstract syntactic operation gender agreement’ in the generative theory sense - to refer to the abstract syntactic representation of a DP as a constituent whose elements share abstract features and feature values that are void of phonological content, and the phrase ‘surface gender agreement’ to refer to the speakers’ ability to fill the abstract syntactic structure with concrete gender morphemes of a specific language according to their gender values while processing DPs in that language. The distinction between the abstract syntactic operation gender agreement and the language-specific surface gender agreement is exemplified in (2), where GenX stands for a gender X which is shared across the elements of the DP, and in (3), where the noun *casa* ‘house’, the adjective *blanca* ‘white’ and the determiner *la* ‘the’ bear a feminine value which is marked through a regular feminine gender marker *-a*.

(2) D_{GenX} N_{GenX} A_{GenX}

(3) L_a casa blanca

DEF.F house-F white-F

“the white house”

The abstract features drive the abstract syntactic operation gender agreement in any language that has gender, while the lexical and morphological instantiation of these features varies from language to language (Spanish has two gender classes while Russian has three, *o*-ending in Spanish often marks masculine while in Russian it marks neuter, etc.).

In summary, the gender feature is a highly complex feature that is represented as an abstract *lexical* feature on nouns, as an abstract *syntactic* feature on all of the DP-elements and is manifested through language-specific feature values like "F", "M", "N", and morphophonologically through language-specific gender markers. The following table sketches the levels of gender representations and which parts are abstract and which are language-specific.

Table 1. *Gender feature representation levels*

GENDER	
Abstract level:	Language-specific level:
Lexical: gender X gender Y	X=fem Y=masc
Syntactic: N _X A _X D _X N _Y A _Y D _Y	Formal (morphophonological): N _O A _O D _O N _A A _A D _A

In this dissertation, I will raise research questions and put forward hypotheses based on these four essential distinctions – the distinction between the lexical-syntactic gender feature on nouns vs. syntactic features on other elements of the DP, the distinction between the lexical-syntactic gender features and the lexical-semantic features on nouns, the distinction between lexical-syntactic gender features on nouns and their concrete

morphological manifestation, and the distinction between the abstract syntactic gender agreement and the language-specific surface gender agreement.

I will suggest that the acquisition of surface gender agreement depends on having the abstract syntactic gender agreement, on the one hand, and having acquired language-specific knowledge, on the other hand. Having the abstract syntactic operation gender agreement is the first and possibly an indispensable step in the acquisition of gender, because it allows one to represent the Determiner Phrase (DP) as a unit whose elements share features and values, which subsequently allows the learners to attend to and process morphological markers that the unit exhibits. Processing morphological markers will subsequently lead to assigning appropriate gender values to all nouns of the language. After a significant amount of exposure and practice in the L2 the access to the gender values of nouns should become automatic, and thus indistinguishable from that of the native speakers.

Hence, having the abstract operation gender agreement makes it possible for a language learner to attend to the morphological markers that a syntactic unit exhibits, attending to morphological markers, in its turn, makes it possible for them to incorporate lexical gender values of nouns into the lexicon, and having these values in the lexicon should lead to a target-like gender value retrieval and, hence, to a target-like surface gender agreement. Thus, the first cue to start acquiring gender for L2 learners (and in fact for L1 learners, too) is the morphological cue. In the next paragraph I will discuss the morphological characteristics of the Spanish gender system with an intent to suggest which specific morphemes should be acquired faster and lead to a more accurate surface gender agreement in the L2.

It has been classically assumed that the morphological marking in Spanish generally follows a formal rule according to which nouns that end in *-o* are masculine while nouns ending in *-a* are feminine, as exemplified in (4) and (5).

(4) El ques-o blanc-o

DEF.M cheese-M white-M

(5) La cas-a blanc-a

DEF.F house-F white-F

However, the masculine “*o*” vs. feminine “*a*” rule is not universal. There are two types of gender markers that do not follow this rule. First, there are morphemes that *contradict* this rule, where nouns with *-o* ending actually carry a feminine lexical gender value and, thus, share the feminine value with the modifying elements, and nouns with *-a* ending have masculine gender values, and share this value with the rest of the DP, as is exemplified in (6) and (7).

(6) La man-o derech-a

DEF.F hand right-F

“The right hand”

(7) El dí-a caluroso

DEF.M day hot-M

“The hot day”

Second, there are *ambiguous* gender morphemes such as *-e*, *-u* and *-d*, *-r*, as in (8), (9), (10) and (11), that can have both feminine and masculine gender values.

(8) La leche

DEF.F milk.F

“the milk”

(9) La tribu

DEF.F tribe.F

“the tribe”

(10) La edad

DEF.F age.F

“the age”

(11) El color

DEF.F color.M

“the color”

Some of these gender markers are not completely arbitrary, because they follow a pattern of subregularity, whereas other markers are completely uninformative with respect to gender. For example, the endings *-ad* and *-ción* always mark feminine gender, as in (12), (13), (14), and (15), while endings *-e*, *-z* and *-is* can mark both feminine, as in (16), (18), (20) and masculine, as in (17), (19), (21).

(12) la universidad

DEF.F university.F

“the university”

(13) la igualdad

DEF.F equality.F

“the equality”

(14) la inspiración

DEF.F inspiration.F

“the inspiration”

(15) la reclamación

DEF.F claim.F

“the claim”

(16) la consonante

DEF.F consonant.F

“the consonant”

(17) el determinante

DEF.M determiner.M

“the determiner”

(18) la paz

DEF.F peace.F

“the peace”

(19) el disfraz

DEF.M disguise.M

“the disguise”

(20) la dosis

DEF.F dose.F

“the dose”

(21) el análisis

DEF.M analysis.M

“the analysis”

While the uninformative gender markers such as in (16)-(21) should be the most difficult to acquire, the subregular gender markers may also present difficulties for lower proficiency L2 learners. These learners have a tendency to assign masculine gender to all nouns that end in a consonant, thus producing gender agreement errors such as *el revolucion* ‘the.M revolution.F’ and *el universidad* ‘the.M university.F’ (personal observations). Even *-a* for feminine and *-o* for masculine endings are not fully reliable, precisely because there are nouns that have contradicting gender markers such as (6) and (7); hence it is more appropriate to refer to this rule not as a *rule* but rather as a general *pattern*. This means that there is a general ‘*-a* for feminine vs. *-o* for masculine’ pattern, and there are subregular patterns such as *-ción* for feminine. However, since the nouns that follow feminine *-a* vs. masculine *-o* pattern by far outnumber the nouns with contradicting gender markers, L2 learners tend to overgeneralize this rule to all nouns that have *-a* or *-o* at the end, and generally perform better on nouns that follow this pattern.

It should be mentioned here that although the feminine vs. masculine distinction based on the *-a* and *-o* endings respectively for Spanish is classical in the literature, it is not supported unanimously. Harris, for example, argues very strongly that this distinction is *hopelessly simplistic* (Harris, 1991, p. 32). He maintains that these endings are two of several other word markers (that is, not only noun markers, but also adjective and even adverb markers) that belong to declensional classes. As for gender specifically, Harris categorizes nouns into 3 gender classes: an inner core of prototypes with *-a* as a feminine and *-o* as a masculine marker, an outer core of words of slightly aberrant cases and a

residue that contains all words not in the core which he refers to as *motley* (Harris, 1991, p. 32).

While the theoretical issue of declensional class vs. gender raised by Harris is beyond the scope of this dissertation, his categorization of nouns is relevant for its purposes. In particular, I will suggest that the acquisition of gender of the inner core of prototypes (such as in (4) and (5)) is facilitated by their (relative) morphological transparency, while the acquisition of the outer core and the residue is hindered by the unreliability of the morphological markers. In order for the correct gender values to be assigned to nouns with unreliable gender markers, L2 learners will either have to internalize the patterns of subregularity (for nouns that exhibit such patterns), and these are more challenging than the general *-a* vs. *-o* pattern, or simply memorize the gender values of specific nouns whose values are completely unreliable. In both cases, the acquisition of gender assignment will be exacerbated compared to the acquisition of the (relatively) reliable gender markers. This does not mean, however, that the learner's linguistic system lacks the *abstract syntactic operation agreement*.

Syntax may be actively combining gender features on nouns with determiners and adjectives, but the language-specific gender values may be absent from the lexicon, and in cases where gender morphemes are not reliable, no cues would be available to establish the gender value of the noun, which would result in an incorrect *surface* gender agreement. Therefore, L2 learners should perform surface gender agreement more accurately on nouns that have reliable morphological forms, although the abstract syntactic operation gender agreement could be active for all items. This will reveal that the process of incorporating lexical-syntactic gender values into the lexicon is facilitated

by morphological cue reliability.

If morphological markers are cues for the acquisition of lexical values, what other cues could potentially be used to determine the lexical-syntactic gender of nouns? One may expect that the lexical-semantic gender could provide such a cue, because many lexical-semantic features are shared across languages. So, when L2 learners hear the word *chica* ‘girl-F’, they may assign a feminine value to it faster and more easily than when they assign the same value to the word *casa* ‘house’, because the word ‘girl’ refers to a feminine entity in any language, while the word ‘house’ does not have any conspicuous inherent features in their L1 that would help classify this word as “feminine” in their L2. Alternatively, L2 learners of Spanish may prioritize lexical-syntactic and lexical-phonological cues to lexical-semantic cues, similarly to L1 learners. Spanish-speaking children have been shown to prioritize intralinguistic (gender on determiners and gender morphemes on the noun) information over extralinguistic information (semantic gender). In general, semantic gender is not acquired before lexical gender (Socarras, 2011). The same pattern holds true for Russian children – they prioritize morphological rules over semantic rules (Ceytlin, 2009; Popova, 1973). For example, if a noun has the typical feminine gender ending –a, it will be assigned feminine gender even if it refers to a male entity. This finding lends support to the idea of modularity of syntax – although children have not yet acquired semantic gender distinctions, they already perform syntactic gender operation. If L2 learners have the same preference, it will indicate that L2 learners exhibit similar developmental patterns.

Whereas L2 learners whose L1 does not have gender should rely predominantly on morphological characteristics of nouns and possibly on lexical-semantic gender to

acquire lexical-syntactic gender values, L2 learners whose L1 has gender already have gender values in their L1, and their gender agreement may be affected by these L1 lexical values *even in the presence of the abstract syntactic gender features*. This is so because the lexicon of such L2 learners will already have specific lexical-syntactic features on nouns, and these features may differ from those in the L2. For example, the noun ‘cup’ has a feminine value both in Spanish (*taza.FEM*) and Russian (*chashka.FEM*); that is, it has *congruent* gender values in Spanish and Russian. Conversely, the noun ‘house’ is feminine in Spanish (*casa.FEM*) and masculine in Russian (*dom.FEM*); that is, it has *incongruent* gender values in the two languages.

Thus, although Russian L1 Spanish L2 learners may transfer the abstract syntactic operation gender agreement to the L2, their surface gender agreement should be affected by incongruency, because the L1 and not the L2 lexical gender values would be distributed among the elements of the DP, and in cases where the values are incongruent, they would lead to a non-target-like surface gender agreement. Thus, contrary to what many SLA researchers have assumed, it is possible that even speakers of an L1 with gender may have problems acquiring gender in their L2.

This question about the (un)availability of the UG after puberty has been one of the central topics in the generative SLA, and while some theories propose that UG is available for adults (Schwartz & Sprouse, 1996), others claim that it is only available for L1 learners (Hawkins & Chan, 1997; Hawkins, 2009) and others suggest it is partially impaired (Eubank, 1994; Vainikka & Scholten, 1996).

Outside of the generative SLA field, there is yet a more general discussion between the proponents of the generative theories, on the one hand, and proponents of

non-generative theories such as Connectionism - an umbrella term for the theories that explicitly deny any innate linguistic capacity (Ellis, 1998). Lidz and Gagliardi (2015) refer to the non-generative approach to language acquisition as *input-driven* tradition, because the main driving force behind any language learning in this approach is the quantitative and qualitative properties of input.

In addition to demonstrating that non-target like performance on surface gender agreement is not an appropriate argument to claim unavailability of UG in L2, I indirectly explored the connectionist idea that input is the main driving force for acquisition of both L1 and L2 and hence that there is no need to postulate a specific innate ability to acquire language (Ellis, 1998; Mariscal, 2008). There is research that shows that the quantity and the quality of input are reliable predictors of gender acquisition in child L2 and adult L2 (Unsworth, 2008; Unsworth et al., 2012). However, I suggest that such a finding is not incompatible with the idea of UG because, as it has been mentioned, UG provides the abstract syntactic operations such as Merge and Agree and endows us with syntactic productivity and creativity, but it does not provide language-specific gender values and morphemes. Since gender is partially a lexical feature, and lexical features are stored in the long-term memory, its acquisition has to correlate with the quantity and the quality of input, and even to a larger extent with the intake and processing for comprehension and production (Putnam & Sanchez, 2013).

The idea that gender is a lexical feature was expressed in Unsworth (2008), “This is to be expected if the acquisition of gender is for a large part word-learning” (p. 365). It needs to be emphasized, however, that while acquisition of gender involves learning lexical and morphological knowledge, it would hardly be possible for L2 learners to learn

this knowledge if they did not represent a DP as a syntactic unit whose elements share certain features and values. This idea can be tested in the following way. If L2 learners are capable of representing the DP as a syntactic unit whose elements share features and to perceive gender morphemes as mapping onto the elements of this syntactic unit, as soon as the lexical-syntactic value for a certain noun has been incorporated into the lexicon, the surface gender agreement should potentially be performed on all elements of the DP, regardless of how often the learners encounter these elements in the input.

More specifically, although agreement between the definite article and the noun is the most frequent in the input (Mariscal, 2008), it should not be more accurate than the agreement between the noun and other elements of the DP, as long as these elements are considered DP-elements in the interlanguage grammar. Mariscal (2008) claimed that children only exhibited target-like agreement on the most frequent DP types – the sequences of a definite article followed by a noun, and that they made errors or refrained from using other types of DPs such as noun/adjective sequences or the sequences of possessives/demonstratives followed by a noun. She used this as evidence against the idea that children acquire gender agreement through access to UG. Although I certainly do not undertake to resolve the long-standing generative vs. non-generative debate, I will test Mariscal's idea on L2 learners by comparing their performance on definite article/noun sequences with their performance on noun/adjective sequences.

In conclusion, the main research pursuit of this dissertation is to show that surface gender agreement is not an appropriate phenomenon to claim unavailability of abstract syntactic features. For this, I will show that both L2 learners whose L1 has gender and L2 learners whose L1 does not have it can be affected by the lexical-

morphological characteristics of gender in the L2. Crucially, I will also show that once the hindering effect of these lexical-morphological characteristics is eliminated, gender agreement can be target-like in L2 learners, both in those learners whose L1 has gender and in those whose L1 does not have it.

Thus, the research questions of this dissertation are as follows:

Research question 1: Are abstract syntactic gender features acquirable in the L2 if they were not activated in the L1?

Research question 2: Does the lack of exposure to abstract *syntactic* gender features in the L1 compromise gender acquisition in the L2, as it is maintained in Hawkins and Chan (1997), Hawkins and Francheschina (2004), Francheschina (2001), and Hawkins (2009)?

Research question 3: Does the exposure to the abstract *syntactic* gender features in L1 facilitate gender acquisition in L2?

Research question 4: Is surface gender agreement affected by the lexical-morphological complexity of the L2 gender system (cue unreliability)?

Research question 5: Is surface gender agreement affected by a presence of preexisting L1 lexical knowledge (L1/L2 congruency), even if the abstract syntactic gender features are available?

Research question 6: Do lexical-semantic gender values (as in *chica* ‘girl-F’ vs. *chico* ‘boy-M’) facilitate the acquisition of lexical-syntactic gender values of these nouns compared to the acquisition of lexical-syntactic gender values of the nouns that lack lexical-semantic gender values (as in *casa* ‘house-F’ vs. *queso* ‘cheese’)?

Research question 7: Is gender agreement more target-like on more frequent DP-types (definite article followed by a noun) than on less frequent DP-types (noun followed by an adjective)?

I conducted two experiments in order to address the research questions. Thirty eight L1 Russian L2 Spanish speakers (19 low proficiency and 19 high proficiency) and 35 L1 English L2 Spanish speakers (19 low proficiency and 16 high proficiency) participated in Experiment 1, where they were asked to perform a speeded Grammaticality Judgment Task as well as a speeded Picture-Matching task in order to find out the extent with which the existing L1 lexical representations affect the syntactic operation gender agreement in L2. On the basis of the results obtained in this study, I challenge the widespread assumption that having a syntactic operation in L1 necessarily facilitates performance on that operation in L2 by showing that speakers of a gendered L1 (Russian) may not have an advantage over speakers of an L1 that lacks gender (English) as far as gender agreement is concerned, precisely due to the interference from the lexical representations. Another highly relevant finding is that, contrary to what Mariscal (2008) claimed for L1 gender acquisition, the amount of input does not seem to predetermine the L2 acquisition of gender agreement between different elements of the DP.

In the second study, 33 English speaking L2 learners of Spanish (16 high proficiency and 17 low proficiency) and 16 native controls performed a timed grammaticality judgment task that included nouns that have lexical-semantic gender vs. nouns that do not have it as well as nouns with reliable vs. unreliable gender morphology. The study was designed to reveal the effect of the lexical-semantic gender and the effect of morphological marking on surface gender agreement as well as to determine whether

L2 learners could potentially perform at a target-like level once the hindering effect of the lexical component of gender was neutralized. It was also designed to show that the lexical representations are not acquired as a whole for a given language, but rather that the gender values of specific nouns can be acquired earlier or later depending on their lexical and morphological profile. As I will discuss later in the dissertation, these findings provide evidence for how closely intertwined the lexical and the syntactic aspects of gender are.

The overall results will be interpreted as supporting the view that both previously existing lexical knowledge and the lack thereof affect performance on gender. This is because L1 English learners of Spanish were more accurate on nouns with reliable gender markers and because L1 Russian learners of Spanish were more accurate on nouns with congruent gender values. This effect decreased significantly with proficiency. Importantly, based on the results of the two studies, I will argue that the abstract syntactic operation gender agreement *per se* is not impaired in the L2, because high proficiency L1 English learners behaved like controls virtually on all tasks both in terms of accuracy and reaction times, despite the fact that their L1 does not have gender features. Moreover, having gender in L1 does not necessarily facilitate gender acquisition in L2, because L1 Russian learners did not outperform L1 English learners of Spanish on any of the tasks. In broad terms, the findings inform and supplement the current understanding of the way language in general and gender in particular are represented and acquired.

The dissertation will be organized as follows: in the second chapter, a review of the existing research relevant for gender acquisition will be provided; in the third chapter, the methodological design of the studies will be described; in the fourth chapter, the

results from the studies will be presented followed by the last chapter where the results will be discussed.

CHAPTER 2: Gender as an interface feature

One of the most crucial questions on research agenda in the field of SLA is the issue of acquirability of syntactic features in L2. A number of researchers have concluded that such features are not available after a critical period, and suggested that this syntactic impairment could account for the presumed inability of L2 learners to acquire such syntactic features as gender (Hawkins, 2009; Hawkins & Chan, 1997; Hawkins & Francheschina, 2004; Franceschina, 2001; inter alia.).

In this dissertation, I study the relationship between lexical, syntactic and morphological components of gender in order to show that gender is not an ideal candidate to test availability of syntactic features, because performance on surface gender agreement may be affected by a number of factors, not exclusively (and not necessarily) by a lack of syntactic features. As noted by Grüter et al. (2012) in a study on gender acquisition, high proficiency L1 English learners of Spanish made 10 times as many gender assignment errors as gender agreement errors, which led the authors to conclude that, “The observation that L2 learners with advanced to near-native proficiency appear to experience more persistent problems with gender assignment than agreement suggests that the lexical representation of grammatical gender, and its acquisition in L1 development, should be reconsidered in some detail.” (p.209). This means that performance on gender may be hindered by the factors that affect gender assignment. In this dissertation I explore these factors, and attempt to build a solid theoretical framework

for the acquisition of complex features that are represented in multiple linguistic domains – syntax, lexicon and morphology.

I define gender as a complex lexical-syntactic feature marked morphologically on nouns and other DP-elements in gendered languages. It is represented lexically as a categorization feature that classifies nouns as F, M, and N in most languages. It is lexical, because it forms an intrinsic part of the lexical representation of the noun since the lexical values such as F and M are memorized and stored in the syntactic lexicon (the syntactic lexicon stores lexical features that have effect on a phrase level; these are lexical-syntactic gender and word class) (Levelt et al., 1999). This inherent lexical-syntactic feature on nouns enters syntactic derivations and shares its values with other DP-elements, which is why gender is also a syntactic feature. In this chapter, I will provide an overview of gender as a syntactic feature in the Minimalist program (Chomsky, 1995, 2000; Pesetsky & Torrego, 2004), of the different levels of lexical representation involved in gendered nouns postulated by theories of lexical retrieval (Levelt et al., 1999) and of the main findings in studies of L1 and L2 acquisition of gender (Alarcón, 2011; Bottari, Cipriani & Chilosi, 1993; Clark, 1986; Franceschina, 2001; Grüter et al., 2012; Hawkins, 1998; Hawkins & Franceschina, 2004; Hernandez Piña, 1984; Hopp, 2012; Karmiloff-Smith, 1979; Lopez-Ornat, 2003; Mazeika, 1973; Mariscal, 2008; Pizzuto & Caselli, 1992; Sabourin & Stowe, 2008). I will conclude with the hypotheses I will be exploring in the dissertation.

Gender in Minimalism

As it was mentioned in the Introduction chapter, I turn to generative linguistic theories to make theoretical assumptions about the syntactic component of gender. However, it should be noted here that not only do these theories represent a powerful source of information about syntax as a core linguistic property of the language faculty, but they also make predictions about the relationship between syntax on the one hand and phonology, semantics and the lexicon on the other hand. In this dissertation, I will discuss and adopt the Minimalist Program perspective (Chomsky, 1995, 2000, 2002) as a framework for the study of the acquisition of gender by second language learners.

According to the Minimalist Program (Hauser, Chomsky & Fitch, 2002; Chomsky, 1995, 2000, 2013), there are three components of the human language faculty in its broader sense: “a sensory-motor system, a conceptual-intensional system, and the computational mechanisms for recursion, providing the capacity to generate an infinite range of expressions from a finite set of elements” (Hauser, Chomsky & Fitch, 2002, p.156). Simply put, the sensory-motor system deals with sounds, the conceptual-intensional system deals with meaning and the computational mechanism serves the two systems by deriving syntactic structures. Importantly, out of the three components it is only the computational mechanism – syntax - that is innate, modular and uniquely human and what is referred to as Universal Grammar in the modern generative theory (Hauser et al., 2002; Fitch et al., 2005).

Crucially for the purposes of this dissertation, syntax performs computations based on the input provided by the lexicon, which is defined as the storage of bundles of phonological, semantic and syntactic features such as gender, animacy, person, etc. The bundles of features represent lexical items, all of which have a category label such as N

(noun), V (verb), A (adjective), and act as a whole when they enter the syntax. Syntax can only “see” features that are relevant for syntactic operations such as agreement features (gender and number), but it cannot “see” semantic and phonological features, because these are irrelevant for the generation of syntactic structures. In other words, syntax is not responsible for anything but the generation of abstract syntactic structures devoid of phonological content and irrespective of semantic meanings (Hauser, Chomsky & Fitch, 2002).

The computational system (syntax) derives structures through two basic syntactic operations: Merge and Agree (Chomsky, 2000, 2001; Pesetsky & Torrego, 2004). Merge combines two syntactic elements into one and does it recursively to create (potentially infinite) complex structures. For example, it combines N with a D to derive a DP. With respect to Agree, I will not assume the classic Minimalist analysis put forward in Chomsky (2000, 2001) where uninterpretability of features drives Agree operation, but will rather assume Pesetsky and Torrego (2004)’s analysis, according to which Agree is driven by unvalued features as a result of which it assigns values to unvalued features in the course of the syntactic derivation. The same feature can be distributed among different elements at different syntactic locations. It can be so that while one instance of a feature is valued, another instance is unvalued. In this case, the unvalued feature ($F []$) will have to find the identical, but valued feature (F_{val}) to enter in an Agreement relationship with it, to obtain the value from it and eventually to delete. The unvalued feature is said to probe the goal – the valued feature. The operation is represented in (22):

(22) $F []$ F_{val} F_{val}

probe goal



Agree operation

The masculine, feminine, and neuter gender values of nouns are exemplified in (23), (24) and (25):

(23) libro (Spanish)

book-M

“book”

(24) ventana (Spanish)

window-F

“window”

(25) okno (Russian)

window-Neut

“window”

Modifiers and determiners within the DP have gender features as well, but these are unvalued; hence, they search their c-command domain for an element with the same, but valued, feature. The first such element they encounter is the noun. They probe the valued gender feature in order to get valuation and delete as shown in (26):

(26) *Gen* [] *Gen* val *Gen* val

probe

goal



Agree operation

This probe-goal relationship between the gender feature on the noun and the gender feature on its modifiers and determiners is referred to as Concord and represents a type of Agreement (Baker, 2013). (27) - (31) are examples of Concord:

(27) es_a taz_a blanca (Spanish)

this-F cup-F white-F

“this white cup”

(28) eta belaya chashka (Russian)

this-F white-F cup-F

“this white cup”

(29) es_e suelo_o negro (Spanish)

this-M floor-M black-M

“this black floor”

(30) etot chyorniy pol (Russian)

this-M black-M floor-M

“this black floor”

(31) eto_o bolshoye_e okno_o (Russian)

this-N big-N window-N

“this big window”

The three syntactic trees in Fig.1-3 below show the way in which the DP is represented in the grammars of Spanish and English native speakers before any movement operations have taken place. I assume that in all three languages the DP hosts a number of FPs, one of which takes the NumP as a complement. I will follow Bernstein (2001) and Cinque (2010) and assume that Russian and Spanish adjectives are generated

in the Spec position of designated functional categories within the DP higher than the NumP. These functional categories are semantically determined and are aligned according to the fixed hierarchy of adjectives proposed by Cinque (1994). Distinct word order in Spanish (Det-N-A) and in Russian (Det-A-N) is accounted for by assuming that Russian exhibits the universally unmarked word order (Bernstein, 1993) whereas the Spanish Det-Noun-Adjective word order is the result of NP-movement to a position higher than the FP hosting AdjP (Cinque, 2010). It is worth mentioning here that the research hypotheses proposed in the current study should not be affected by the differences in word order in the two languages under consideration, because the phenomenon studied here is agreement and not movement/word order acquisition. The NP is a complement of the NumP in this analysis, although positing a specific projection for number is not a matter of principle here. As can be seen in Fig.1 and Fig.2, in Russian and Spanish both the adjectives and determiners have unvalued gender features. These are the features that probe the gender feature on the noun and thus motivate gender agreement. An important difference between the two languages is that while the Spanish gender system has two gender classes, Russian has neuter values in addition to feminine and masculine, and these lexical values in a large number of cases are not congruent in Russian and Spanish.

There is an ongoing debate as to whether article-less languages like Russian include a DP (Boškovic, 2008; Caruso, 2011; Rappaport, 2000). Boškovic, for example, argues that languages without articles differ fundamentally from languages with overt articles, and claims that only the latter, and not the former, give evidence for a functional layer DP (2008). However, although Russian lacks articles, its nominal domain includes a

range of elements such as demonstratives, possessive pronouns and quantifiers (Rappaport, 2000). Moreover, Franks & Pereltsvaig (2004) argue that a DP layer in Russian is necessary to bear morphological case. This indicates that there has to be *some kind* of a functional layer above the NP that would provide a locus for those elements and for their formal features. Therefore, I will assume that there is a functional projection above the NumP/NP. I will not claim this functional projection to be identical to the DP-layer in languages with articles, but for the sake of simplicity I will refer to it as DP.

Thus, Spanish and Russian are similar with respect to the syntactic operation Concord (type of Agree), but not with respect to the language- specific lexical values. English, on the other hand, is what is called a pronominal gender language (Audring, 2008), because the only gender-marked items in English are pronouns, and their gender is based on semantic principles of natural sex, not on formal rules. English nouns are not labeled lexically as having a certain gender value, and they do not enter in gender-related syntactic relationships with their modifiers. The syntactic tree in Fig. 3 sketches the DP representation in native speakers of English.

Figure 1: *DP-structure in Russian*

Figure 2: *DP-structure in Spanish*

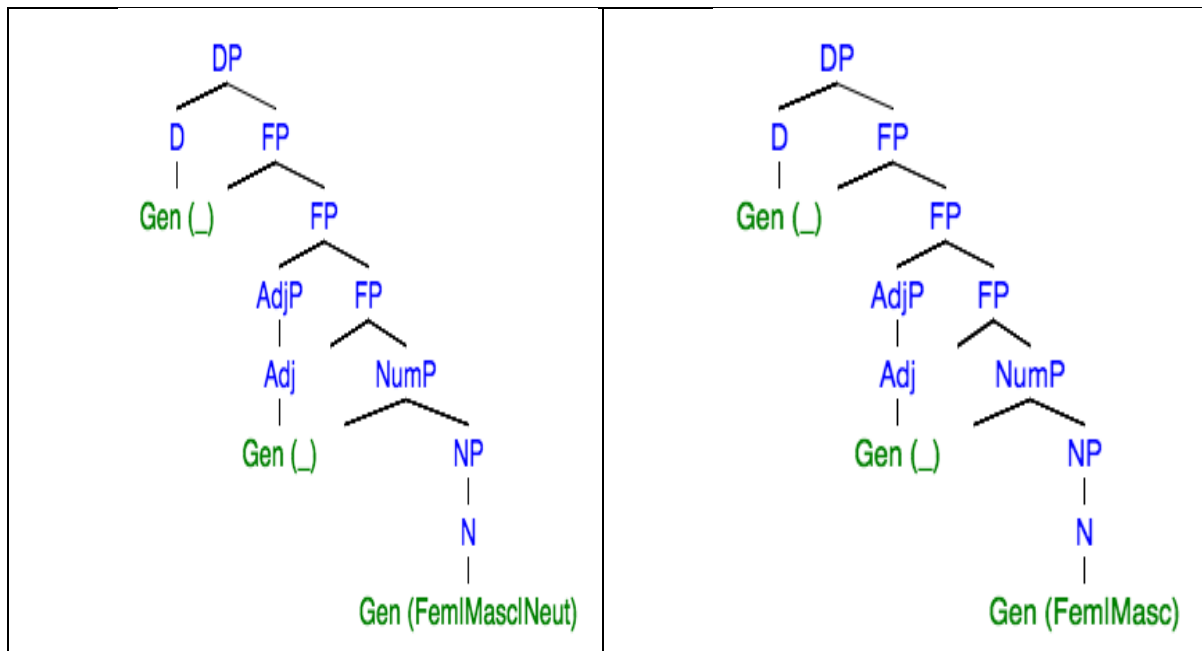
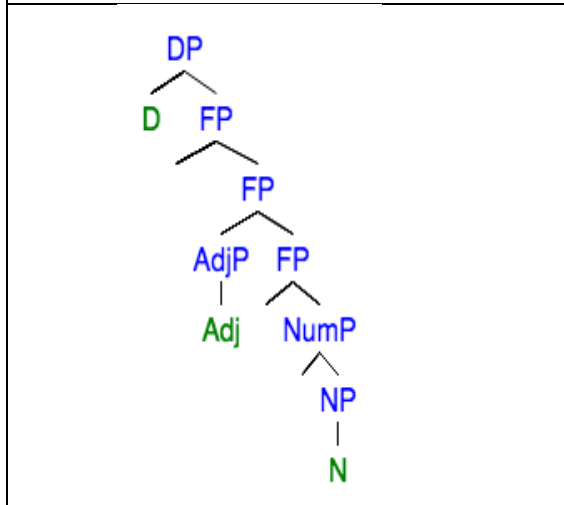


Figure 3: *DP-structure in English*



Thus, while Spanish and Russian nouns enter the syntactic structure with a gender value and the adjectives and determiners have unvalued gender features awaiting valuation in the course of Agree operation, English nouns do not bear any gender values and determiners and adjectives have no unvalued features to drive probe-goal

relationships with nouns. However, Russian and Spanish nouns often do not match in gender values, which can potentially lead to an incorrect gender value being shared in the course of the syntactic operation gender agreement.

In the next section I will provide an overview of theories of lexical retrieval to establish the basis of how gender as a syntactic feature relates to the lexical configuration of nouns.

Gender in Theories of Lexical Retrieval

Theories of lexical retrieval (Bock and Levelt, 1994; Caramazza, 1997; Garrett, 1976; Levelt et al., 1999; Roelofs, 1992) have focused on the question of how the lexicon is organized and retrieved for comprehension and production purposes. More specifically, they identify components of the lexical representation and they attempt to model the way in which information flows from one component to another in native speakers. They provide the bases to our understanding of how language learners access and acquire these components, and, importantly for this dissertation, provide insights into the process of acquisition of lexical-syntactic features such as gender. In this section, I will present the main proposals about how gender is integrated in the lexicon and the main findings of lexical retrieval studies conducted with monolingual and bilingual populations.

With respect to the components of the lexical representation, these models suggest that there are three stages of lexical retrieval and, therefore, of lexical representation. First, a conceptual representation is formed depending on the speakers' communicative intention, then, it is shaped into a lexical-semantic representation by matching bits of the

communicative message to separate semantic concepts and, lastly, the morphological and phonological (segmental and metrical) information about the word is activated (Bock & Levelt, 1994; Levelt et al., 1999; Roelofs, 1992). Importantly for the purposes of this dissertation, there is a stage between the lexical-semantic and the lexical-phonological stages that captures syntactic information. This level of representation is referred to as the “lemma” level. It includes information about the grammatical category of the item (N, V, A, etc.) and about features such as gender - a type of syntactic information that is language-specific (Garett, 1992; Levelt, 1989). As mentioned in the Introduction chapter, information on syntactic categories and features such as gender stored in the lemmas of all lexical items and are linked to specific category and gender nodes (Roelofs, 1992; Levelt, 1999). The lexicon will have as many category nodes as available in the language and as many gender nodes as gender values are available in the language. The notion of nodes is similar to the notion of formal features in generative theory, because both are abstract and both drive syntactic operations. So, the Spanish lexicon will have two gender nodes – one for masculine and one for feminine gender values, while the Russian lexicon will have three gender nodes – for masculine, feminine, and neuter gender values. A brief sketch of the types of features that a lexical item such as *abuela* ‘grandmother’ includes are as follows:

- a) Semantic features: ANIMATE, HUMAN BEING, FEMALE, RELATIVE, OLD, COUNTABLE, etc.
- b) Syntactic features (lemma): N, GEN (FEM)
- c) Formal features (lexeme):
 - Morphological form: <abuel > <a>

- Phonological form: /a-ˈbue-la /

Out of all of these features only the syntactic features are relevant for the computational mechanism (syntax). However, *production* of a single word involves accessing three types of lexical representation - lexical-semantic, lexical-syntactic (lemma) and lexical-phonological (also referred to as lexeme). This means that *acquisition* of gender for production requires having learned (or transferred from the L1) the three types of lexical representation.

Most lexical theories assume that lexical items compete for selection in the course of language production (Bock and Levelt, 1994; Caramazza, 1997; Garrett, 1976; Levelt et al., 1999; Roelofs, 1992). The chances of a word winning the selection process are in direct ratio to the number of its lexical-semantic, lexical-syntactic and lexical-phonological features activated during the process. For example, gender has been shown by Schriefers (1993) to be one of such features because of an observed gender congruency effect. In that study, monolingual Dutch speakers were tested on a picture-word interference task, where they had to name words while they heard or read a distractor word. The participants had shorter naming latencies if the gender of the distractor word and the target word were congruent (both common or both neuter gender) – an effect called gender congruency effect.

Identification of the components of the lexical representation led to numerous questions about the flow of information between these three components. First of all, are these lexical components stored separately or together? Second, how are they accessed and is there a precedence of one over the other as far as the retrieval process is concerned? That is, once the lexical-semantic features are selected, do they exclusively

activate the lemma level where different lemmas compete for selection until one of them is finally selected or do the lexical-semantic features activate both the lemma level and the lexeme level simultaneously? Also, does the lemma level receive activation exclusively from the lexical-semantic network or both from the lexical-semantic network and formal (lexical-phonological and lexical-morphological) network? According to serial models (Levelt, 1989), in the process of lexical retrieval the lemma level is accessed first where a single lemma is selected, and this lemma is subsequently encoded phonologically. This implies that phonology cannot affect the selection/activation of syntactic features during the production of a lexical item. On the contrary, according to interactive activation models (Dell, 1986), both lexical-syntactic and lexical-phonological information can be accessed simultaneously, which implies that the latter can influence the selection of grammatical features such as gender. Caramazza's (1997) Independent Network model makes similar predictions about the flow of information between the lexical levels. In this model, the three types of lexical information are represented as three separate networks. In the process of lexical retrieval, the lexical-semantic network sends activation to both the lexical-syntactic and the lexical-phonological networks simultaneously. Activation from the lexical-semantic network alone does not suffice to activate syntactic features such as gender; additional activation from the phonological network is required for the feature to be selected.

Considerations about the flow of information during lexical production are relevant here because they led researchers to make hypotheses about lexical *perception* (Levelt et al., 1999), and lexical perception is one of the phenomena through which language learners acquire lexical representations. Levelt et al. (1999) report that the

specific details about lexical perception are not completely clear, but they suggest that when the flow of information is reversed, as it happens during comprehension, the orthographic or acoustic form of the word will first lead to the phonological decoding which only subsequently will activate the syntactic and semantic features of this word. That is, although the lexical retrieval process in the research tasks in this dissertation is triggered by an exposure to the written word¹, not by a necessity to express a communicative message, as it happens when one plans to produce speech, lexical retrieval theories help identify the components of the lexical representation and how these components are accessed and, potentially, how they are acquired. Additionally, the fact that exposure to the written or spoken word obligatorily involves phonological processing illustrates the importance of phonological cues in gender acquisition.

The hypotheses made by the linguistic theories and the lexical retrieval theories discussed above are tested using different populations and methodological designs. The prevalent methodological tool to test linguistic theory hypotheses is the grammaticality judgment task (GJT) obtained from L1 speakers. However, although L1 speakers are an appropriate population to tap into syntactic representations, there is a limit to the insights that data from L1 speakers can provide about the nature of the interaction between the three types of lexical features, precisely because the association between these features is difficult to break, and hence it is difficult to tease apart the different types of features based on L1 speakers' linguistic behavior (Putnam & Sanchez, 2013). This issue has been addressed in psycholinguistics by using Tip-of-the-tongue (TOT) experiments (discussed in detail below) and in neurolinguistics by studying aphasic patients, whose strength of

¹ More information about the research tasks will be provided in Methodology section.

association between the different types of features may be weakened or some of the feature representations may be impaired, as reported below. Remarkably, while initially mainly employed to test hypotheses about syntactic development, second language learners and bilingual speakers represent an ideal population for testing hypotheses about relations between lexical-semantic, lexical-syntactic and lexical-phonological features, because the strength of association between these features is not yet established in L2 learners, or may be more flexible than in L1 speakers and in bilinguals. In the following section, I will report empirical evidence that supports the hypotheses made by the discussed theories.

Gender in Monolingual Studies of Lexical Organization

Lexical retrieval theories have been tested by studies in neuropsychology and cognitive psychology (Badeker et al., 1995; Biran & Friedman, 2012; Caramazza & Miozzo, 1997; Costa et al., 2003; Herbert & Best, 2010; Kulke & Blanken, 2001; Vigliocco et al., 1997). Gender is broadly used in this type of research as a tool for tapping into the organization of the lexicon, because according to the models of lexical retrieval, it is represented as a lexical feature and hypothesized to be stored in the lemma. Through investigating lexical-syntactic gender in healthy individuals and in aphasic patients, the authors of these studies test hypotheses about the levels of lexical representation (semantic, syntactic and phonological) and the relationships between them. The first finding of these studies was that the lexical-syntactic information is stored separately from the lexical-phonological representation.

One type of evidence in favor of separate lexical-phonological representation comes from tip-of-the-tongue (TOT) experiments, where participants are offered a definition of a concept or an object and then are asked to name it. Sometimes participants experience the so-called tip of the tongue phenomenon, where they cannot retrieve the word, but have a strong feeling that they are about to recall it. In many cases they can provide certain characteristics of the word, such as some details about its phonology or orthography, etc. Some TOT studies (Caramazza & Miozzo, 1997, Vigliocco et al., 1997) show that people are able to provide the gender of nouns (language-specific syntactic information) that they are unable to name (phonological form), which suggests that lemma and lexeme in fact represent two distinct types of lexical information.

Caramazza and Miozzo (1997) ran two experiments where they used TOT states to retrieve different types of lexical information from the participants. Fifty-three Italian L1 speakers participated in the Experiment 1 and forty-four Italian L1 speakers participated in Experiment 2. They were asked to provide information about the gender of the word, on the one hand, and its phonological characteristics such as the final and the initial phoneme and the number of syllables. The participants were able to indicate the gender of the noun correctly in 73.8% of cases in Experiment 1 and in 67.8% of cases in Experiment 2. They were also able to name the initial phoneme in 57.5% cases in Experiment 1 and 28.3% in Experiment 2. An important finding is that there was a lack of correlation between the cases where the participants were able to name gender and the cases where they were able to provide information about the phonological form of the

word. This implies that the phonology of the word is accessed independently of the grammatical information.

Vigliocco et al. (1997) conducted a study with 60 undergraduate students also using the TOT phenomenon in order to determine whether the lemma level is accessed separately from the lexeme level. That is, the research question was whether a grammatical form is available independently from the phonological form. The participants were able to name the gender of words correctly in 84% of TOT states, although they were not able to adduce any information about its phonological form (last/first phoneme, the number of phonemes, etc.), thus confirming the hypothesis that lemma level information (gender) is independent from the information about the lexeme.

Another piece of evidence with respect to the question of how the lexicon is organized comes from the studies on aphasic patients. Similarly to the TOT studies, they have supported the idea that the grammatical form and phonological form of the same word constitute different lexical representations. For example, Badecker et al. (1995) conducted an exhaustive series of five experiments with an Italian anomic patient named Dante. In the first experiment, the patient was presented with 344 pictures of entities with masculine (184) and feminine (160) gender. These items included the canonical (but not non-canonical) gender nouns (*-o* endings for masculine and *-a* endings for feminine). He was asked to name the pictures, and if he was unable to do so, he had to indicate the gender of the noun by pointing at two cards with labels “feminine” and “masculine” and to indicate any kind of phonological/orthographic information about the word such as what other words the target item sounds/looks like or what its final and/or initial

letter/phoneme is. The patient was unable to name 111 pictures, but he correctly indicated gender of 106 of those nouns (95% of the cases). At the same time, he was unable to provide any information about orthographic/phonological form of the word.

A sentence-completion task with items that cannot be represented graphically (such as ‘idea’, ‘respect’, etc.) replicated the results of the picture-naming task: Dante correctly indicated gender of the 100 (out of 103) words that he had been unable to name, morphological canonicity did not affect his performance (in fact, the 3 errors he made were all with canonical items), but he could recover no phonological information about the target items whatsoever before he was presented with more than half of their phonemes. Since a forced-choice task was used in order to obtain information about gender (there are only two genders in Italian, and one has to choose one out of the two), but not about phonology of the word (there are 27 phonemes in Italian), the researchers ran another experiment where forced choice task was used across the two conditions. For initial/final phoneme identification, Dante was presented with two phonemes – a distractor and a target phoneme. Therefore, he had to choose between two items similarly to the gender condition. The results were such that while performing with 97.7% accuracy on gender identification, Dante performed below chance on all types of phonological/orthographic information: 53.4 on first letter, 46.6% on last letter and 47.7 on a rhyming word identification. These results confirm the notion of separate access to lemma and lexeme representations.

The finding about the dissociation of grammatical and phonological information in the lexicon has also been supported by a more recent study by Biran and Friedmann (2012), who studied syntactic and lexical retrieval impairments in aphasia in order to tap

into the representation of lexical-syntactic information. A relevant finding was that some of their participants had intact lexical-syntactic knowledge (grammatical gender and predicate-argument structure), but were impaired with respect to phonology. Together with the finding reported above about the dissociation between syntactic and phonological knowledge, this finding suggests that the lexical-syntactic information is not stored in the semantic or phonological lexicon; instead, it is stored separately in a syntactic lexicon. A curious finding of the study is that the impairment of the lexical-syntactic feature such as grammatical gender leads to impaired grammatical gender agreement, but does not necessarily involve impaired semantic gender agreement. More specifically, a patient with lexical-syntactic deficits in the study failed to perform gender agreement on nouns that have lexical-syntactic gender (such as ‘book’ or ‘lamp’), but he performed flawlessly (100% correct) on nouns that have lexical-semantic gender (such as ‘boy’ or ‘girl’). This finding is puzzling, because lexical-semantic gender is a semantic feature, and words that have semantic gender should also have a lexical-syntactic representation at the lemma level so that they can provide values to the syntactic module. If the lemma is not accessible, it should not matter whether the lexical item also has a semantic gender. This hypothesis will be brought up for discussion and tested empirically below.

The studies outlined above confirm the lexical retrieval models’ hypotheses about the organization of the lexicon: the lexicon has been empirically shown to include the lexical-semantic, lexical-syntactic and lexical-phonological representations, which are accessed in the course of word production. Each of these levels contains a bundle of relevant features: the lexical-semantic level stores semantic features such as color, size,

shape, etc., the lexical-syntactic level stores syntactic features such as gender and category, and the lexical-phonological level stores phonological features such as the number of phonemes and syllables. Based on the TOT-studies and studies on aphasic patients, the researchers concluded that gender is one of the lexical-syntactic features that are stored at the lemma level and may be stored separately from information at the lexeme level. This is how gender is represented in the monolingual lexicon. This opens up the question of what happens in the L2 lexicon. How does the lexical representation of gender change in L2 learners depending on proficiency levels? How is the representation of lexical-syntactic features different in learners whose L1 has gender and in those whose L1 does not have gender? I will discuss these questions in the next section.

Studies of Gender in the Bilingual Lexicon

Lexical retrieval is a complex phenomenon, which is still not fully understood in monolinguals, and it is not surprising that it is even less understood in bilinguals. In general, research on the bilingual lexicon has been fairly prolific in the area of L1/L2 lexical-semantic and L1/L2 lexical-phonological interactions, but not in the area of lexical-syntactic interactions. Studies have shown that the *semantic* lexicon is shared across the two languages of a bilingual speaker, and the two languages are activated simultaneously (Kroll & Stewart, 1994; Green, 1998; Costa et al., 2003; Kroll & Tokowitz, 2005). Similarly, data from cognate studies indicate that some of the lexical-*phonological* representations are also shared across the two languages, because lexical retrieval of a word in one language is facilitated if the word in the other language is a

cognate (De Groot & Nas, 1991; Costa et al., 2000; Vigliocco et al., 2002; Salamoura & Williams, 2007; Lemhofer et al., 2008).

Research findings are not as definitive with respect to lexical-*syntactic* features such as gender. The main question is whether the L1 and L2 lexicons share gender nodes. According to the gender-integrated hypothesis (Salamoura & Williams, 2007), both L1 and L2 lemmas are linked to the same gender nodes in bilinguals, which means that activation of a noun with feminine gender in L1 will automatically lead to an activation of the L2 nouns with the same gender. According to the gender autonomous representation hypothesis (Costa et al., 2003), on the other hand, each of the languages has its own entirely separate gender nodes, and this implies that activating a gender node in one of the languages will only cause activation of the nouns in that specific language, and not in the other language. There are studies on the gender congruency effect that support gender-integrated hypothesis (Bordag, 2004; Bordag & Pechmann, 2007; Lemhofer et al., 2008; Paolieri et al., 2010), but there are also studies that support gender autonomous representation hypothesis (Costa et al., 2003).

All of these studies on lexical interactions have been conducted with balanced bilinguals highly proficient in both of their languages. While research on the bilingual lexicon is inconclusive, research on L2 lexicon representation and acquisition is scarce. Lemhofer et al. (2008) have argued that in L2 learners, unlike in balanced bilinguals, it is not about whether the L1 and L2 gender systems are shared/separate, but about *one* system – the L1 system that transfers into the L2. Hence, if the L1 does not have gender, there is no gender system at all, and if it appears, it is very unstable.

A similar idea has been proposed in Jiang's (2000) psycholinguistic model of lexical acquisition. The author adapted Levelt's (1989) earlier hypothesis about the internal structure of the lexical entry, whereby the lexical entries consist of two components – lemma and lexeme. Semantic and syntactic information about the lexical entry is stored in the lemma, while the morphological and phonological information is stored in the lexeme. According to Jiang's (2000) model, L2 lexical acquisition proceeds in three stages. The first stage is referred to as a formal stage, when the L2 phonological form of the word is acquired. The second stage is called L1 lemma mediation stage, because during this stage all of the L1 lemma representations are transferred to the L2 and applied to the L2 lexemes. The third stage is called L2 lemma integration stage, since it involves an integration of all L2 lemma and some L2 lexeme (morphological) representations into the L2 lexical entry. The fundamental argument proposed and tested in Jiang (2000) is that L2 learners fossilize at the second stage of lexical learning, because the L1 lemma content blocks acquisition of the L2 lemma. Thus, the L2 lexical representations include the L2 lexeme that contains mainly phonological forms, only a few L2 morphological specifications, and the L1 lemma. Such a hypothesis may account for the difficulties exhibited by L2 learners with respect to gender agreement acquisition, since performance on surface gender agreement hinges upon mastery of the L2 lemma, and the L2 lemma is acquired late or may even never be acquired for certain lexical items, as it has been suggested by Jiang (2000). In this dissertation, I will test this hypothesis by investigating Russian L1 and English L1 learners of Spanish.

To summarize so far, in all aforementioned theories, lexical items are represented as bundles of semantic, syntactic and phonological features. In Minimalism, lexical items containing semantic and syntactic features enter the syntactic derivation (which can only see the syntactic features but not the semantic ones) and then are sent to the phonological component where the abstract syntactic structures obtain phonological forms. In theories of lexical retrieval, there are two hypotheses in this connection: either serial flow of information from lexical-semantic information to lexical-syntactic and finally to lexical-phonological level or parallel activation of lexical information when lexical-semantic information activates both lexical-syntactic and lexical-phonological levels. An essential difference between the linguistic theories and the lexical retrieval theories presented is that the latter do not make any predictions about the syntactic module per se, nor do they take any explicit stance at whether syntax is modular and innate. Again, however, all of the above mentioned theories assume that lexical items are comprised of three types of features: semantic, syntactic and phonological features. The relevance of these considerations for my dissertation and my theoretical assumptions with respect to the minimalist view and the lexical theories will be discussed below in the Lexical/syntactic hypothesis chapter.

Having provided a basic introduction to theories of lexical retrieval, I will present in the next section previous works on L1 and L2 acquisition of gender.

The Study of Gender in Language Acquisition

In this section, I will review existing research on L1 and L2 gender acquisition, based on which I will emphasize two main ideas. First, the results of both L1 and L2 research are inconclusive and still leave room for interpretation. Second, despite the classic juxtaposition between L1 and L2 gender acquisition, adults and children acquire gender similarly in several respects. For example, they exhibit incipient agreement soon after they are exposed to language. Their gender errors often reflect a search for a perfect rule-based system and, in fact, are not syntactic gender agreement errors, but rather lexical gender assignment errors. Based on this literature review, I will propose a hypothesis for L2 gender acquisition, where I will principally distinguish between the lexical and the syntactic components of gender.

There are different theories that hypothesize about how language is acquired. Generative theories propose that at least some part of language is modular and innate. According to these theories, all L1 learners have access to the innate language acquisition capacity, but for L2 learners this capacity may or may not be available. Among generative SLA theories are Full Transfer Full Access Hypothesis or FTFA (Schwartz & Sprouse, 1996), Minimal Trees Hypothesis (Vainikka & Scholten, 1996), Failed Functional Features Hypothesis or FFFH (Hawkins & Chan, 1997) and Local Impairment Hypothesis (Eubank, 1994).

Non-generative theories deny or remain agnostic with respect to the innate linguistic capacity. Among those non-generative theories that do not make strong claims about the innate linguistic capacity is the Sociocultural Theory (Lantolf & Thorne, 2007), according to which both L1 and L2 acquisition happens through peer interaction and cultural immersion. Another such theory is the Skill Acquisition theory (Dekeyser, 2007;

VanPatten & Benati, 2010), which proposes that adults learn language as well as any other skill – through explicit learning and practice, which then may lead to implicit knowledge.

Among the non-generative theories that explicitly deny any innate linguistic capacity are usage-based theories of language acquisition, which reject any preexisting knowledge of language, and claim that (both L1 and L2) language is acquired solely through understanding intentions of other people, picking up patterns in their interactions and *imitating* others (Tomasello, 2003). Similarly, Connectionist theories (Ellis, 1998) argue strongly against innateness and modularity of language and posit that language acquisition happens through emergence – a process in which smaller units interact and combine to form larger units and where the interaction between the units leads to spreading activation. Simply put, the amount of input is the core predictor of language acquisition. As Lidz and Gagliardi (2015) state, “On the input-driven view, what is learned is the recapitulation of the inputs to the learner. The acquired representations are a compressed memory representation of the regularities found in the input” (p.334).

Innateness has been one of the key topics in the language acquisition field, but positing an innate linguistic capacity does not imply negating the importance of input for certain aspects of language, as was mentioned in the Introduction. In fact, input is considered one of the key factors for language acquisition in generative theories as well (Lidz & Gagliardi, 2015), and it is only the function of input that is conceived of as different. More specifically, in the knowledge-driven tradition, “the learner searches the input for cues to help *choose* an abstract representation” (Lidz & Gagliardi, 2015, p.334).

In this dissertation I assume a generative theoretical framework because I test hypotheses that concern the acquisition of syntax and the lexicon and not, for example, the effect of peer interaction on the acquisition of syntax and the lexicon, or whether explicit knowledge of grammar can turn into the implicit knowledge of it. I do, however, indirectly test the effect of the amount of input on gender agreement. As the findings will show, a generative framework can account for the syntactic aspects of gender as separate from the lexical ones although both are needed for comprehension and acquisition purposes.

L1 gender acquisition. L1 gender acquisition has been classically contrasted with L2 gender acquisition with the emphasis on the difficulties and the variability of the latter. However, a comprehensive literature review reveals that child gender acquisition is also not a uniform process, and that different researchers have obtained inconclusive results with respect to at least some of its aspects. In this subsection I will focus on L1 acquisition research that has explored the age by which children acquire gender, the types of errors they make as they acquire gender and the reasons why they make these specific gender errors.

Researchers have not been unanimous with respect to the question of the age of L1 acquisition of gender. L1 learners have been shown to acquire the Spanish determiner/noun agreement as early as by the age of three by some authors (Hernandez Piña, 1984; Soler, 1984; Clark, 1986; Snyder et al., 2001). Ample research on L1 gender in other Romance languages shows the same (Clark, 1986; Karmiloff-Smith, 1979; Heinen & Kadow, 1990; Pizzuto & Caselli, 1992; Bottari, Cipriani & Chilosi, 1993; inter

alia). On the other hand, some authors such as Mariscal (2008) argue that these studies mostly explore agreement between the noun and the most frequent determiner – the definite article and maintain that although three year olds may have acquired determiner/noun agreement, they still may not have mastered agreement between the noun and other agreement elements such as demonstratives, possessives, attributive and predicative adjectives, and clitics. Mariscal (2008) argues that children master agreement between nouns and definite articles before anything else precisely because these determiners are so omnipresent in the input. In this dissertation I test this idea with L2 populations.

Similarly, while some studies show that L1 gender acquisition is almost error-free from the very early ages - from the onset of 2 word stages (Aguirre, 1995; Schnell de Acedo, 1994), other studies show that children go through a period when they make different types of agreement errors (Hernandez Piña, 1984; Lopez-Ornat, 2003; Clark, 1986; Mazeika, 1973; Mariscal, 2008). For example, Mariscal (2008) argues that gender acquisition process is typically characterized by a large amount of inter-individual variability - children of the same age producing different types of errors or being at different developmental stages as well as intra-individual variability - the same child can show target-like and non-target like agreement on the same structure.

Two lines of reasoning have been proposed to account for the L1 gender errors. Some authors suggested that the gender agreement errors that children make are in fact performance errors and are not caused by their lack of agreement features (Lleo, 1997,

2001; Aguirre, 1995)². More recently, two L1 online comprehension studies have been designed to test this hypothesis (Johnson, 2005; Lew-Williams & Fernald, 2007). Johnson (2005) used an online methodology (Split-screen Preferential Looking Paradigm) to find out whether 28-month old Dutch-speaking babies show sensitivity to determiners (definite articles) and use gender cues in their interpretation of Dutch nouns. The participants were able to use gender cues on the common gender article *de*, but not on the neuter gender article *het*. According to the authors, this finding may be accounted for by a significantly higher frequency of the article *de* in the input. Moreover, the article *de* is more regular than the article *het*, because *het* serves several grammatical functions. Such small children may need more time to map the form of the word to all of its functions. Lew-Williams and Fernald (2007) conducted an eye-tracking study with 2 to 3 year old Spanish children with similar research questions. These children were able to identify nouns faster when nouns were preceded by a gender-congruent determiner (Lew-Williams & Fernald, 2007).

Other authors emphasize that although children do not master agreement completely until 5-7 years of age, they exhibit at least incipient gender agreement at very early ages (Ceytlin, 2009; Clark, 1986; Karmiloff-Smith, 1979; Muller, 1994; Perez-Pereira, 1991). They argue that the majority of L1 gender errors are not arbitrary, but rather are caused by the *lexical and morphological* complexity of specific gender systems (Ceytlin, 2009; Clark, 1986; Karmiloff-Smith, 1979; Muller, 1994; Perez-Pereira, 1991). These systems exhibit numerous exceptions and unreliability of forms, which causes

² The same argument has been proposed with respect to the L2 learners' errors and has been theoretically enunciated in the Missing Surface Inflection Hypothesis (Haznedar and Schwartz, 1997; Prévost and White, 1999), which will be discussed below.

children to spend years making sense of the system and deducing its rules. This line of reasoning also finds support in research on L1 Russian gender acquisition which shows that Russian children exhibit target-like noun/adjective agreement on nouns ending in *-a* and masculine nouns ending in an unpalatalized consonant as early as by the age of 2;6 (Ceytlin, 2009). That is, as soon as they have established a straightforward rule to assign gender, they perform agreement. However, they only master agreement on end-stressed neuter nouns between the ages of 3-4, and agreement on stem-stressed neuter nouns and on feminine palatalized nouns is fully acquired as late as by the age of 6-7 (Ceytlin, 2005, 2009). These are the types of nouns that are the least transparent with respect to gender assignment. It seems, therefore, that children exhibit target-like determiner/noun agreement by the age of three, but still have a long way to go before they master agreement on all nouns of the language. Thus, most of the errors reflect children's search for a dependable and stable system, while in reality the system is unreliable due to the lexical/morphological complexities.

In general, the errors that children make often stem from their reliance on morphological gender markers on nouns (Clark, 1986; Karmiloff-Smith, 1979; Perez-Pereira, 1991; Muller, 1994), and as reported in the Introduction chapter, these markers are not reliable in Spanish. More specifically, Clark, (1986), Karmiloff-Smith (1979), Perez-Pereira (1991), and Muller (1994) show that younger children (roughly before age of 6) mostly use gender markers on nouns to determine gender of nonce words whereas older children (up to age 9) rely on gender on determiners or choose masculine as a default for the words whose gender they do not know. For example, when nonce words with typically feminine gender forms (*-elle*, *-aise*) are presented to French children in a

DP with the masculine indefinite determiner “un”, children before 5 years of age tend to use the feminine definite determiner ‘*la*’ to match it to the morphological form of the noun, and vice versa for typically masculine gender forms (Clark, 1986). However, they change their gender strategies and start relying on the indefinite article with which the nonce words appear at older ages (after 5), irrespective of the form of the nouns.

Similarly, Spanish-speaking children (and L2 Spanish adults, as it will be discussed later) treat the feminine *-a* and masculine *-o* pattern as a general rule to determine gender of nouns (Perez-Pereira, 1991). Again, this changes after 5 years of age, when children start relying on gender on determiners rather than on gender markers on nouns.

Thus, L1 learners acquire basic gender as early as by the age of three both in Spanish and Russian, and the errors they make after this age are caused by a lexical-morphological complexity of the specific gender systems rather than by an inability to perform gender agreement.

L2 gender acquisition. The fundamental question within the SLA field is the explanation of the differences between the native grammars and the L2 interlanguage. In this section, I will review some SLA theories and the factors that they propose to account for these differences. Among such theories are, on the one hand, those that propose a critical period for the acquisition of *syntactic* features, and, on the other hand, the theories according to which the only predictor of language acquisition is the quality and quantity of input received by L1 and L2 learners. In this dissertation, I will challenge both of the theories, and will attempt to show that syntactic features are acquirable in L2, although a lexical-morphological restructuring of the L2 grammar is necessary.

Connectionists and function-based theory supporters maintain that adults are in no way compromised compared to children, and that the preponderant factor in the L2 acquisition difficulties is the decreased amount and quality of input compared to the L1 (Cameron-Faulkner, Lieven & Tomasello 2003; Tomasello, 2003; Tomasello & Brooks, 1999). They report that while L1 learners receive numerous hours of high-quality, child-directed input every day (approximately 5,000 to 7,000 utterances daily), L2 learners are exposed to incommensurably less input, which often does not address the learners' communicative goals. In fact, Mariscal (2008) claims that even for children input is the only predictor of the rate of gender acquisition. She argues that children exhibit target-like agreement between the determiner and the noun, but not on the noun and other elements of the DP such as adjectives precisely because D/N sequences are far more frequent in the input than (D)/N/A sequences. This should be true for adults as well, if input is the mere or at least the key factor for gender acquisition. I will test this idea by comparing gender agreement between D and N, on the one hand, and between N and A, on the other hand.

On the other hand, some generative SLA theories attribute the L2 difficulties to a critical period after which access to Universal Grammar becomes unavailable. More specifically, Minimal Trees Hypothesis (Vainikka & Scholten, 1996), Failed Functional Features Hypothesis or FFFH (Hawkins & Chan, 1997), Impaired Representation Hypothesis or IRH (e.g. Eubank, 1993/94; Meisel, 1997) and more recently the Representational Deficit Hypothesis (Hawkins, 2009) posit a syntactic impairment at the level of functional categories, features or feature strength. A number of L2 studies on

gender production have provided support for these theories, specifically for FFFH (Hawkins, 1998; Hawkins & Franceschina, 2004; Franceschina, 2001). For example, Hawkins (1998) showed that L1 English L2 French learners make persistent but inconsistent errors – they use either masculine or feminine as a default gender and optionally produce the masculine or the feminine determiner with the same noun. Also, in Franceschina (2001), the three L1 English L2 Spanish participants produced gender concord less accurately than the three L1 Italian L2 Spanish participants. An example of a gender concord error made by the English L1 speakers is “una sistema” (a.F system.M).

At least two types of empirical evidence can challenge the idea of the critical period for grammar acquisition. First, although the critical period theories propose that L2 learners fail or perform poorly on grammatical operations, accuracy rates across the impressive body of existing research on gender range as high as 80-90%, regardless of whether the L1 has or does not have gender, and even in the studies arguing strongly for an L2 representational impairment the accuracy rates range between 75-80% (Dussias et al., 2013). For example, Bruhn de Garavito & White (2000) and White et al. (2004) have shown that L2 learners whose L1 does not have gender (English) behave similarly to L2 learners whose L1 has gender (French), thus challenging the generative critical period theories and lending support to non-syntactic impairment models such as Full Transfer Full Access Hypothesis (1996), according to which UG is available in any age and thus all L2 functional categories and features are acquirable in L2, provided that sufficient input is present to activate grammar reorganization. Additionally, some of the studies that claim maturational constraints on the acquisition of syntax used production data, which might not be completely appropriate methodologically for tapping into syntactic

representations. This idea is reflected in the Missing Surface Inflection Hypothesis (Haznedar & Schwartz, 1997; Prévost & White, 1999), according to which L2 learners do not lack syntactic features but rather occasionally lag behind in producing (not representing) their morphological exponents in real time. Similarly, according to Lardiere (2000), the errors that L2 learners make in production are explained in terms of problems at the syntax-morphology interface referred to as the “Mapping Problem.”

Second, while the “critical period” studies posit that the difficulties in L2 acquisition stem from a critical period on acquisition of grammar, there is a growing amount of evidence that shows that gender errors in fact are caused by a lack of lexical knowledge, not knowledge of grammar. More specifically, several psycholinguistic and neurolinguistic studies indicate that gender agreement errors that L2 learners make are gender assignment errors (Alarcón, 2011; Grüter et al., 2012; Hopp, 2012; Sabourin & Stowe, 2008), caused by having assigned an incorrect value to the noun and subsequently sharing this value among the elements of the DP. In the following paragraphs I will report on such studies and expand on the idea of the lexical deficit vs. syntactic impairment.

Grüter et al. (2012) ran an experiment to find out whether L2 learners whose L1 lacks gender (L1 English) showed difficulties on a production task and, if they did, whether these difficulties came from syntactic (gender agreement) or lexical (gender assignment) domains. For that, they included three types of nouns in the study design: 1) transparent gender nouns (*a*-feminine, *o*-masculine), non-transparent gender nouns (consonant and *e*-ending) and irregular gender nouns (*o*-feminine and *a*-masculine). Participants were offered two images of the same object with slight alterations (color, size, etc.) as they heard sentences such as “Which of the _____ do you prefer?” For

example, they saw an image of a red butterfly and an image of a yellow butterfly and heard a question “¿Cuál mariposa prefieres?” (Which butterfly do you prefer?). They were expected to reply with D/A sequences such as “La roja” (The.F red-F) or “La amarilla” (The.F yellow-F), where the gender of the noun should be reflected on the determiner and the adjective. Results showed that gender assignment errors were 10 times as frequent as gender agreement errors. Gender assignment errors were considered those where both the determiner and the adjective had an incorrect (but matching) gender (*El mano pequeño* – The.M hand small-M instead of the target-like *La mano pequeña* – The.F hand small-F) while gender agreement errors were considered those where the gender on the determiner did not match that on the adjective or vice versa (*El mano pequeña* – The.M hand small-F or *La mano pequeño* – The.F hand small-M).

Hopp (2012) conducted an eye-tracking study to find out whether advanced to near-native L1 English L2 learners of German can use gender as a cue for noun recognition during language processing. Twenty L2 German learners and 20 native speakers of German participated in the study. First of all, their lexical gender assignment was tested. For that, the participants looked at displays of four tangible objects (e.g., a dress, a car, a button and a card) on a computer screen and named the objects as well as their color. The results of this production task showed that some of the L2 participants showed target-like overall assignment, while others did not.

After the production task, the participants had to look at the same displays again while they heard a question such as “Where is the.M/F/N yellow Noun?” Gender cues were available on fifteen of the displays, where two of the four objects were of the same color/quality/size and the only difference between them was their gender (e.g., a display

with pictures of a red car, blue card, yellow dress and a yellow button was accompanied with the question “Where is the yellow dress?”). The participants could only rely on the determiner to predict the noun, because German nouns are non-transparent with respect to gender.

The results showed that L2 learners are capable of using gender as a predictive cue, and that this ability depends crucially on whether the noun has been assigned the correct gender value. Essentially, the facilitating power of gender diminished significantly for the group of L2 learners who had shown variable or inconsistent gender assignment.

Importantly, even speakers of a gendered language may seem to have difficulties with gender agreement in L2, if the gender systems are different in the two languages, as it is the case in French and Dutch or Italian and Dutch (Dewaele and Véronique, 2001; Sabourin & Stowe, 2008). The authors of these studies have shown that the difficulties that L2 learners experience stem from gender reassignment – the process of assigning L2 gender values to nouns that already have an L1 value in the lexicon. For example, if masculine value has already been assigned to a noun in the L1, it will be difficult to reassign this value to the same noun if the noun has a different value (feminine/neuter, common, etc.) in the L2. That is, in these studies the L1 appears to affect the L2, but not syntactically in terms of inability to acquire syntactic features after puberty, but rather lexically in terms of reassigning new gender values to the nouns.

Sabourin and Stowe (2008) compared cortical reactivity in two experimental groups both of which came from a gendered language (14 L1 German L2 Dutch and 8 L1 Romance L2 Dutch speakers) and a control group of 23 L1 Dutch native speakers. All

three languages have the gender feature, but gender assignment is different for Dutch and Romance. For example, the German and the Dutch gender systems are very similar in both gender assignment and agreement, because German neuter gender nouns match Dutch neuter nouns and German masculine and feminine gender become common gender in Dutch. On the other hand, Romance languages have entirely different lexical gender: both masculine and feminine gender nouns can have both neuter and common gender values in Dutch. Therefore, the authors hypothesized that L1 German learners will transfer processing strategies and gender assignment from their L1 into Dutch and so will show target-like P600 effect whereas L1 Romance will transfer their lexical gender assignment into Dutch which will lead to non-target like processing. Results of two experiments confirmed the hypotheses: German L1 learners of Dutch showed target-like sensitivity to gender agreement violations, while Romance speakers did not.

Based on the results, the authors conclude that it may be not transfer of grammatical representation of an L1 structure into L2 that ensures target-like processing, but rather transfer of processing routines. Similar neural processing cannot occur in structures such as Dutch gender agreement for Romance L1s because it depends on lexical gender assignment, which is fundamental for subsequent syntactic operation gender agreement.

The studies by Alarcón (2011), Grüter et al. (2012), Hopp (2012) and Sabourin and Stowe (2008) are innovative in their distinction between gender agreement and gender assignment, but, as it was noted in the Introduction chapter, they do not aim to provide a specific theoretical foundation for the representation and acquisition of complex lexical-syntactic features. In this dissertation, based on the insights from the

linguistic theories and the theories of lexical retrieval, I have built and tested a theoretical foundation as to what gender is, precisely what its lexical representations are, what the difference between these lexical representations and abstract syntactic gender features is, how these lexical representations are acquired, and how and why they affect the performance on the surface gender agreement.

In summary, L2 acquisition is considered considerably less successful than L1 acquisition. For some researchers this observation is a motivation to posit critical periods for grammar acquisition. For others it is a big piece of evidence to argue that input is the only factor for language acquisition. However, the literature review of L1 and L2 gender acquisition has shown that L1 gender is not as flawless and rapid as it has been considered, and that L2 gender acquisition is not as compromised, against the widespread opinion. Moreover, L1 and L2 gender acquisition seem to exhibit similar developmental patterns: both L1 learners and L2 learners exhibit gender agreement fairly soon after the initial exposure to language and although they still may make errors years after, these errors seem to be triggered not by an incapability to acquire grammar, but rather by problems with lexical learning. Thus, the critical period idea and the input as the only factor idea may be incorrect, and the lexical-morphological complexity of the specific linguistic systems may be an additional (if not a central) factor that leads to problems with acquisition of such complex features as gender.

Lexical/syntactic hypothesis. Having reviewed the large body of existing research on gender acquisition, I propose and attempt to test a set of hypotheses that may account for some of the findings of this research. First of all, I propose a binary

distinction between two separate, but closely interrelated phenomena – abstract syntactic operation gender agreement and language-specific gender assignment. I suggest that the abstract syntactic operation gender agreement is based on core universal linguistic abilities (more details in the next paragraph) and that it reveals itself in the L2 shortly after the initial exposure to DPs containing gender. Gender assignment, on the other hand, is a lengthy incremental process of learning language-specific gender values such as M, F, N, etc. for specific nouns, a process that depends on the amount of exposure and practice and that may take years and even decades to complete. The presence of the abstract syntactic operation gender agreement combined with the correct gender assignment ensures target-like *surface* gender agreement. In other words, surface gender agreement cannot be target-like if one of these two components is missing.

The first component - the abstract syntactic gender agreement - is a parameterized (more details below) syntactic operation present in some languages but not in others (Carroll, 1989). It is a type of Agree operation, which is also referred to as Concord (Baker, 2013). This operation is made available by the UG in the following way. Agree is a universal syntactic operation (Hauser et al., 2002; Fitch et al., 2005). As it was discussed above, it is a feature-sharing operation in the sense that it distributes the same features and their values among different syntactic elements that are bound to a certain syntactic domain (Chomsky, 2000; Pesetsky & Torrego, 2004). There is a universal repertoire of syntactic features such as tense, number and gender that may be subject to Agree and that different languages can draw from. Thus, both Agree and the repertoire of syntactic features are universal and made available by the UG. However, which of these features is instantiated in a given language is not universal but rather parameterized in the

sense that some languages have a [+] parameter for this operation, and others have [-] parameter for it. For example, Spanish and Russian have gender agreement, whereas English does not.

Then, if gender is a parameterized syntactic feature subject to Agree operation, the question is – is it available in the L2 if it was not activated in the L1? Different linguistic SLA theories have made different predictions with respect to the acquisition of syntactic features in the L2. Some of these theories have been discussed earlier in the chapter. For example, Failed Functional Features Hypothesis or FFFH (Hawkins & Chan, 1997) and Representational Deficit Hypothesis (Hawkins, 2009) predict that all lexical and functional categories and syntactic features are transferred from L1 into L2; however, there is no access to UG after the critical period and, consequently, no resetting of parameters is possible in L2 acquisition. On the hand, Full Transfer Full Access Hypothesis or FTFA (Schwartz & Sprouse, 1996) predicts that in case the L1 and the L2 have parametric differences, the L2 parameters are in fact acquired through access to UG. In other words, the innate language faculty is always available, and consequently, features that were not activated in the L1 can be activated in the L2. The following predictions can be made with respect to gender based on these two theories. If FFFH were correct, and if abstract syntactic gender features were not available in the L2 if they were not activated in the L1, such a syntactic impairment would inevitably lead to permanent non-target-like performance on surface gender agreement. Conversely, if FTFA were correct, and abstract syntactic features could be acquired in the L2 even if they were not activated in the L1, it could lead to a target-like surface gender agreement *provided that the language-specific gender assignment is acquired*. This leads us to the

same conclusion that was made in the Introduction chapter of this dissertation – the abstract syntactic gender features are indispensable for the surface gender agreement to be target-like; however, they do not suffice, because language-specific gender assignment (the lexical-morphological characteristics of the gender system) has to be acquired in addition to the abstract syntactic features. I will test the predictions of these two theories by studying surface gender agreement in L2 learners of Spanish whose L1 lacks gender (English). If these learners are capable of behaving target-like on surface gender agreement, it would indicate that the abstract syntactic operation gender agreement is in fact available in the L2. If they are not, however, this would not mean that they are necessarily impaired syntactically, because the non target-like behavior can be caused, among other factors, by an incorrect gender assignment.

Thus, the second component necessary for surface gender agreement to be target-like is gender assignment. Gender assignment is the process (and the result) of incorporating language-specific gender values (M, F, N, etc.) of specific *nouns* into the lexicon. As it was suggested in the introduction chapter, these values can be incrementally acquired based on three types of cues. The first type of cue is the gender morphemes on nouns. More reliable (transparent) gender morphemes should facilitate gender assignment, and less reliable (opaque) gender morphemes may exacerbate the process. Again, there are no completely reliable gender morphemes on nouns or reliable gender deduction rules based on these morphemes; instead, there are *patterns*. Some gender systems such as Italian, Russian or Hebrew exhibit more organized or fewer and more inclusive patterns, whereas other systems such as German could follow a more vague pattern of gender deduction (Bordag et al., 2006). Consequently, the rate of the

gender assignment process should depend on such characteristics of specific gender systems. For example, a system that exhibits more exceptions to a general pattern should be acquired slower, and this idea is brought up in Bordag et al. (2006, p. 1091), "...not only that opacity of gender marking affects gender learning of such nouns but also that the higher the proportion of these nouns is, the slower the learning of the whole system becomes." Nevertheless, establishing the patterns is also based on the morphological markers, because these patterns are built upon an association of gender morphemes such as *-a* and *-o* with feminine and masculine gender values.

The second type of cue in addition to gender morphemes on nouns is the gender on determiners. This type of cue is the only unambiguously reliable cue to establish gender values of nouns. The problem for gender assignment comes from the tendency that language learners have to rely on gender morphemes on nouns instead of on determiners. This is the case in L1 Spanish and L1 French acquisition, where children initially rely on gender morphemes on nouns to deduce gender, and only later switch to gender on determiners (Clark, 1986; Karmiloff-Smith, 1979; Muller, 1994; Perez-Pereira, 1991). One could expect that adults, like children, initially prefer morphemes on nouns to gender on determiners. If this idea is applicable to adults as well, L2 learners will need to have learned a large number of nouns and they would have had to be exposed to many instances of violation of the morphological feminine *-a* vs. masculine *-o* pattern (*la mano* 'the hand', *la torre* 'the tower', *el problema* 'the problem', etc.) in order to shift to the strategy of using gender on determiners as a cue to establishing gender of the noun. The extent with which language learners prefer to rely on gender morphemes on nouns or on determiners should depend on availability and transparency of gender morphemes on

nouns. For example, if nouns of a specific language are morphologically non-transparent with respect to gender (like Dutch), the learners of this language should rely on determiners, because they do not have morphemes to rely on. This was the case in Hopp's (2012) study, where the L2 learners of Dutch were able to use determiners as cues for noun interpretation. On the other hand, if nouns are systematically marked for gender, L2 learners should rely on these gender morphemes. In fact, it may be the case that morphologically transparent systems such as Spanish and Russian (unlike German or Dutch) are more difficult to acquire, because morphological markers and patterns based on them do not provide such an unequivocal gender cues as determiners.

Lexical-semantic gender features might serve as the third type of cue to acquire gender assignment, because lexical-semantic features transfer from L1 to L2. For example, it may be easier for an L2 learner to assign feminine gender to a noun that refers to a female entity such as "lady" than to a noun that has no conspicuous semantic characteristics that would help classify the word as of feminine or of masculine gender. In this dissertation, I will test the idea that the (more) reliable gender morphemes that follow the general *-a* for feminine/ *-o* for masculine pattern (defined as 'the inner core' in Harris, 1991) are acquired more easily and thus lead to a more target-like surface gender agreement, while the rest of gender morphemes (defined as 'the outer core' and the 'residue' in Harris, 1991) are acquired later and thus lead to non-target-like surface gender agreement. Such a finding would also indicate that L2 learners, like children, initially rely on gender morphemes on nouns before they switch to the determiners to deduce gender values. I will also test the idea that lexical-semantic gender features may serve as an additional cue to acquire gender values.

The crucial idea in this dissertation is that both gender morphemes on nouns and gender on determiners can only be attended to in the input provided that the abstract syntactic operation gender agreement is present in the learners' grammars. If it were not, the learners would not be able to represent the DP as a constituent (a syntactic domain) that shares features and feature values, and thus would not single out the language-specific morphemes on nouns, adjectives and determiners that map onto the abstract syntactic structures. For example, it is precisely because a speaker processes (32) as a syntactic constituent whose elements must share features and feature values that s/he attends to the morphological markers on all of the elements of the DP and subsequently associates this marker with a certain (in this case, feminine) gender value.

(32) La casa

DEF.F house-F

This initial association will lead to categorizing nouns into two classes (for Spanish), or, speaking in terms of WEAVER++ (Levelt et al., 1999), to connecting the nouns to two incipient gender nodes in the lexicon. This association between the gender morpheme (the lexical-phonological representation) and the gender value (the lexical-syntactic representation) of the noun will be weak, which will reveal itself in a non-target-like surface gender agreement, and it has to be constantly reinforced through accessing and producing the noun. That is, language learners have to be exposed to, comprehend, and produce nouns in order for this association between a specific morphological form, specific gender value, and finally a specific semantic meaning to become automatic – that is, for the correct *lexical* representation of the noun to be created. The more often they

process for comprehension and production, the stronger the association becomes between the gender morpheme, the gender value and the semantic meaning of the noun.

Importantly, once the lexical-syntactic representation on a specific noun has been formed, surface gender agreement should be performed on all elements of the DP that includes that noun, and not only on those that are more frequent in the input such as D/N sequences, as it is claimed by Mariscal (2008). This is so because if abstract syntactic gender features are present, they should be present on all DP-elements. This idea will be tested in this dissertation by comparing agreement between the definite article and a noun – a type of agreement that is considered by Mariscal (2008) as the most frequent in the input - with agreement between a noun and an adjective, a type of agreement that is considered as less frequent.

The idea that while gender agreement always remains present in L1 grammars, gender *assignment* may be lost due to a lack of exposure and practice can be supported by the data from heritage speakers, who have been shown to undergo attrition with respect to gender (Montrul et al., 2008). If at some point early in life the constant reinforcement of the association between the lexeme and the lemma becomes scarce or absent, the association may weaken or even get lost. Thus, in the case of heritage speakers what is attrited is not abstract syntactic operation gender agreement, but gender assignment – the language-specific gender values associated with specific forms and meanings stored in the lexicon³.

Furthermore, since gender assignment is language specific, it should proceed differently depending on whether the L1 has grammatical gender. If the L1 does not have grammatical gender, there are three basic factors that might hinder gender assignment.

³ This idea is not going to be tested in this dissertation.

The first factor is the potential unavailability of abstract syntactic features in the L2, as claimed by the FFFH (Hawkins & Chan, 1997) and by the Representational Deficit Hypothesis (Hawkins, 2009). Phonological constraints are the second factor that could affect gender assignment. The third factor has to do with the absence of lexical-syntactic gender features in the lemmas of nouns that would be connected to gender nodes in the lexicon.

If the L1 has gender, on the other hand, gender assignment in the L2 should be affected by different factors. Some researchers have assumed that if the L1 has grammatical gender, L2 learners should not have difficulties or at least should have fewer difficulties acquiring L2 gender agreement (Carroll, 1989; Hawkins & Chan, 1997, among others). For instance, Carroll (1989) maintains that L2 learners who have gender in L1, "...will have relatively few problems learning French gender even when the morphosyntactic, phonological, or semantic properties of gender in the L1 differ radically..." I will challenge this claim by comparing L2 learners of Spanish whose L1 has gender (Russian) with the L2 learners whose L1 lacks gender (English).

I will argue that the presence of the abstract syntactic operation gender agreement in the L1 alone may not be able to guarantee seamless gender acquisition, because along with the transfer of *syntactic* operations, the *lexical*-syntactic representations of the L1 will transfer to the L2 and may thus significantly affect the acquisition of the L2 gender agreement.

The lemma content of nouns and the gender nodes present in the lexicon of this particular language will transfer from L1 to the L2 and will affect *surface* gender agreement. In case L1 and L2 gender values match, it should provide an advantage, but in

case they do not, it should lead to non-target-like performance on surface gender agreement in L2. This is so because L2 learners, like L1 learners, acquire lexical-phonological features (referred to as *lexeme* in this dissertation) first, followed by the lexical-syntactic features (referred to as *lemma* here) (Ellis, 2003). That is, L1 Russian learners of Spanish will first learn that their L1 Russian word *dom* ‘house’ is pronounced [*ka-sa*] in Spanish, and only later will (probably) acquire that this Spanish word has a feminine gender value in the L2, unlike in the L1. More specifically, when they hear a word in the L2 such as [*ka-sa*], they will translate this word into their L1 - [*dom*] *without accessing its meaning*, and will automatically access the L1 lexical-syntactic and lexical-semantic representations for the lexeme [dom] (Jiang, 2000).

As learners become more proficient through frequent processing for comprehension and production in their L2, they will learn to directly link the L2 phonological form /*ka-sa*/ to its meaning (lexical-semantic representation) (Kroll & Stewart, 1994), but they may or may not acquire the L2 lexical-syntactic representation of this noun, because the L1 masculine value could block the integration of the feminine L2 gender value into the L2 lemma (Jiang, 2000). That is, although gender is present in L1 Russian and hence the abstract syntactic gender features should be present in the L2, the L2 Spanish values may not be acquired or may take a long time before being acquired. In addition to the lexical incongruency, phonological constraints may affect processing of gender morphemes in the L2 both for L2 learners whose L1 has gender and for those who do not have it. Therefore, the fact that the L1 is similar to the L2 may not be enough for the surface gender agreement to be target-like.

In summary, gender is represented syntactically - as the value distributing abstract syntactic operation gender agreement that is bound to a certain syntactic domain - DP, morphologically - as gender morphemes on the elements of the DP, and lexically - as gender nodes to which lemmas of all nouns of a given language are connected. Hence, mastery of gender requires being able to build a syntactic structure that has a gender feature on all DP-elements, to rapidly retrieve a gender value from the lemma level of the specific noun (provided that the value has been incorporated), to share the value with the DP-elements and finally to retrieve the morphological markers corresponding to the gender features.

While generating abstract structures is in the domain of syntax (abstract syntactic operation gender agreement) and may or may not be available in the L2 (FTFA vs. FFFH), the lexical values and morphological forms are a matter of the long-term memory, and hence should require sufficient exposure and practice. Since lexical values are learnt through exposure to gender morphemes, L2 learners will show differential surface gender agreement depending on the (relative) reliability of morphological cues. As it was discussed in the Introduction chapter, this may account for the finding of some researchers that gender is not acquirable in L2 – it is not the syntactic features that are not acquirable, but rather the gender assignment has not been acquired. Additionally, lexical-semantic gender features may facilitate the acquisition of gender values, which would lead to a more target-like surface gender agreement on nouns that have lexical-semantic features such as ‘girl’ vs. ‘boy’ compared to nouns that do not have such features, such as ‘house’. Moreover, contra to what some researchers have assumed, even if the L2

learners transfer abstract syntactic gender features from L1 to L2, they may still perform non-target like on surface gender agreement due to an L1/L2 gender incongruency.

Thus, the research questions explored, corresponding research hypotheses tested and the predictions made with respect to the hypotheses in this dissertation are:

Research question 1: Are abstract syntactic gender features acquirable in the L2 if they were not activated in the L1?

Hypothesis 1: Contra FFFH (Hawkins & Chan, 1997) and Representational Deficit Hypothesis (Hawkins, 2009), and following FTFA (Schwartz & Sprouse, 1996), abstract syntactic gender features should be available in the L2 regardless of age of acquisition of these features, even if the L1 lacks them.

PREDICTION 1: High proficiency English L1 learners of Spanish, whose L1 lacks gender, will be capable of performing target-like on the research tasks in this dissertation. Such a finding will indicate that both the abstract syntactic features and the language-specific lexical-morphological features are acquirable in the L2. It will also indirectly indicate that the phonological constraints can be overcome in the L2.

Research question 2: Does the lack of exposure to abstract *syntactic* gender features in the L1 compromise gender acquisition in the L2, as it is maintained in Hawkins and Chan (1997), Hawkins and Francheschina (2004), Francheschina (2001), and Hawkins (2009)?

Hypothesis 2: Lacking abstract *syntactic* gender features in the L1 does not compromise surface gender agreement in the L2. Because access to abstract *syntactic* gender features is hypothesized to be available, as predicted by FTFA Hypothesis (Schwartz and Sprouse, 1996), L2 language-specific lexical and morphological features *can* be acquired with

proficiency, which allows for target-like surface gender agreement even in the case that gender is absent in the L1.

PREDICTION 2: High proficiency English L1 learners of Spanish, whose L1 lacks gender, will be capable of performing target-like on the research tasks in this dissertation.

Research question 3: Does the exposure to the abstract *syntactic* gender features in L1 facilitate gender acquisition in L2?

Hypothesis 3: Having acquired abstract *syntactic* gender features in the L1 does not facilitate surface gender agreement in L2. Preexisting lexical knowledge can affect surface gender agreement, because L2 gender assignment may fossilize at the L1 mediation stage without passing to the L2 lemma integration stage, as suggested by Jiang (2000). Additionally, even if the L2 lemmas are integrated into the lexicon, L1 Russian L2 Spanish learners may experience interference from the L1 lemmas, which can also affect surface gender agreement.

PREDICTION 3: High proficiency L1 Russian learners of Spanish may not exhibit target-like behavior on L1/L2 incongruent nouns.

Research question 4: Is surface gender agreement affected by the morphological characteristics of the L2 gender system (cue unreliability)?

Hypothesis 4: Surface gender agreement is affected by the morphological characteristics of the L2 gender system (cue unreliability).

PREDICTION 4: L1 English speakers will make the majority of gender errors on nouns that do not follow the general *-a* for feminine / *-o* for masculine pattern (the outer core and residue, in Harris' (1991) terms), and will have significantly fewer errors on nouns that follow this pattern (inner core of prototypes, in Harris' terms).

Research question 5: Is surface gender agreement affected by the presence of preexisting L1 lexical knowledge (L1/L2 congruency), even if the abstract syntactic gender features are available?

Hypothesis 5: Preexisting L1 lexical knowledge affects surface gender agreement in L2. L2 learners transfer this knowledge and initially assign a new lexeme (e.g., *casa* ‘house’) to the L2 lexical items while keeping all of the L1 lexical-syntactic features (*dom* ‘house’ – masculine). With proficiency, this effect may diminish for the majority of incongruent nouns, but there still may be interference from L1 lemma, which might affect surface gender agreement.

PREDICTION 5: Proficiency mediates the state of lemma acquisition in the L2. More specifically, low-proficiency L1 Russian L2 Spanish learners will transfer L1 lemmas into Spanish and will perform significantly more accurately on nouns with Russian/Spanish congruent gender, while high proficiency L1 Russian L2 Spanish learners may have integrated the L2 lemmas, and thus may behave target-like on L1/L2 incongruent nouns. Alternatively, interference from L1 lemmas may lead L1 Russian learners to flawed surface gender agreement.

Research question 6: Do lexical-semantic gender values (as in *chica* ‘girl-F’ vs. *chico* ‘boy-M’) facilitate the acquisition of lexical-syntactic gender values of these nouns compared to the acquisition of lexical-syntactic gender values of the nouns that lack lexical-semantic gender values (as in *casa* ‘house-F’ vs. *queso* ‘cheese’)?

Hypothesis 6: Lexical-semantic gender values may or may not facilitate gender assignment.

PREDICTION 6: If L2 learners use lexical-semantic gender as an additional cue to learn gender assignment, they will perform more accurately on nouns with lexical-semantic gender.

Research question 7: Is the most frequent type of gender agreement - agreement between a definite article and a noun - more target-like than the less frequent type of agreement – agreement between a noun followed by an adjective?

Hypothesis 7: If the abstract syntactic gender features are present on all DP-elements and provided that the correct lexical gender value has been incorporated into the lexicon, L2 learners should not be affected by the type of agreement (D/N vs. N/A). That is, L2 learners should not perform more accurately on D/N agreement than on N/A simply because D/N sequences are more frequent in the input.

PREDICTION 7: L1 Russian and L1 English L2 learners of Spanish will behave similarly on determiner/noun gender agreement and on noun/adjective gender agreement.

The research questions will be addressed using a methodological design described in the next chapter. Detailed information about two research studies conducted to test the hypotheses will be provided followed by the results obtained in these studies.

CHAPTER 3: Methodology

As mentioned in the Introduction chapter and in Chapter 2, research questions of this dissertation explore the factors that may affect L2 acquisition of surface gender agreement. The first and the main research question in this dissertation concerns the effect of Age of Exposure on the acquisition of gender agreement and gender assignment. I use the term ‘Age of Exposure’ to refer to the age of first exposure to the *abstract* gender features, not to the language specific lexical, phonological and morphological manifestation of gender. For example, while L1 Russian L2 Spanish learners do not have knowledge of gender values for specific nouns or knowledge of concrete morphemes for F or M in Spanish, they do have the abstract gender representation, because they had activated it in their L1 before the critical period. L1 English L2 Spanish learners, on the other hand, were exposed to gender only in their L2, after the critical period had ended. Alternatively, instead of the age of exposure, proficiency in the L2 could affect surface gender agreement; hence, another research question that the dissertation explores is the role of proficiency for ultimate attainment in gender acquisition.

Furthermore, this dissertation investigates how certain linguistic factors such as the way gender is manifested lexically and morphologically in a given language (Spanish) could affect surface gender agreement. The next research question in this dissertation concerns the role of input for L2 acquisition of gender agreement. Last but not least, the dissertation investigates whether lexical-semantic gender features facilitate the acquisition of the syntactic operation gender agreement in the L2.

Thus, the hypotheses in this dissertation test effects of two types of factors in the acquisition of gender: a) acquisitional factors (effects of Age of Exposure and proficiency in the L2) and b) linguistic factors (effects of input, congruency, gender type and cue type). The acquisitional factors represent between subjects variables, because they are examined between separate groups (early vs. late Age of Exposure groups, low vs. high proficiency groups), while the linguistic factors represent within subjects variables, because they are tested in all of the groups across the board.

Two separate experiments were conducted to test the hypotheses. Experiment 1 was designed to test the two acquisitional hypotheses (effects of Age of Exposure and proficiency) and two of the four linguistic hypotheses (effects of input and congruency), while Experiment 2 was designed to test one of the acquisitional hypotheses (proficiency) and two of the linguistic hypotheses (gender type and cue type). Effects of input and congruency were not investigated in Experiment 2 because clear results had been already obtained for both variables in Experiment 1, and effects of gender type and cue type were not investigated in Experiment 1 because these variables were added to the design after Experiment 1 had been conducted and a clear lexical effect on gender agreement had been established for L1 Russian, so it was relevant to find out whether there was a similar lexical effect for a non-gendered L1.

Accuracy and reaction times were the two dependent variables in both of the experiments. The analysis of reaction times was performed in addition to the analysis of accuracy in order to establish whether L2 learners' processing of gender was affected by the linguistic variables (type of agreement, congruency, gender type, and cue type). Furthermore, they were analyzed to see whether low proficiency learners' processing of

gender is slower because of a less automatic lexical retrieval. Last but not least, RTs were measured because slower RTs on ungrammatical items compared to grammatical ones are considered indicative of actual processing of gender agreement violations.

Experiment 1

The two acquisitional (between subjects) variables in this experiment are Age of Exposure (AoE) of abstract gender features, where English L1 are the late Age of Exposure group and Russian L1 are the early Age of Exposure group, and Proficiency in the L2 (Spanish), which was manipulated across the two L1 backgrounds – L1 English L2 Spanish low proficiency, L1 Russian L2 Spanish low proficiency, L1 English L2 Spanish high proficiency, and L1 Russian L2 Spanish high proficiency. The two linguistic (within subjects) variables are Type of Agreement (the more frequent D/N agreement type vs. the less frequent N/A agreement type, as reported in Mariscal (2008), and Congruency (Spanish/Russian congruent vs. incongruent items).

Participants. Four groups of L2 Spanish learners (19 low proficiency L1 Russian, 19 high proficiency L1 Russian, 19 low proficiency L1 English, and 16 high proficiency L1 English) and a control group of 13 native speakers of Spanish were involved in the Experiment 1. Age and proficiency scores are reported in Table 2.

Table 2. *Age and proficiency scores*

	n= of participants	Age (years)	Proficiency score (DELE)
Russian low proficiency	19	18-41	18-33
Russian high proficiency	19	25-40	40-45

English low proficiency	19	18-24	19-32
English high proficiency	16	20-35	40-49
Control group	13	26-42	n/a

The groups were matched on measures of education, age, and socioeconomic status. The four experimental groups were also comparable in terms of L2 learning background: they all had started learning Spanish after puberty (at least 10 years old) and had learned Spanish in a classroom. Most of the high proficiency L2 learners of Spanish (both L1 Russian and L1 English) had lived abroad. Most of them had not learned other gendered foreign languages other than Spanish. Some high proficiency L1 English speakers had studied other Romance languages such as French, Italian and Portuguese, but their proficiency was low or at least (in case of one participant) lower than their proficiency in Spanish. It would be ideal if no participants spoke any gendered language other than Spanish, but it was particularly difficult to find high proficiency L2 learners of Spanish who had not learnt other Romance languages. Since fourteen out of 48 test items of the Experiment 1 had incongruent gender values in Spanish and Italian, and 14 items had incongruent gender in Spanish and French, I controlled for an effect of L2/L3 incongruency on the behavior of these participants. No significant differences were found between the performance of these participants on L2/L3 congruent and incongruent items; that is, their performance on Spanish gender was not affected by their L3. They performed equally well on both L2/L3 congruent and incongruent items.

Proficiency levels of the two experimental groups were determined by the results of a written portion (grammar and reading) of a version of the DELE proficiency exam. There are 50 total points possible on the test. The following proficiency groups were

differentiated for the purposes of the study: low-proficiency group (scores from 18 to 33) and high-proficiency group (scores from 40 to 49). This division may seem arbitrary, but a One-way ANOVA revealed significant differences between the two low-proficiency groups and the two high proficiency groups ($F(3,69)=130.297, p<.5$). Post hoc tests (Bonferroni) did not reveal significant differences 1) between the DELE scores of the L1 Russian low proficiency group and the English L1 low proficiency group or 2) between the DELE scores of the L1 Russian high proficiency group and the L1 English low proficiency group. However, it showed that the two low proficiency groups were significantly different from the two high proficiency groups. Considering that all other variables were equal, these results confirm that the proficiency division was appropriate and that the experimental groups could be used to test the hypotheses.

In order to test the effect of Age of Exposure on gender agreement, the two L1 Russian groups (19 low proficiency and 19 high proficiency participants) were grouped together to form an early Age of Exposure group (early AoE), because they had been exposed to and had activated the abstract gender features in their L1. The two L1 English groups (19 low proficiency and 16 high proficiency participants), on the other hand, were grouped together to form a late Age of Exposure group (late AoE), because they were only exposed to abstract gender features in their L2. The early AoE group consisted of 38 participants and the late AoE group consisted of 35 participants.

The control group was comprised of 13 native speakers of Spanish from Spain and from different countries in South and Central America. They all spoke both Spanish and English on a daily basis; however, most of them (11 out of 13) self-reported to be

dominant in Spanish. The other two participants were more or less balanced in both of their languages.

Research Tasks. The two research tasks employed in the Experiment 1 are: a timed self-paced reading grammaticality judgment task (GJT) and a timed picture-matching task (PMT). Both tasks were created using a PsychoPy software, were executed on a PC laptop computer and required reading, but not speaking, writing or listening.

Grammaticality Judgment task. The GJT in this experiment was self-paced. Self-paced reading tasks are *online* tasks that are often used to tap into processing of different linguistic structures (Marinis, 2010). Self-paced reading tasks can be cumulative, non-cumulative and center non-cumulative. For the GJTs in this experiment, the center non-cumulative presentation was used. This is a type of a self-paced reading task presentation where one word appears in the center of the screen and then disappears followed by another word appearing in the center of the screen. This type of presentation is considered to provide more accurate information about sentence processing than a cumulative presentation, because it does not allow the participant to regress to the preceding words and to build expectations about the length of the sentence (Marinis, 2010). Longer RTs are indicative of difficulties in processing, and thus provide insights into how certain linguistic structures are processed.

After reading detailed instructions on the computer screen and taking a practice test, the participants in this experiment saw sentences in a center non-cumulative moving window format: they saw one word of the sentence at a time and every time they pressed

the relevant key, the word disappeared and was followed by the next word of the sentence. After every word the participants saw, they had to indicate whether it looked acceptable in the context of the sentence. There were 48 target sentences and 48 distracter sentences. All target sentences consisted of 9-12 words and contained determiner-noun-adjective sequences such as in (33).

(33) Puedes usar el bolígrafo rojo que está en la mesa.

You-can use the.M pen-M red-M that is on the table.

“You can use the red pen that is on the table.”

The sequences were of four types:

1. 12 grammatical sequences as in (34)

(34) la casa blanca
the.F house-F white-F
“the white house”

2. 12 determiner/noun error sequences as in (35)

(35) el casa blanca
the.M house-F white-F
“the white house”

3. 12 noun/adjective error sequences as in (36)

(36) la casa blanco
the.F house-F white-M
“the white house”

4. 12 determiner/noun/adjective error sequences as in (37)

- (37) *el casa blanco*
 the.M house-F white-M
 “the white house”

The critical regions were after the noun (Noun position), where participants had to judge whether the noun that they saw at the moment “agreed” (had the same gender value) with the determiner that had preceded it (*la casa*), and after the adjective (Adjective position) where they had to judge whether the adjective “agreed” with the noun that had preceded the adjective (*casa blanca*). The noun position showed agreement between the determiner and the noun. For example, if participants saw *el* (the.M) followed by *casa-F* (house), they were expected to judge *casa-F* as unacceptable.

The adjective position showed agreement between the noun and the adjective. For example, if participants saw *casa-F* followed by *blanca-F* (white-F), they were expected to judge *blanca* as acceptable.

To test Hypothesis 3 about the effect of input on gender agreement, I compared groups’ scores and reaction times on D/N sequences such as (38) and (39) against their scores and reaction times on N/A sequences such as (40) and (41).

D/N sequences:

- (38) *la casa ~~blanca~~*
 (39) *el casa ~~blanca~~*

N/A sequences:

- (40) ~~La~~ *casa blanca*
 (41) ~~La~~ *casa blanco*

In order to avoid potential spillover effects, no N/A sequences with a preceding determiner/noun gender mismatch (*el casa blanca*, *el casa blanco*) were used for the analysis of N/A sequences.

To test Hypothesis 5 about the L1 lemma effect on L2 gender assignment, I compared the groups' scores and reaction times on congruent and incongruent test items. Half of the test items had the same gender values in Spanish and Russian (L1/L2 congruent gender items) and the remaining half had different gender in Spanish and Russian (L1/L2 incongruent gender items) (Table 3 below). While Russian L1 groups were expected to perform more accurately and react faster on congruent items, English L1 groups and the control groups were expected to be immune to such an effect. Additionally, proficiency was expected to modulate the congruency effect.

Table 3. *Examples of congruent and incongruent items*

#	Congruent items	Incongruent items
1.	English: 'cup' Spanish: 'taza-F' Russian: 'chashka-F'	English: 'orange' Spanish: 'naranja-F' Russian: 'apelsin-M'
2.	English: 'spoon' Spanish: 'cuchara-F' Russian: 'lojka-F'	English: 'car' Spanish: 'coche.M' Russian: 'mashina-F'
3.	English: 'cucumber' Spanish: 'pepino-M' Russian: 'ogurets-M'	English: 'book' Spanish: 'libro-M' Russian: 'kniga-F'
4.	English: 'suit' Spanish: 'traje.M' Russian: 'kostyum-M'	English: 'magazine' Spanish: 'revista-F' Russian: 'jurnal-M'

Accuracy was calculated by assigning a score of one (1) for each correct answer (accepting a correct item and rejecting an incorrect item) and by assigning a zero (0) for

an incorrect response. Scores from one participant from the low proficiency Russian L1 group were removed because they represented an outlier, which is why only data from 19 low proficiency L1 Russian participants were analyzed instead of the initial twenty who had participated in the study. RTs were analyzed for outliers per participant and per condition (grammatical vs. ungrammatical condition, congruent vs. incongruent condition), as it is suggested in Blom and Unsworth (2010). All values two standard deviations above or below the mean of each participant per each condition were excluded from further analysis.

In this dissertation I attempted to prevent the participants from using their metalinguistic (explicit) knowledge of gender. The question of the specific distinctions between implicit and explicit knowledge and the specific methodological tools to measure them has been investigated in a number of studies (Bialystok, 1979; Bowles, 2011; DeKeyser, 2003, 2009; Ellis, 2005; Ellis et al., 2009; Godfroid et al., 2015; Rebuschat, 2013; Suzuki & DeKeyser, 2015). Some of the criteria to distinguish between implicit and explicit knowledge are the *degree of awareness* (the issue of whether the learners are aware of the grammatical rule or are driven by a “feeling”), the *time available* (the idea that L2 learners will be less likely to retrieve their explicit knowledge under time constraints) and the *focus of attention* (the idea that L2 learners cannot attend to both form and meaning) (Ellis et al., 2009).

While some researchers consider that the time available is the crucial criterion to tap into implicit knowledge (Bowles, 2011; R. Ellis, 2005; Erlam, 2006; Spada et al., 2015; Zhang, 2015), others suggest that explicit knowledge can be used even under time pressure, and that the degree of awareness is thus a more reliable criterion in this regard

(DeKeyser, 2003, 2009; Suzuki & DeKeyser, 2015). Moreover, according to Suzuki and DeKeyser (2015), no behavioral methodologies can distinguish between implicit knowledge and a highly automatized metalinguistic knowledge. Nevertheless, while the question of tapping into implicit knowledge still remains open, a timed GJT is considered by many researchers as an appropriate tool to measure implicit knowledge (Bialystok, 1979; Bowles, 2011; R. Ellis, 2005; Ellis et al., 2009; Erlam, 2006; Godfroid et al., 2015; Rebuschat, 2013; Spada et al., 2015; Zhang, 2015).

I attempted to create conditions that would force the participants to use their implicit knowledge. First, although the GJT was self-paced, the participants in the experiments of this dissertation were requested to give their answers intuitively, as soon as possible and without thinking. They were told that their reaction times were recorded. Second, in Experiment 2 the sentences were followed by a comprehension check, so the participants had to focus on meaning in order to give the correct answer. Moreover, in the GJTs of both of the experiments the participants had to direct all of their attention to reading the words and keeping them in their working memory, because the sentences appeared on the screen one word at a time, and they had to hold the part of the sentence they have read in working memory. It is highly likely that the cumulative moving window task format imposed an increased processing load and potentially prevented L2 learners from being able to retrieve their memorized knowledge of L2 rules.

In addition, because of the non-cumulative presentation of the sentences the participants were not able to regress (move their gaze from right to left to see the previous words in the sentence) – and regressing is one of the ways in which L2 learners can reflect on their metalinguistic knowledge when they are taking a GJT (Godfroid et

al., 2015). With respect to the awareness criterion, the participants were naïve to the purpose of the experiment, and even after completion of it many of them remained unaware of the linguistic structure in question (gender agreement). Thus, given that there is no consensus among researchers about the criteria to distinguish between implicit and explicit knowledge, and with the understanding that this is a behavioral study, every effort has been made to force the participants in this dissertation to refrain from using their metalinguistic knowledge.

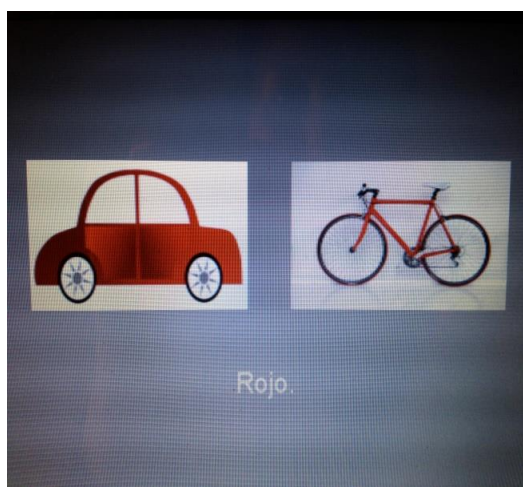
Picture-matching task. The picture-matching task was added as an alternative tool to test the effect of age of exposure, the effect of proficiency, and the effect of congruency on L2 surface gender agreement.

During the task, the participants were presented with 40 sets of two pictures of concrete tangible objects, one of which was of masculine and the other of feminine gender. One of the pictures was placed on the left side and the other one on the right side of the screen. Below the pictures participants saw a phrase which either consisted of a verb and a clitic as in *Dámelo* (Give me it.M) or of a single adjective such as *Nueva* (new.F). There were 40 phrases in total, 20 of which were of congruent gender and 20 of incongruent gender (See Table 3 above). The phrase on the screen semantically could refer to both pictures, but grammatically it only matched with one of them because only one of the objects was of the relevant gender.

Figure 4. *A slide with verb+clitic phrase on the picture-matching task*



Figure 5. A slide with adjective on the picture-matching task



For one of the test items (See Fig. 4), participants were expected to press the left arrow because the clitic *la* (her) in the phrase *cómela* (eat her) below the two objects signals feminine gender and correlates with gender of the noun *fresa/frutilla*. On the left of the screen. This is an example of a congruent item: strawberry is feminine both in Spanish and in Russian. On another test item (see Fig. 5) the adjective *rojo-M* (red) can only refer to the noun *coche/carro-M* (car) on the left of the screen although semantically it could refer to both (both the car and the bicycle are red). This is an example of an incongruent item: car is masculine in Spanish (*coche/carro*), but feminine in Russian (*mashina*). The

position of appropriate objects was counterbalanced so that the participants could not build expectations as to whether the right object would be on the left or on the right of the screen.

In order to find out whether the Age of Exposure affected surface gender agreement, the late and early Age of Exposure groups were compared between each other and against the control group. Similarly, to find out whether proficiency had an effect, the low and high proficiency groups' scores and RTs were compared. In order to determine whether the linguistic variable Congruency affected surface gender agreement, the groups' scores and RTs on congruent and incongruent items were compared. In addition, the PMT was expected to cause the participants to follow a lexical-semantic to lexical-syntactic to lexical-phonological route, as it happens during the word production, rather than the reversed lexical-phonological to lexical-syntactic to lexical-semantic route followed in the GJT. The same procedure for calculating Accuracy and RTs and establishing outliers as in the GJT was used in the PMT.

Experiment 2

Experiment 2 was designed to test the effect of proficiency, the effect of lexical-semantic vs. lexical-syntactic gender type, and the effect of cue reliability. That is, it was designed to reveal whether the lexical-morphological complexity of gender in a specific gender system (Spanish) affects gender assignment and, thus, surface gender agreement. The experimental task included a) lexical-semantic vs. lexical-syntactic gender items and b) reliable vs. unreliable cues items. Cues were considered reliable if the masculine *el*

determiner was followed by a noun with the transparent morphological *-o* ending and the feminine *la* determiner was followed by the transparent *-a* ending (*the inner core of prototypes*, according to Harris (1991)), whereas unreliable cues were those where the masculine *el* determiner was followed by a noun with the contradicting morphological *-a* ending, the feminine *la* determiner was followed by the contradicting *-o* ending (*the residue*, according to Harris' terminology), or when either of the determiners was followed by a noun with an uninformative morphological ending such as *-e*, *-d*, *-n*, and *-z* (*the outer core*, according to Harris).

Therefore, the study design of Experiment 2 was 2 (Gender Type: lexical-semantic vs. lexical-syntactic) x 2 (Cue Type: reliable vs. unreliable) x 3 (Group: low proficiency L2 x high proficiency L2 x native-speakers). As in Experiment 1, Accuracy and Reaction Times were the dependent variables.

Participants. Two groups of L1 English L2 Spanish learners (17 low proficiency L1 English and 16 high proficiency L1 English) and a control group of 17 native speakers of Spanish were tested in the Experiment 2. The same high proficiency group as in the Experiment 1 was involved in the Experiment 2. Age and proficiency scores are reported in Table 4.

Table 4. *Age and proficiency scores*

	n= of participants	Age (years)	Proficiency score (DELE)
English low proficiency	17	18-29	18-30
English high proficiency	16	20-35	40-49
Control group	17	26-42	n/a

The groups were matched on measures of education, age, and socioeconomic status. The two experimental groups were also comparable in terms of L2 learning background: they all had started learning Spanish after puberty (at least 9 years old) and had learned Spanish in a classroom. I tested the high proficiency speakers who had studied French, Italian and Portuguese, because ten items out of 40 test items had incongruent gender values in Spanish and Italian, and 11 items had incongruent gender in Spanish and French. I found no significant differences between the performance of these participants on L2/L3 congruent and incongruent items; that is, their performance on Spanish gender was not affected by their L3 lexical knowledge. They performed equally well on both L2/L3 congruent and incongruent items. Proficiency levels of the two experimental groups were determined by the results of a written portion (grammar and reading) of a version of the DELE proficiency exam. There are 50 total points possible on the test. The following proficiency groups were differentiated for the purposes of the study: low-proficiency group (scores from 18 to 30) and high-proficiency group (scores from 40 to 49). This division may seem arbitrary, but an independent samples t-test revealed significant differences between the low-proficiency group ($M=26.06$, $sd=3.944$) and high-proficiency group ($M=43.50$, $sd=3.141$), $t(31) = -13.993$, $p < .05$. Considering that all other variables were equal, these results confirm that the two experimental groups had different proficiency levels and could be used to test the research hypotheses.

The control group was comprised of 17 native speakers of Spanish from Spain and from different countries in South and Central America. They all spoke both Spanish and English on a daily basis; however, most of them (14 out of 17) self-reported to be dominant in Spanish. The other three participants were more or less balanced in both of

their languages. The fact that the native speakers come from different L1 backgrounds (from different countries in Latin America and Spain and with variable amount of exposure to English) may reflect the variety in the input received by the L2 Spanish L1 English learners.

Research tasks. The research task employed in the study was a self-paced reading grammaticality judgment task (GJT). It was similar to the one used in the Experiment 1. The task was created and executed on a laptop computer using PsychoPy software. It required reading, but not speaking, writing or listening. After reading detailed instructions on the computer screen and taking a practice test, the participants saw sentences in a center non-cumulative moving window format: they saw one word of the sentence at a time and every time they pressed the relevant key, the word disappeared and was followed by the next word of the sentence. After every word the participants saw, they had to indicate whether it looked acceptable in the context of the sentence. Half of the experimental sentences included DPs with gender agreement violations. Again, as in the Experiment 1, the task was presented in a center non-cumulative moving window format to minimize the use of metalinguistic knowledge by the participants. The RTs were recorded and analyzed to reveal any potential differences in the processing of grammatical vs. ungrammatical items, lexical vs. semantic gender items, and reliable vs. unreliable cue items.

There were a few differences between the GJT in Experiment 1 and Experiment 2. First, in Experiment 2, the test items contained DPs only with determiner/noun sequences, but no adjectives were included. Second, the GJT in the Experiment 2

included 60 target sentences and 60 distracter sentences, unlike 48 target and 48 distracter sentences in the Experiment 1. Third, unlike Spanish/Russian congruent/incongruent conditions in the Experiment 1, the 60 experimental sentences in the Experiment 2 involved the following three conditions: 1) lexical-semantic gender conditions with 20 nouns such as *abuela* ‘grandmother-F’, as in (42) 2) lexical-syntactic gender condition with 20 nouns with regular gender endings (-a vs. -o) such as *cerveza* ‘beer-F’, as in (43) and 3) lexical-syntactic gender with 20 nouns with unreliable cues such as *la mano* ‘the.F hand.M’, as in (44). Last but not least, the task sentences were followed by a comprehension check – a translation of the original sentence in Spanish.

(42) Me preocupa que el abuela se haya quedado sola en casa.

I am worried that the.M grandmother-F stayed alone at home.

‘I am worried that the grandmother stayed at home alone.’

Comprehension check: I am worried that the grandmother stayed at home alone.

(43) No me gusta el cerveza porque tiene un sabor amargo.

I don’t like the.M beer-F because it has bitter taste.

Comprehension check: I really like wine, because it tastes sweet and fruity.

(44) Mark dice que sabe leer la mano y predecir el futuro.

Mark says he can read the.F hand.M and predict the future.

Comprehension check: Mark says he has met his girlfriend's parents the other day.

The example in (42) is followed by a comprehension check translation, which in this case corresponds to the Spanish original, while the examples in (43) and (44) do not. The comprehension check translations matched the original sentences in Spanish only half of the time both for experimental and distracter sentences. The example in (42) contains a

DP with a gender agreement violation (*el abuela* – the.M grandmother.F). The critical region is after the noun, where participants had to judge whether the noun that they saw at the moment (*abuela* – grandmother.F) “agreed” (had a gender value that matched the value of the determiner) with the determiner that had preceded it (*el* – the.M). For example, in this case the participants were expected to judge *abuela.F* as unacceptable, because they had seen the determiner *el* (the.M), whose gender value did not match that of the noun *abuela*.

In order to find out whether cue reliability enhances gender assignment, I compared sentences with reliable cues such as in (43) with unreliable cues such as in (44). In order to find out whether lexical-semantic gender serves as a cue for gender assignment, I compared sentences with lexical-semantic gender nouns such as in (42) with sentences with lexical-syntactic gender nouns such as in (43). In addition, I tested whether the effects of the linguistic variables were modulated by proficiency.

To sum up, the variables studied in Experiment 1 are linguistic variables Type of Agreement (D/A vs. N/A) and Congruency (congruent vs. incongruent) and acquisitional variables Age of Exposure (early vs. late) and Proficiency (low vs. high). The variables studied in Experiment 2 are linguistic variables Gender Type (lexical-semantic vs. lexical-syntactic) and Cue Type (reliable vs. unreliable) and an acquisitional variable Proficiency. Statistical tests conducted for these experiments and results obtained from the tests are reported in the next chapter.

CHAPTER 4: Results

The organization of the chapter is as follows. First, results for Accuracy and Reaction times for Experiment 1 will be provided followed by results for Accuracy and Reaction times for Experiment 2. For Experiment 1, within the subsection of results for Accuracy and the subsection of results for Reaction times, results for age of exposure and results for proficiency will be presented separately. Finally, within the subsection of results for age of exposure and the subsection of results for proficiency, results from the grammaticality judgment task (GJT) and the picture-matching task (PMT) will also be reported separately. In other words, I will first report how accurate the AoE groups were on GJT, and then how accurate they were on PMT. After that, I will report how accurate the Proficiency groups were on GJT, then how accurate they were on PMT. Then, I will report results in the same order for Reaction Times – GJT results and PMT results for AoE groups followed by these types of results for Proficiency. For Experiment 2, I will first report results for Accuracy and then results for RTs.

Experiment 1

Accuracy. As it has been mentioned above, the results in this section of the chapter are reported for two different variables – Age of Exposure and Proficiency for each of the research tasks (GJT and PMT). I will first present results for Age of Exposure from the GJT and PMT and then present results for Proficiency from the GJT and PMT.

Age of Exposure.

Grammaticality Judgment Task.

Descriptive data. This task consisted of 48 research items, so the total possible score that the participants could have is 48. Table 5 presents the distribution of scores (group means) for overall condition, D/N agreement condition, N/A agreement condition, congruent condition, and incongruent condition across the early AoE, late AoE groups and the control group. As expected, the control group was more accurate than the two experimental groups in all conditions. Furthermore, the early AoE group seems to be slightly more accurate than the late AoE on all conditions but the incongruent one. As hypothesized, the early AoE group appears to be more accurate on congruent than on incongruent condition. Finally, none of the groups seem to be affected by type of agreement (D/N vs. N/A).

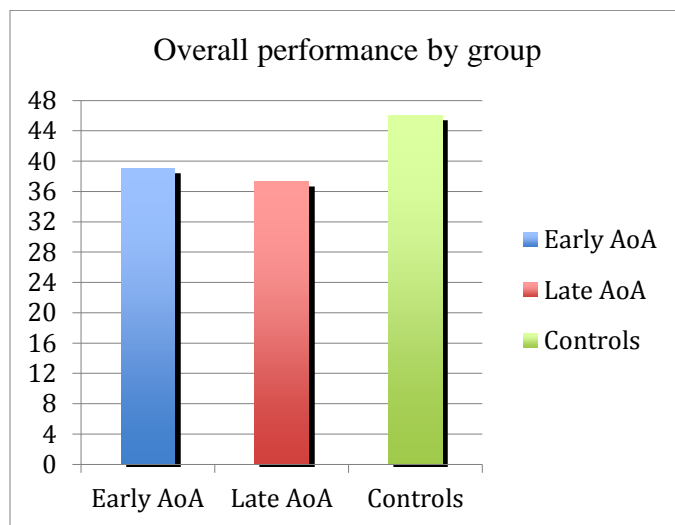
Table 5. *Descriptive data. Group mean scores and total possible scores for the task (indicated next to the means) for specific conditions across the two AoE groups and the control group*

	Mean group scores /total possible score		
	Early AoE	Late AoE	Controls
Overall items	39.08 / 48	37.29 / 48	46.00 / 48
D/N agreement	20.24 / 24	19.17 / 24	23.31 / 24
N/A agreement	19.47 / 24	18.91 / 24	23.31 / 24
Congruent items	20.39 / 24	18.06 / 24	22.85 / 24
Incongruent items	18.63 / 24	18.20 / 24	23.15 / 24

In order to determine whether the between-subjects and within-subjects differences were significant, a set of statistical analyses was carried out.

Statistical analyses. First, I compared the overall scores from the early AoE group (L1 Russian group), late AoE group (L1 English group), and the control group. Group means are presented in Fig. 6.

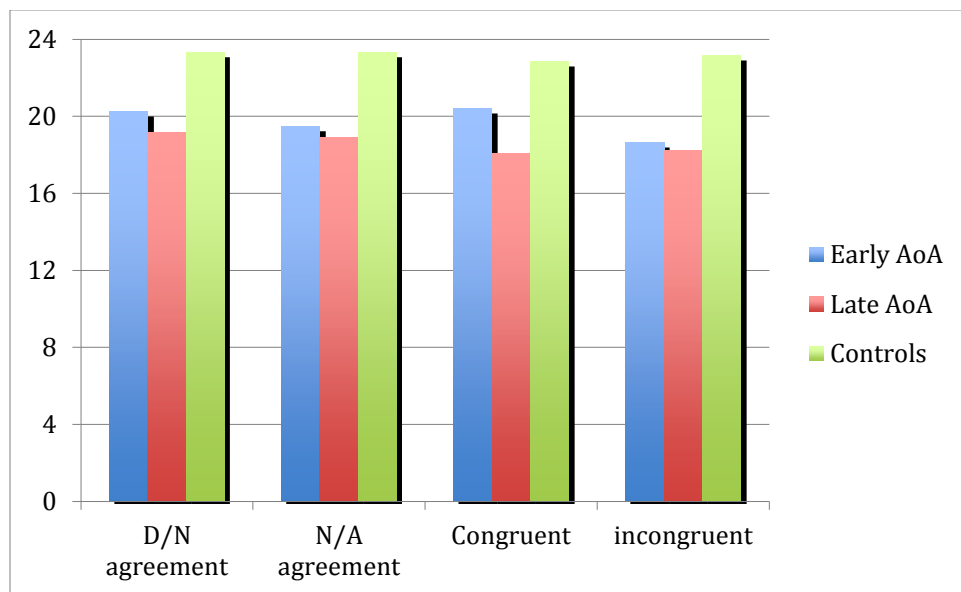
Figure 6. *Group mean scores for the overall condition (total possible score 48) across the two AoE groups and the control group*



Tests of normality (Kolmogorov-Smirnov test) showed that distributions in all three groups were normal. A one way ANOVA revealed a significant main effect, $F(2,83)=6.376$, $p=.003$ (Levene's $p=.000$). Post Hoc tests (Dunnett T3) showed that the differences between the two experimental groups were not statistically significant, $p=.721$. On the other hand, the control group was significantly more accurate than both experimental groups, $p=.000$. Thus, early Age of Exposure did not lead to an enhanced gender agreement on the task.

The next step was to reveal any potential interactions between the acquisitional variable Age of Exposure and the linguistic variables Type of Agreement and Congruency. Half of the 48 research items were congruent and half of them were incongruent, so the total possible score that the participants could get on congruent items is 24 and the total possible score for incongruent items is also 24. The same applies for D/N/ agreement type and N/A agreement type – each of the conditions has 24 items, and hence the highest possible score for each of the conditions is 24. Group means across the four conditions are presented in Fig. 7.

Figure 7. *Group mean scores for D/N condition (total possible score 24), N/A condition (total possible score 24), congruent condition (total possible score 24), and incongruent condition (total possible score 24) across the two AoE groups and the control group*



While in Fig. 7 it may seem that D/N agreement was more accurate for Early AoE group, this difference was not statistically significant, because a repeated measures ANOVA with a 2 (Type of Agreement) x 2 (Congruency) x 3 (Age of Exposure) factorial design showed no significant effect for Type of Agreement, $F(1,83)=1.447$, $p=.232$, as

well as no interaction between Type of Agreement and Age of Exposure, $p=.761$. However, the statistical test showed a significant main effect for Congruency, $F(1,83)=3.995$, $p=.049$, and a significant interaction between Congruency and Age of Exposure, $p=.002$. A series of paired samples t-tests showed that the only group affected by Congruency (performed more accurately on congruent than on incongruent items) was the early AoE group (L1 Russian). The ANOVA also revealed a significant main effect for the between-subjects variable Age of Exposure, $F(2,83)=9.546$, $p=.000$, and Post Hoc tests (Dunnett T3) showed that the differences were between the two experimental groups, on the one hand, and the control group, on the other hand, but no statistically reliable differences were found between the early and late AoE groups.

Picture Matching task.

Descriptive data. Results obtained from the PMT were largely similar to the ones obtained from GJT. Table 6 presents distributions of scores for overall condition, congruent condition, and incongruent condition across the early AoE and late AoE groups and the control group. The task did not include D/N agreement condition and N/A agreement condition because of the nature of the task. A preliminary inspection of scores in Table 6 reveals that the control group was considerably more accurate than the two experimental groups, and that the early AoE group is the only group whose scores were modulated by Congruency. There were 40 research items in the PMT, which means that the total possible score that the participants could get on the task is 40. There were 20 items in each of the two conditions – 20 congruent items and 20 incongruent items, which means that the total possible score that the participants could get in each condition is 20.

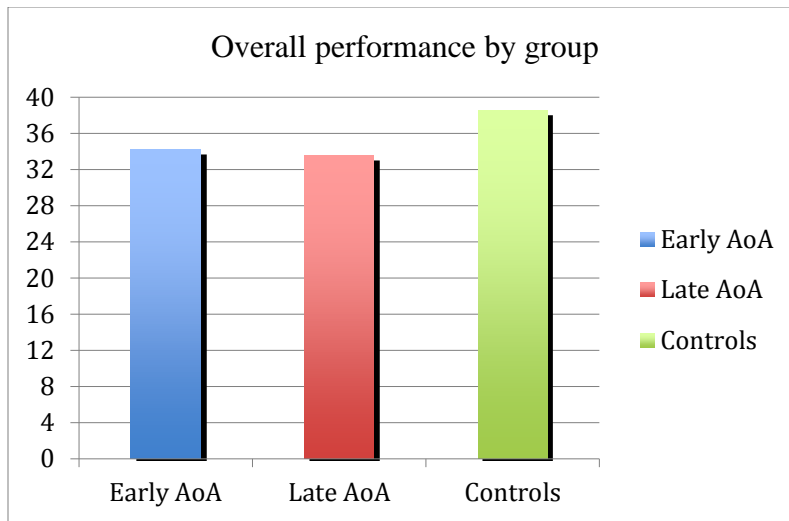
Table 6. *Descriptive data. Group mean scores and total possible scores (indicated next to the means) for the task for overall condition (total possible score 40), D/N agreement condition (total possible score 20), N/A agreement condition (total possible score 20), congruent condition (total possible score 20), and incongruent condition (total possible score 20) across the two AoE groups and the control group*

	Mean group scores / total possible score		
	Early AoE	Late AoE	Controls
Overall items	34.16 / 40	33.54 / 40	38.54 / 40
Congruent items	18.34 / 20	16.54 / 20	19.08 / 20
Incongruent items	15.84 / 20	17.00 / 20	19.54 / 20

To find out whether these preliminary observations were statistically reliable, a series of statistical tests was carried out, which are reported below.

Statistical analyses. First, I compared the early AoE group (L1 Russian), the AoE group (L1 English) and the control group on their overall scores. Tests for normality (Kolmogorov-Smirnov test) showed that the data in all three groups were normally distributed. A graphic presentation of the overall scores is provided in Fig. 8.

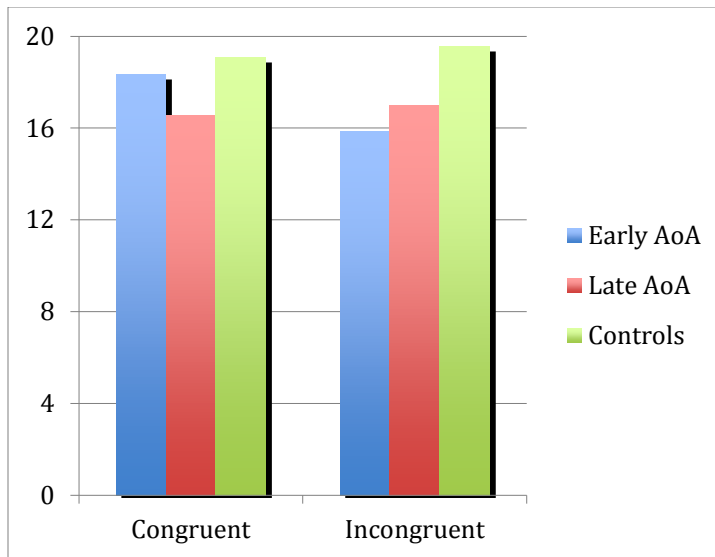
Figure 8. *Group mean scores for the overall condition (total possible score 40) across the two AoE groups and the control group.*



A one-way ANOVA revealed a significant main effect for Group, $F(2,84)=6.057$, $p=.004$, and post hoc tests (Dunnett T3) showed no significant differences between late and early AoE groups, $p=.868$. Both groups were less accurate than the control group, $p=.000$. Thus, early Age of Exposure did not lead to an enhanced *surface* gender agreement on this task, and by surface gender agreement I mean the learners' ability to match gender morphemes of a specific language according to their values while producing or processing DPs in that language.

The next step was to reveal any potential interactions between Age of Exposure and Congruency. Group means across the four conditions are presented in Figure 9.

Figure 9. *Group mean scores for congruent and incongruent conditions (total possible score 20) across the two AoE groups and the control group.*



A repeated measures ANOVA with a 2 (Congruency) x 3 (Age of Exposure) factorial design revealed no significant main effect for Congruency, $F(1,83)=3.551$, $p=.063$, but it showed a significant interaction between Congruency and AoE, $F(2,83)=17.372$, $p=.000$, and subsequent paired samples t-tests detected that the only group affected by Congruency (that is, the group that performed more accurately on congruent than on incongruent items) was early AoE group, $p=.000$. ANOVA also revealed a main effect for the between-subjects variable, and post hoc showed that this effect came from differences between the two experimental groups, on the one hand, and the control group, on the other hand, and not from any statistically reliable differences between the early AoE group and the late AoE group.

Proficiency.

Grammaticality Judgment task.

Descriptive Data. The same set of tests was conducted for Proficiency as for Age of Exposure. Table 7 presents the distribution of scores (group means scores) for overall condition, D/N agreement condition, N/A agreement condition, congruent condition, and incongruent condition across the four proficiency groups (low proficiency English, low proficiency Russian, high proficiency English, high proficiency Russian) and the control group.

Table 7. *Descriptive data. Group mean scores and total possible scores (indicated next to the means) for overall condition (total possible score 48), D/N agreement condition (total possible score 24), N/A agreement condition (total possible score 24), congruent items condition (total possible score 24), and incongruent condition (total possible score 24) across the four proficiency groups and the control group*

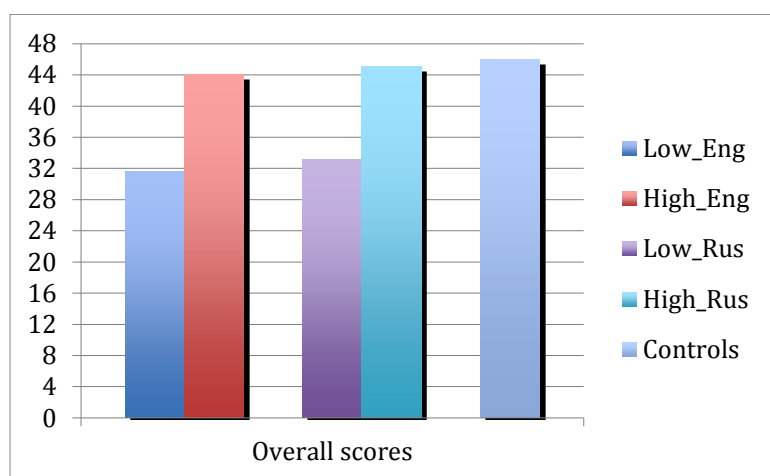
	<i>Mean group scores / total possible score</i>				
	Low proficiency		High proficiency		Controls
	English	Russian	English	Russian	
Overall	31.58 / 48	33.11 / 48	44.06 / 48	45.05 / 48	46.00 / 48
D/N agr	16.74 / 24	17.32 / 24	22.06 / 24	23.16 / 24	23.31 / 24
N/A	16.53 / 24	17.37 / 24	21.75 / 24	21.58 / 24	23.31 / 24
Congr	15.16 / 24	18.00 / 24	21.50 / 24	22.79 / 24	22.85 / 24
Incongr	15.21 / 24	15.11 / 24	21.75 / 24	22.16 / 24	23.15 / 24

As hypothesized, the two high proficiency groups and the control group appear to behave more accurately than the two low proficiency groups. Furthermore, while the high proficiency groups do not appear to be different from each other, the low proficiency L1 Russian group exhibits a slightly higher mean in Congruent condition than the low proficiency L1 English group. Importantly, Type of Agreement (D/N vs. N/A) does not

seem to have affected the scores, although the high proficiency Russian L1 group has a slightly higher mean on D/N than on N/A agreement. I conducted a set of statistical tests in order to find out whether these preliminary observations were statistically reliable.

Statistical analyses. First, I compared the overall scores of each of the proficiency groups and the control group (See Fig. 10).

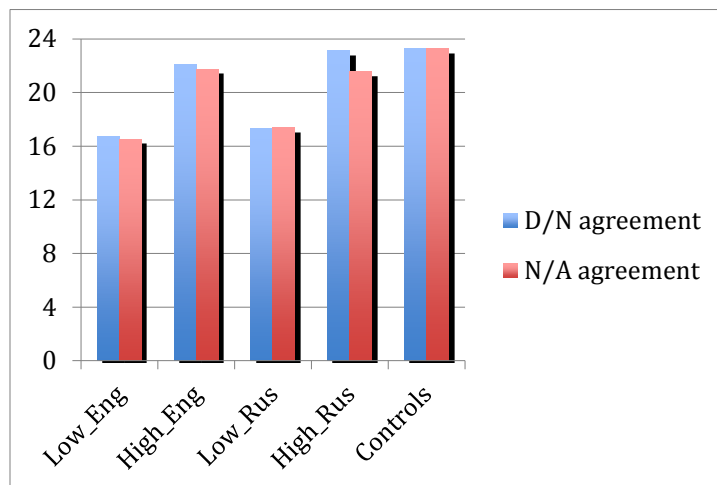
Figure 10. *Group mean scores for the overall condition (total possible score 48) across the four proficiency groups and the control group*



A one-way ANOVA showed a significant main effect, $F(4,81)=34.442, p=.000$ (Levene's $p = .000$). Post Hoc tests revealed no differences between the two low proficiency groups, $p=1.000$, no differences between the two high proficiency groups, $p=1.000$, no differences between the two high proficiency groups and the control group, $p = 1.000$, but they showed a statistically reliable difference between the low proficiency groups, on the one hand, and the high proficiency groups and the control group, on the other hand, $p = .000$. That is, the higher proficiency groups were significantly more accurate than the lower proficiency groups in their overall performance on the task.

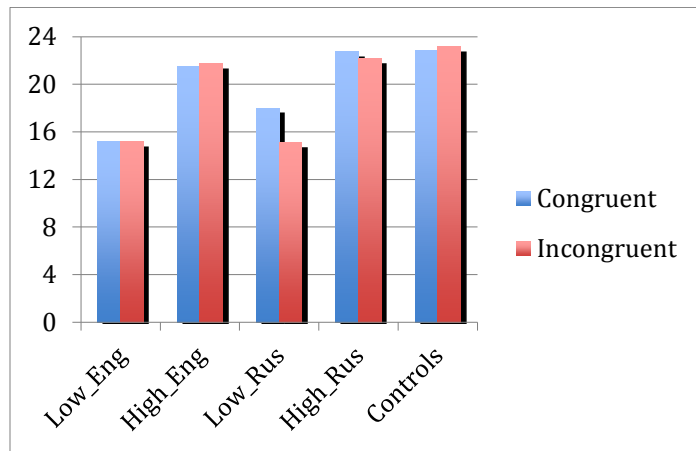
In order to reveal potential interactions between Proficiency and Type of Agreement and Congruency, I conducted a repeated measures ANOVA with a 2 (Congruency) x 2 (Type of Agreement) x 5 (Group) factorial design. Tests for normality (Kolmogorov-Smirnov Test) showed that data from each of the five groups for all of the variable levels (Congruent, Incongruent, D/N agreement, N/A agreement) were normally distributed, $p > .05$. Fig. 11 and Fig. 12 present the groups scores across the within-subjects conditions.

Figure 11. *Group mean scores for D/N vs. N/A conditions (total possible score 24) across the four proficiency groups and the control group*



The ANOVA showed no significant main effect for Agreement Type, $F(1,81)=1.929$, $p=.169$ and no significant interaction between Agreement Type and Group (the four different proficiency groups and a control group), $F(4,81)=.512$, $p=.727$, which means that the higher frequency of occurrence of determiner/noun sequences vs. noun/adjective sequences in the input did not lead any of the groups to a more accurate gender agreement on the task.

Figure 12. *Group mean scores for congruent vs. incongruent conditions (total possible score 24) across the four proficiency groups and the control group*



On the other hand, the test revealed a significant main effect for Congruency, $F(1,81)=7.866, p=.006$, as well as a significant interaction between Congruency and Group, $F(4,81)=3.342, p=.014$. Expectedly, series of paired samples t-tests failed to reveal a statistically reliable difference between scores on congruent and incongruent items for the low-proficiency English L1 group, $t(18)=-.094, p=.929$, for the high-proficiency L1 English, $t(15)=-.939, p=.362$, or the control group $t(12)=-.805, p=.436$. Crucially, while the low proficiency Russian group behaved significantly more accurately on congruent than on incongruent items, $t(18)=4.811, p=.000$, the high proficiency L1 Russian group was not affected by Congruency, $t(18)=1.935, p=.069$. This seems to indicate that the L1 lemma does affect L2 gender performance at the early stages of gender acquisition, but the effect fades away with proficiency.

In addition, there was a significant effect for Group, $F(4,81)=56.655, p=.000$. Post hocs (Dunnett T3) failed to show statistically significant differences between the two low proficiency groups, $p=.886$, but they showed significant differences between the low proficiency groups, on the one hand, and the high proficiency groups, on the other hand,

$p = .000$. Moreover, the two high proficiency groups were not significantly different from each other, $p = 1.000$, and from the control group, $p = .485$ for L1 English group and $p = 1.000$, for L1 Russian group. Thus, proficiency did modulate L2 performance on gender agreement in this task.

Picture matching task.

Descriptive data. Distributions of scores (group mean scores) for overall condition, congruent condition, and incongruent condition across the four proficiency groups and the control group are presented in Table 8.

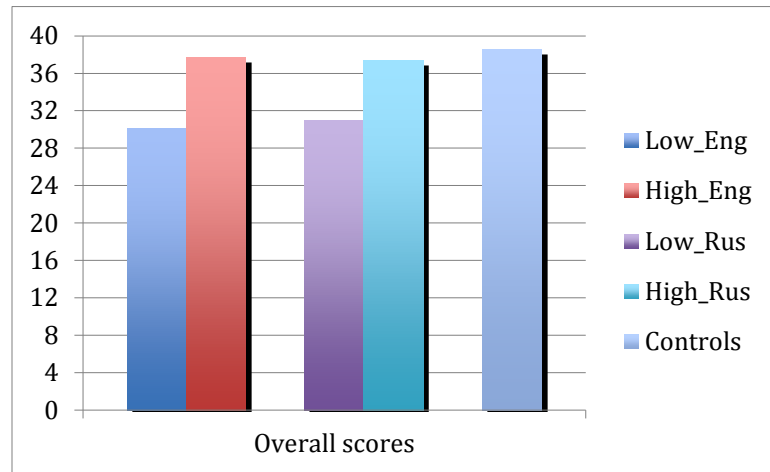
Table 8. *Descriptive data. Group mean scores and total possible scores (indicated next to the means) for overall condition (total possible score 40), congruent condition (total possible score 20), and incongruent condition (total possible score 20) across the four proficiency groups and the control group*

	<i>Mean group scores / total possible score</i>				
	Low proficiency		High proficiency		Controls
	English	Russian	English	Russian	
Overall	30.05 / 40	30.95 / 40	37.69 / 40	37.37 / 40	38.54 / 40
Congr.	14.84 / 20	17.16 / 20	18.56 / 20	19.53 / 20	19.08 / 20
Incong.	15.21 / 20	13.84 / 20	19.13 / 20	17.84 / 20	19.54 / 20

In order to find out whether the differences between the scores reflected in the Table 8 were statistically significant, I ran a series of statistical analyses for the four proficiency groups and the control group.

Statistical analyses. First, I compared the groups' overall scores. A graphic presentation of the overall scores is provided in Fig. 13.

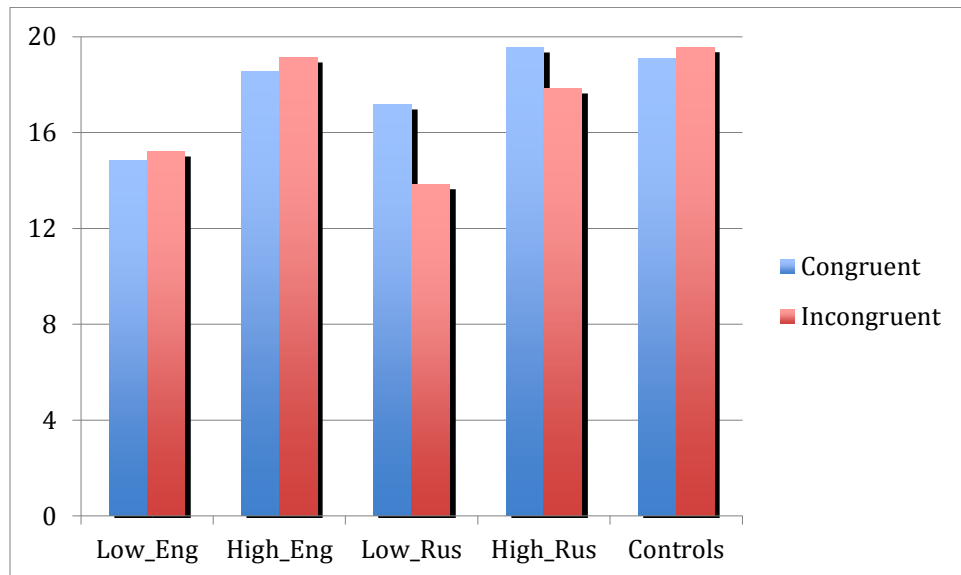
Figure13. *Group mean scores for the overall condition (max score 40) across the four proficiency groups and the control group*



To find out whether the differences in scores reflected in Fig. 13 were statistically reliable, I compared the groups' overall scores in a one-way ANOVA. The test revealed a significant main effect, $F(4,81)=25.420$, $p=.000$. Post Hoc tests (Dunnett T3) showed no statistically reliable differences between the two low proficiency groups, $p=.999$, between the two high proficiency groups, $p=1.000$, between the high proficiency L1 English group and the control group, $p=.951$, and between the high proficiency L1 Russian group and the control group, $p=.270$. However, the low proficiency groups were shown to be significantly less accurate than the high proficiency groups and the control group. Thus, proficiency modulated accuracy on gender agreement in this task as well as in the GJT.

Next step was to determine whether Congruency modulated accuracy in this task for any of the groups. Fig. 14 provides a graphic presentation of the group scores across congruent vs. incongruent conditions.

Figure 14. *Group mean scores for congruent condition and incongruent condition (total possible score 20) across the four proficiency groups and the control group*



A repeated-measures ANOVA with a 2 (Congruency) x 5 (Group) factorial design revealed a significant main effect for Congruency, $F(1,81)=8.486$, $p = .005$ and a significant interaction between Congruency and Group, $F(4,81)=10.238$, $p = .000$. There was also a main effect for the between subjects variable Group, $F(4,81)=25.083$, $p = .000$. Post hoc tests (Dunnett T3) showed that L1 English and L1 Russian low proficiency groups were not significantly different from each other, $p = .999$, while they both were significantly less accurate than the L1 English high proficiency group, $p = .000$, and L1 Russian high proficiency group, $p = .000$. Importantly, both high proficiency groups were not significantly different from each other, $p = 1.000$, or from the control group, $p = .919$ for L1 English group and $p = .202$, for L1 Russian group.

A series of paired samples t-tests showed that congruency affected both Russian L1 groups ($p = .000$ for the low proficiency group and $p = .002$ for the high proficiency group), but not the L1 English groups or the control group.

Reaction times.

Similarly to the results for Accuracy, the results for Reaction Times are reported for two different variables – Age of Exposure and Proficiency for each of the research tasks (GJT and PMT). Again, I will first present the results for Age of Exposure from the GJT and PMT and then the results for Proficiency from the GJT and PMT.

Age of Exposure.

Grammaticality Judgment Task.

Descriptive data. Table 9 presents the distribution of RTs (mean RTs and standard deviations in milliseconds) for overall condition, grammatical condition, ungrammatical condition, D/N agreement condition, N/A agreement condition, congruent condition, and incongruent condition across the early AoE (L1 Russian high and low proficiency groups), the late AoE groups (L1 English high and low proficiency groups) and the control group.

Table 9. *Descriptive data. Group mean RTs in milliseconds for the overall condition, grammatical condition, ungrammatical condition, D/N agreement type condition, N/A agreement type condition, congruent condition, and incongruent condition across the AoE groups and the control group*

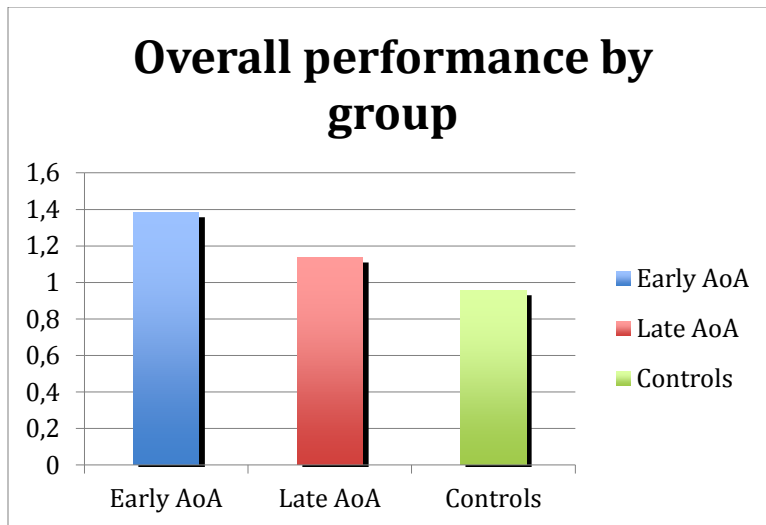
	<i>Mean RTs</i>		
	Early AoE	Late AoE	Controls

Overall RTs	1.383229	1.137112	0.955618
Gramm. RTs	1.243517	0.979727	0.748202
Ungram. RTs	1.713084	1.407122	1.171085
D/N agr. RTs	1.383265	1.115182	0.947471
N/A agr. RTs	1.367369	1.148884	0.976393
Congruent RTs	1.331228	1.103141	0.955258
Incongr. RTs	1.440527	1.157474	0.954331

First and foremost, the data in the table indicate that all of the groups were affected by Grammaticality, and such slower reactions to ungrammatical items compared to grammatical ones imply that the participants did in fact process gender agreement violations. Second, quite surprisingly, the early AoE group (L1 Russian) seems to have reacted slower than the late AoE group (L1 English). The control group appears to have reacted faster than the two experimental groups in all conditions. Furthermore, while none of the groups appear to be affected by the Type of Agreement, Congruency seems to have affected the high proficiency L1 Russian group. In order to determine whether the between-subjects and within-subjects differences were significant, a set of statistical analyses was carried out.

Statistical analyses. First, I compared the overall RTs (in milliseconds) from the early AoE group, late AoE group, and the control group. Group mean RTs are presented in Fig. 15.

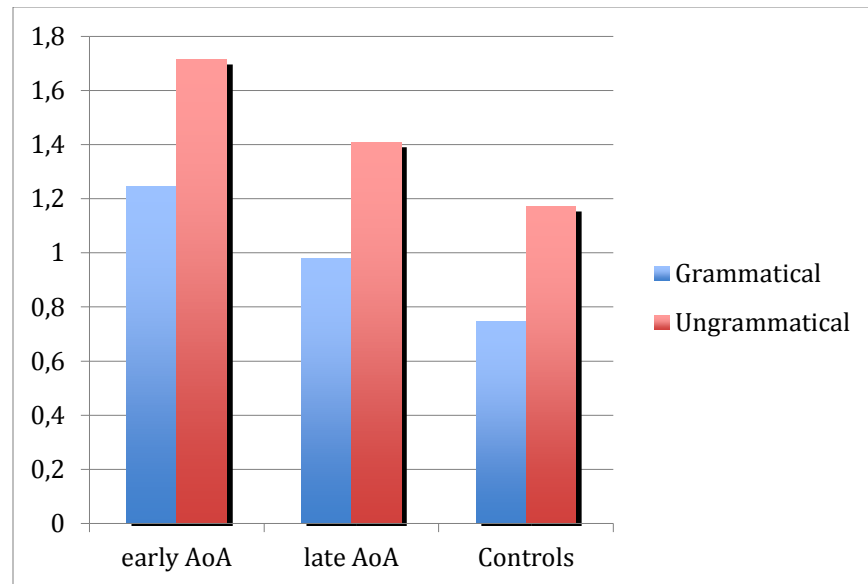
Figure 15. *Group mean RTs in milliseconds for the overall condition across the AoE groups and the control group*



One-way ANOVA revealed a significant main effect, $F(2,83)=12.404$, $p=.000$ (Levene's $p = .039$). Post Hoc tests (Dunnett T3) showed that the differences between the late AoE group (L1 English) and the control group were not statistically significant, $p=.185$. On the other hand, the early AoE group (Russian) was significantly slower than the control group, $p=.000$, and than the late AoE group (L1 English), $p=.002$. Thus, early Age of Exposure did not lead to a faster processing of gender agreement; in fact, late AoE group exhibited significantly shorter RTs than the early AoE group.

Next, in order to ascertain that Grammaticality affected RTs (which would indicate that the groups in fact processed gender agreement violations), I ran a repeated measures ANOVA with a 2 (Grammaticality) x 3 (Groups) factorial design (See Fig. 16).

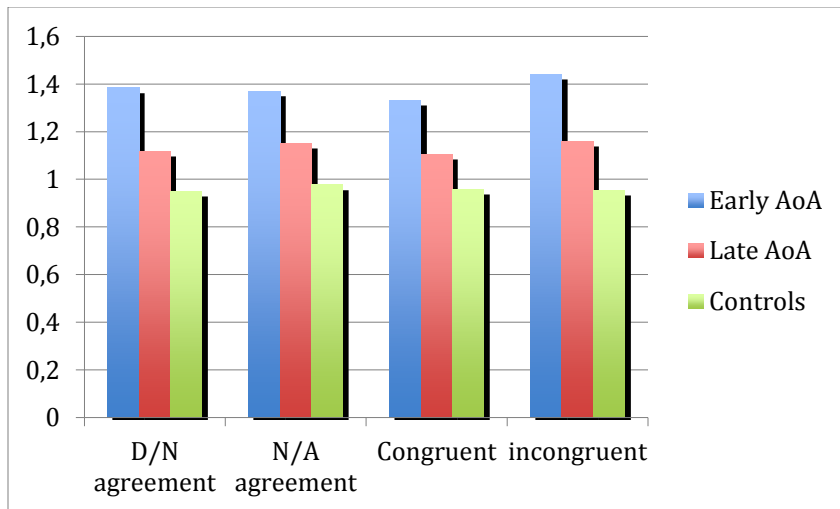
Figure 16. *Group mean RTs in milliseconds for grammatical and ungrammatical conditions across the AoE groups and the control group*



The ANOVA showed a significant main effect for Grammaticality, $F(1,83)=104.103$, $p=.000$, but it did not reveal any significant interaction between Grammaticality and Age of Exposure, $F(2,83)=.101$, $p=.854$, which indicates that all of the groups were affected by the within subject variable. A set of paired samples t-tests (one test per each of the three groups) revealed that all groups reacted significantly slower to the ungrammatical items compared to the grammatical ones, $p=0.000$ for all three groups. Thus, the statistical tests confirmed the preliminary observations made on the basis of the descriptive data.

The next step was to reveal any potential interactions between Age of Exposure and the linguistic variables Type of Agreement and Congruency. Group means across the four conditions are presented in Fig. 17.

Figure 17. *Group mean RTs in milliseconds for D/N agreement condition, N/A agreement condition, congruent condition, and incongruent condition across the AoE groups and the control group*



A repeated measures ANOVA with a 2 (Type of Agreement) x 2 (Congruency) x 3 (Age of Exposure) factorial design showed no significant effect for Type of Agreement, $F(1,83)=0.002$, $p=.968$, but it showed a significant main effect for Congruency, $F(1,83)=5.442$, $p=.022$. A series of paired samples t-tests revealed that the only group affected by Congruency was the early AoE group (L1 Russian), $p=0.004$. The ANOVA also revealed a significant main effect for the between-subjects variable Age of Exposure, $F(2,83)=9.546$, $p=.000$, and Post Hoc tests (Dunnett T3) confirmed the unexpected finding made on the basis of observation of the descriptive data – the early AoE group (L1 Russian) reacted significantly slower than the late AoE group (L1 English), $p=.001$, and than the control group, $p=.000$, but the late AoE group was not significantly different from the control group, $p=.181$.

Picture Matching task.

Descriptive data. Table 10 presents distributions of RTs in milliseconds (means and standard deviations) for overall condition, congruent condition, and incongruent

condition across the early AoE and late AoE groups and the control group. A preliminary inspection of scores in Table 6 reveals two major trends. First, and similarly to the results on GJT, the early AoE (L1 Russian) appears to be the slowest, followed by the late AoE (L1 English) group with the control group reacting the fastest. Second, neither the late AoE group nor the control group appears to discriminate between congruent and incongruent items, while the early AoE group (L1 Russian) exhibits a very clear congruency effect.

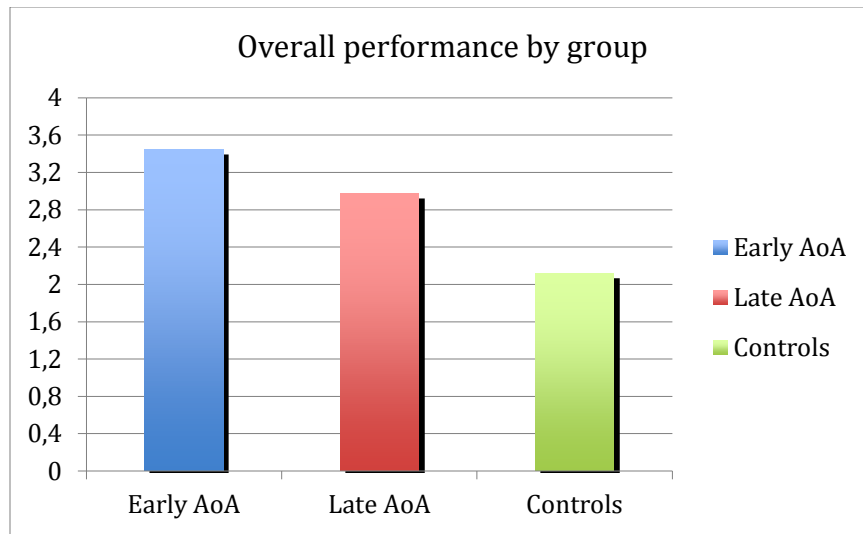
Table 10. *Descriptive data. Group mean RTs in milliseconds for the overall condition, congruent condition, and incongruent condition across the AoE groups and the control group*

	<i>Mean RTs</i>		
	Early AoE	Late AoE	Controls
Overall RTs	3.445836	2.975440	2.119563
Congruent RTs	3.203132	2.915313	2.099708
Incongruent RTs	3.868293	2.984626	2.126162

To find out whether these preliminary observations were statistically reliable, a series of statistical tests was carried out, which are reported below.

Statistical analyses. First, I compared the early AoE group (L1 Russian), the AoE group (L1 English) and the control group on their overall RTs. A graphic presentation of the overall RTs is provided in Fig. 18.

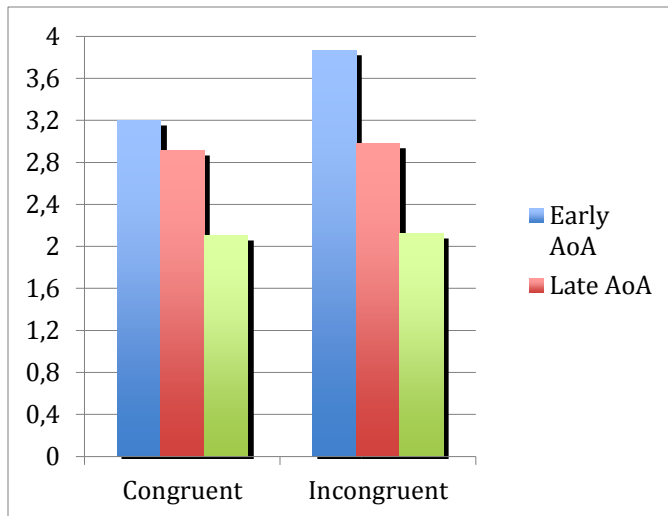
Figure 18. *Group mean RTs in milliseconds for the overall condition across the AoE groups and the control group*



A one-way ANOVA revealed a significant main effect for Group, $F(2,83)=17.933$, $p=.000$, (Levene's $p=.119$) and post hoc tests (Bonferroni) confirmed the preliminary observation made on the basis of the descriptive data – the differences reflected in Fig.18 between the early and late AoE group were statistically reliable, $p=.015$, as well as the differences between the late AoE and the control group, $p=.001$. Thus, early Age of Exposure did not lead to decreased RTs; on the contrary, the early AoE group reacted significantly slower than the late AoE group.

The next step was to reveal any potential interactions between Age of Exposure and the linguistic variables Congruency. Group means across the two conditions are presented in Fig. 19.

Figure 19. *Group mean RTs in milliseconds for congruent and incongruent conditions across the AoE groups and the control group*



A repeated measures ANOVA with a 2 (Congruency) x 3 (Age of Exposure) factorial design revealed a significant main effect for Congruency, $F(1,83)=9.747$, $p=.002$, and a significant interaction between Congruency and AoE, $F(2,83)=8.699$, $p=.000$. Subsequent paired samples t-tests detected that the only group affected by Congruency was early AoE group, $t(37)=-4.366$, $p=.000$. The ANOVA also revealed a main effect for the between-subjects variable, and post hocs (Bonferroni) showed that the differences between all groups were significant. That is, the late AoE group was significantly faster than the early AoE group, $p=.015$, and the control group was significantly faster than both the late ($p=.001$) and the early AoE groups ($p=.000$).

Proficiency.

Grammaticality Judgment task.

Descriptive Data. The same set of tests was conducted for Proficiency as for Age of Exposure. Table 11 presents the reaction times in milliseconds for overall condition, grammatical condition and ungrammatical condition, Table 12 presents the reaction times for congruent condition and incongruent condition, and Table 13 presents reactions times

for D/N agreement condition and N/A agreement condition across the four proficiency groups (low proficiency English, low proficiency Russian, high proficiency English, high proficiency Russian) and the control group.

Table 11. *Descriptive data. Group mean RTs in milliseconds for overall condition, grammatical condition and ungrammatical condition across the four proficiency groups and the control group*

	<i>Mean RTs</i>		
	Overall RTs	Grammatical RTs	Ungrammatical RTs
English Low	1.264796	1.0946344	1.6387672
Russian low	1.51973	1.3997767	2.0163551
English high	0.985488	.84326548	1.1320436
Russian high	1.246728	1.0872576	1.4098128
Controls	0.955618	.74820297	1.171085

Table 11 allows one to make four main observations. First, similarly to the results on Age of Exposure, all five groups seem to have shown sensitivity to gender agreement violations, which has revealed itself in slower reaction times for ungrammatical condition compared to the grammatical condition. Second, while low proficiency groups seem to have reacted slower than their respective high proficiency groups and the control group, the high proficiency L1 Russian group also seems to have reacted slower than the high proficiency English group and the control group. On the other hand, the high proficiency L1 English group RTs do not appear to differ from those of the control group. In general and most importantly, in their respective proficiency groups (low Russian/low English

and high Russian/high English) Russian L1 speakers clearly reacted slower than the English L1 speakers *despite the fact that Russian has gender, and English does not.*

Table 12. *Descriptive data. Group mean RTs in milliseconds for congruent condition and incongruent condition across the four proficiency groups and the control group*

	<i>Mean RTs</i>	
	Congruent RTs	Incongruent RTs
English Low	1.220024	1.306378
Russian low	1.469215	1.578909
English high	0.9640799	0.980651
Russian high	1.193239	1.302146
Controls	0.955258	0.954331

The data in Table 12 seem to indicate that congruency affected both L1 Russian groups. Both high and low proficiency group means for congruent items seem to be shorter than the means for incongruent items, which reveals a processing difficulty for the latter items most likely due to the interference from the L1 (Russian) gender values.

Table 13. *Descriptive data. Group mean RTs in milliseconds for D/N agreement type and N/A agreement type across the four proficiency groups and the control group*

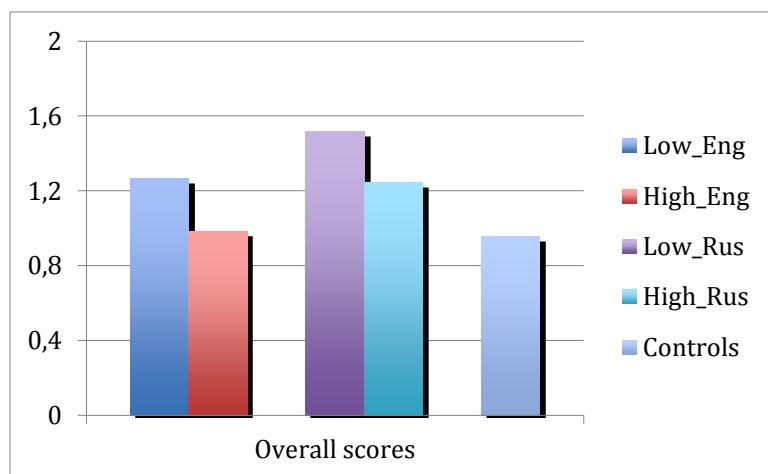
	<i>Mean RTs</i>	
	D/N agreement RTs	N/A agreement RTs
English Low	1.251681	1.241197
Russian low	1.541428	1.473038
English high	0.953089	1.039262

Russian high	1.225101	1.261699
Controls	0.947470	0.976396

The data in Table 13 indicate that Type of Agreement did not affect any of the groups. I conducted a set of statistical tests in order to find out whether these preliminary observations for Grammaticality, Congruency, Type of Agreement and Group were statistically reliable.

Statistical analyses. First, I compared the overall RTs of each of the proficiency groups and the control group (See Fig. 20).

Figure 20. *Group mean RTs in milliseconds for the overall condition across the four proficiency groups and the control group*

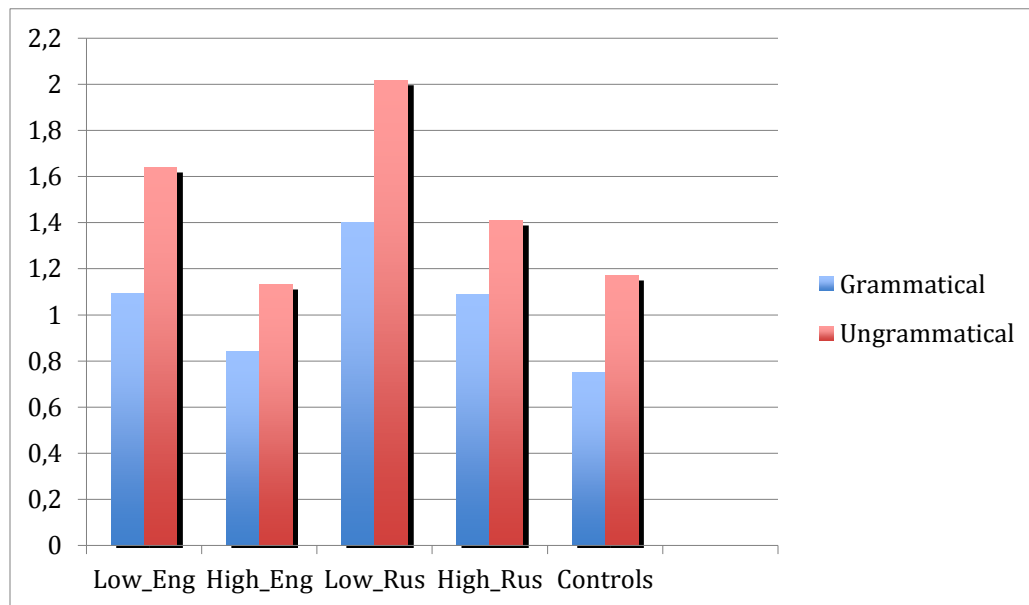


A one-way ANOVA showed a significant main effect, $F(4,81)=12.287$, $p=.000$ (Levene's $p = .012$). Post Hoc tests (Dunnett T3) revealed no significant differences between the two low proficiency groups, $p=0.256$; however, the difference between the low proficiency L1 Russian and the high proficiency L1 Russian groups also failed to reach significance, $p=.081$. Moreover, the high proficiency L1 Russian group reacted significantly slower than the control group, $p=.000$, and even slower than the high

proficiency L1 English group, $p = 0.05$. The high proficiency L1 English group, on the other hand, reacted as fast as the control group, $p = 1.000$.

Next, in order to ascertain that Grammaticality affected the RTs, I ran a repeated measures ANOVA with a 2 (Grammaticality) x 5 (Groups) factorial design (See Fig. 21).

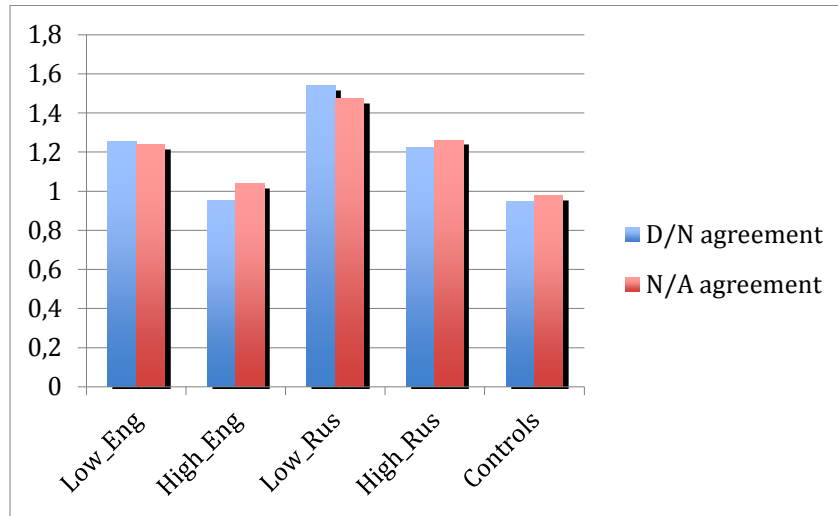
Figure 21. *Group Mean RTs in milliseconds for grammatical and ungrammatical conditions across the four proficiency groups and the control group*



The ANOVA showed a significant main effect for Grammaticality, $F(1,81)=16.204$, and a set of paired samples t-tests (one test per each of the five groups) revealed that all groups reacted significantly slower to the ungrammatical items compared to the grammatical ones, $p=0.000$ for all five groups.

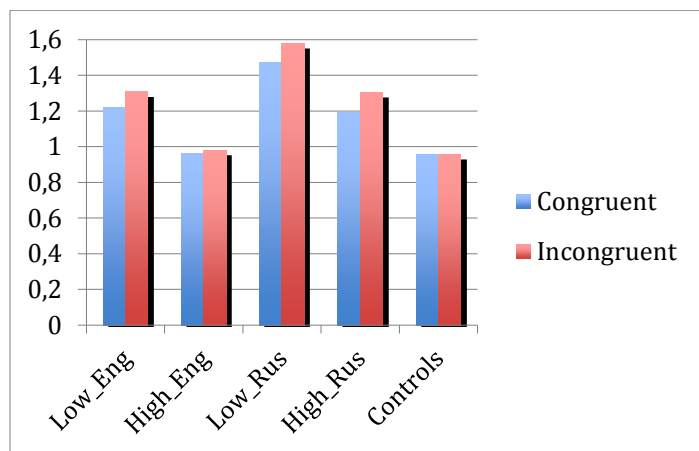
The next step was to reveal any potential interactions between the acquisitional variable Proficiency and the linguistic variables Type of Agreement and Congruency. I conducted a repeated measures ANOVAs with a 2 (Type of Agreement) x 2 (Congruency) x 5 (Group) factorial design. Fig. 22 and Fig. 23 present the group RTs across the within-subjects conditions.

Figure 22. *Group Mean RTs in milliseconds for D/N vs. N/A conditions across the four proficiency groups and the control group*



The slight differences reflected in the D/N vs. N/A bars in Fig. 22 were not statistically significant, because the ANOVA showed no significant main effect for Agreement Type, $F(1,81)=0.012$, $p=.913$, nor a significant interaction between Agreement Type and Group (the four different proficiency groups and a control group), $F(4,81)=.359$, $p=.837$. This means that the Type of Agreement did not affect any of the groups' reaction times.

Figure 23. *Group Mean RTs in milliseconds for congruent vs. incongruent conditions across the four proficiency groups and the control group*



On the other hand, the test revealed a significant main effect for Congruency, $F(1,81)=8.482, p=.005$. Expectedly, a series of paired samples t-tests failed to reveal a statistically reliable difference between scores on congruent and incongruent items for the low-proficiency English L1 group, $t(18)=-1.479, p=.156$, the high-proficiency L1 English, $t(15)=-.863, p=.402$, or the control group $t(12)=0.12, p=.979$. Interestingly, while the low proficiency Russian group was not sensitive to Congruency in terms of RTs, $t(18)=-1.828, p=.084$, the high proficiency L1 Russian group was, $t(18)=-2.759, p=.013$. Plausible explanations to account for such a reversed effect for RTs compared to Accuracy will be provided in the Discussion chapter.

In addition, there was a significant effect for Group, $F(4,81)=56.655, p=.000$. Post hocs (Dunnett T3) showed that the high proficiency L1 English group and the control group reacted with a similar speed, $p=1.000$, and that they both were faster than the high proficiency L1 Russian group, $p=.004$ for high proficiency L1 English group and $p=.000$ for the control group. Moreover, the high proficiency L1 Russian group was not faster than the low proficiency L1 English group, although it was faster than the low proficiency L1 Russian group, which seems to indicate that higher proficiency in the L2 leads to shorter RTs, but on the condition that the L1 background is the same. At the same time, the differences in RTs between the two low proficiency groups were not statistically significant, $p=0.153$.

Picture Matching task.

Descriptive data. Distributions of scores for overall condition, congruent condition, and incongruent condition across the four proficiency groups and the control group are presented in Table 14.

Table 14. *Group mean RTs for overall RTs, congruent RTs, and incongruent RTs across the four proficiency groups and the control group*

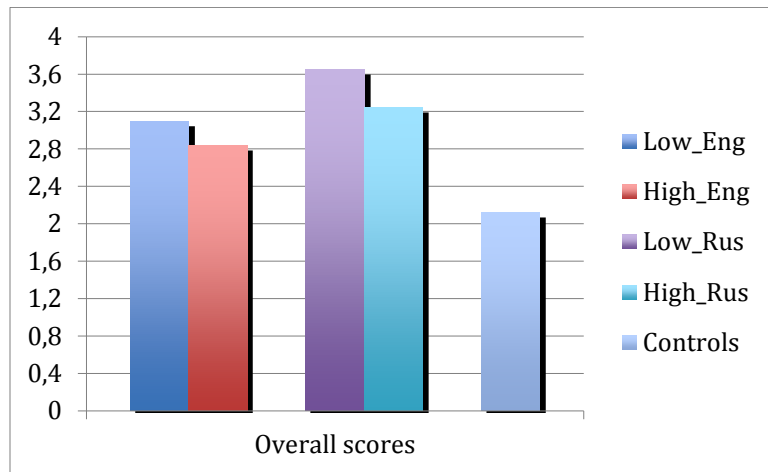
	Overall	Congruent	Incongruent
	<i>Ms</i>	<i>Ms</i>	<i>Ms</i>
English Low	3.092480	3.009807	3.076527
Russian low	3.6513197	3.354132	4.282139
English high	2.836456	2.803102	2.875495
Russian high	3.240475	3.052032	3.454447
Controls	2.119563	2.099708	2.126262

The data reflected in Table 14 seem to indicate that the control group reacted faster than the experimental groups. Furthermore, both L1 English groups (low and high proficiency) seem to have reacted faster than the L1 Russian high proficiency group, and the L1 Russian low proficiency group appears to be the slowest of all groups. Moreover, both L1 Russian groups (low and high proficiency) showed a robust congruency effect, but the high proficiency group was affected to a lesser extent than the low proficiency group (congruent $M = 3.052032$ vs. incongruent $M = 3.454447$ for the high proficiency group and congruent $M = 3.354132$ vs. incongruent $M = 4.282139$ for the low proficiency group).

In order to find out whether the differences between the scores reflected in Table 14 were statistically significant, I ran a series of statistical analyses for the four proficiency groups and the control group.

Statistical analyses. First, I compared the groups' overall RTs. A graphic presentation of the overall RTs is provided in Figure 24.

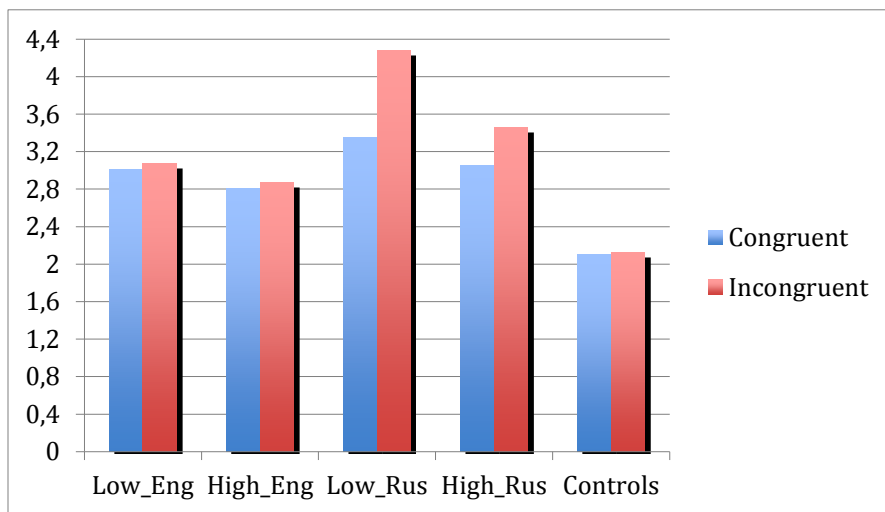
Figure 24. *Group mean RTs in milliseconds for the overall condition across the four proficiency groups and the control group*



To find out whether the differences in scores reflected in Figure 24 were statistically reliable, I compared the groups' overall RTs in a one-way ANOVA. The test revealed a significant main effect, $F(4,81)=25.420$, $p = .000$. Post Hoc tests (Dunnett T3) showed no statistically reliable differences between the two low proficiency groups, $p = .999$, between the two high proficiency groups, $p = 1.000$, between the high proficiency L1 English group and the control group, $p = .951$, and between the high proficiency L1 Russian group and the control group, $p = .270$. However, the low proficiency groups were shown to be significantly slower than the high proficiency groups and the control group. Thus, proficiency modulated accuracy on gender agreement in this task as well as in the GJT.

The next step was to determine whether Congruency modulated RTs in this task for any of the groups. Fig. 25 provides a graphic presentation of the group RTs across congruent vs. incongruent conditions.

Figure 25. *Group mean RTs in milliseconds for congruent vs. incongruent conditions across the four proficiency groups and the control group*



A repeated-measures ANOVA with a 2 (Congruency) x 5 (Group) factorial design revealed a significant main effect for Congruency, $F(1,81)=17.524$, $p = .000$ and a significant interaction between Congruency and Group, $F(4,81)=6.093$, $p = .000$. A series of paired samples t-tests showed that congruency affected both Russian L1 groups, $p = .004$ for the low proficiency group and $p = .000$ for the high proficiency group, but not the L1 English groups, $p=.469$ for the low proficiency group and $p=.347$ for the high proficiency group, or the control group, $p=.690$.

There was also a main effect for the between subjects variable Group, $F(4,81)=11.015$, $p = .000$. Post hoc tests (Bonferroni) showed that the differences between the low proficiency L1 English group, high proficiency L1 English group and high proficiency L1 Russian group did not reach significance, $p=1.000$, but that the low proficiency L1 English group reacted significantly faster than the low proficiency L1 Russian group, $p=.018$. At the same time, the difference between the low proficiency L1 Russian group and the high proficiency L1 Russian group was not significant, $p = .208$. Finally, the control group reacted significantly faster than all groups except for the high

proficiency L1 English group. Thus, proficiency was not the definitive factor influencing RTs.

Experiment 2

Experiment 2 was designed to find out whether L1 English L2 Spanish speakers (the experiment did not include L1 Russian speakers) were affected by the type of gender (lexical-semantic vs. lexical-syntactic) and by morphological cue reliability (reliable cue vs. unreliable cue).

Accuracy. Table 15 presents the distribution of scores (means and standard deviations) for semantic gender condition, lexical gender condition, reliable cue condition and unreliable cue condition across the two experimental groups (low and high proficiency) and the control group. As expected, the L1 English low proficiency group appears to be less accurate than the L1 English high proficiency group and the control group, while the latter two groups do not seem to have behaved differently from each other. Furthermore, performance on semantic gender type and on lexical gender type appears to be similar across all groups, while there are clear differences between reliable and unreliable cue types for the low proficiency group.

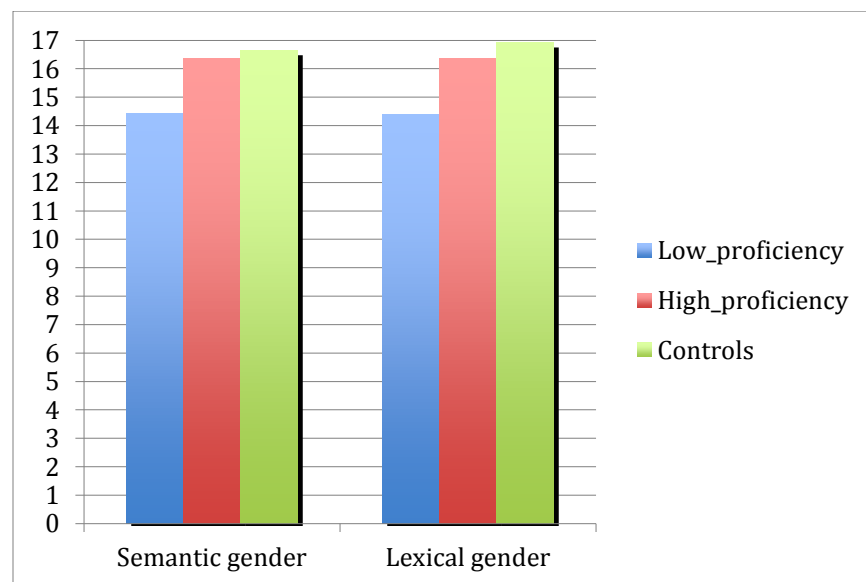
Table 15. *Descriptive data. Group mean scores and total possible scores (indicated next to the means) for semantic gender condition (total possible score 17), lexical gender condition (total possible score 17), reliable cue condition (total possible score 18), and unreliable cue condition (total possible score 18) across the two proficiency groups and the control group*

	Low proficiency	High proficiency	Controls
	<i>M / total possible score</i>	<i>M / total possible score</i>	<i>M / total possible score</i>
Semantic	14.44 / 17	16.38 / 17	16.65 / 17
Lexical	14.39 / 17	16.37 / 17	16.94 / 17

Reliable	15.39 / 18	17.37 / 18	17.94 / 18
Unreliable	10.94 / 18	16.44 / 18	17.29 / 18

Statistical analyses. In order to find out whether the differences observed in the Table 15 were significant, I ran two statistical tests: a repeated measures ANOVA with a 2 (Gender type) x 3 (Group) factorial design and a repeated measures ANOVA with a 2 (Cue type) x 3 (Group) factorial design⁴. Graphic presentation of group means across the within-subjects conditions is provided in Fig. 26 for Gender Type and in Fig. 27 for Cue Type.

Figure 26. *Group mean scores for semantic gender vs. lexical gender conditions (total possible score 17) across the two proficiency groups and the control group*



The ANOVA for Gender type revealed no significant effect for Gender Type,

$F(1,48)=44.987, p > .05$ nor a significant interaction between Gender Type and Group,

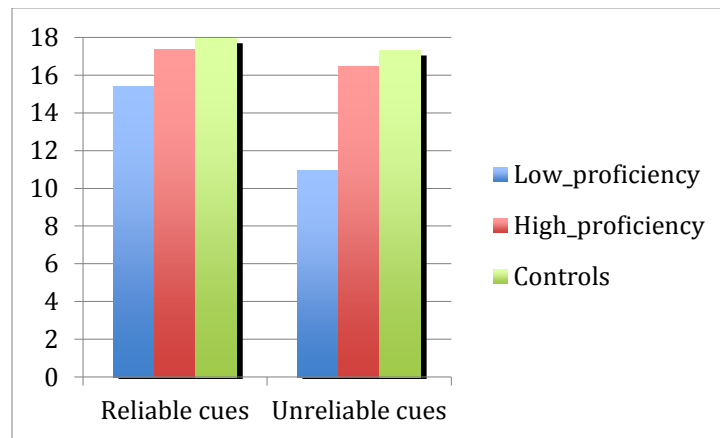
$F(2,48)=17.121, p > .05$, which means that the participants did not discriminate between

⁴ Tests for normality (Kolmogorov-Smirnov test) showed that 5 out of 12 data sets deviated from normality. However, some authors report that ANOVA is not sensitive to non-normality (Glass et al., 1972; Harwell et al., 1992; Lix et al., 1996), which is why it is acceptable to use the parametric test with moderately non-normal data.

semantic vs. lexical gender nouns. In other words, lexical-semantic gender did not facilitate L2 gender agreement in this experiment.

However, the test did reveal a significant main effect for Group, $F(1,48)=43.804$, $p=.000$. Post hoc tests (Dunnett T3) showed that the high proficiency group did not differ from the controls, $p=.212$, but that the low proficiency group behaved significantly less accurately than the controls, $p=.003$ and than the high proficiency group, $p=.015$.

Figure 27. *Group mean scores for reliable cue vs. unreliable cue conditions (total possible score 18) across the two proficiency groups and the control group*



The ANOVA for Cue Type revealed an effect for Cue type, $p=.000$, and Group, $p=.000$, which means that the different groups behaved differently from each other and also they behaved differently on regular and irregular items. Post hoc tests (Dunnett T3) failed to reveal a statistically reliable difference between the high proficiency group and the control group, $p=.099$, but it did reveal that low proficiency learners were significantly less accurate than the high proficiency group and the control group, $p=.000$ in both cases.

After these general main effects were established, I ran a series of additional tests to find out the precise loci of the differences. First, since the repeated measures ANOVA

showed an effect for the Cue Type, I conducted two One-way ANOVAs – one for reliable cue items and another one for unreliable cue items. On reliable cue items, the low proficiency group was significantly less accurate than both the high proficiency group, $p = .007$ and the control group, $p = .000$, while the high proficiency group was as accurate as the control group, $p = 1.000$. Similar results were obtained for unreliable cue items: the controls and high proficiency group did not differ from each other, $p = .602$, but they both were more accurate than the low proficiency group ($p = .000$ in both cases).

Second, I ran three paired samples t-tests to compare each of the group's behavior separately on regular vs. irregular items. The tests showed that cue reliability (or lack thereof) affected all three groups. The low proficiency group behaved significantly more accurately on reliable cue items compared to unreliable cue items, $t(17)=5.788$, $p = .000$, and the high proficiency group showed similar results, $t(15)=2.798$, $p = .014$. This result seems to indicate that syntactic/morphological cue reliability leads to an enhanced ability for gender assignment, which reveals itself in more accurate surface gender agreement. Surprisingly, the control group also behaved similarly to the two experimental groups, $t(16)=3.096$, $p = .007$, displaying more accuracy on regular items. Plausible explanations for this finding will be provided below in the Discussion chapter.

Reaction times. Table 16 presents the distribution of RTs (means and standard deviations) in milliseconds for semantic gender condition, lexical gender condition, reliable cue condition, unreliable cue condition, grammatical condition, and ungrammatical condition across the two experimental groups (low and high proficiency) and the control group.

Table 16. *Group mean RTs in milliseconds for semantic gender condition, lexical gender condition, reliable cue condition, unreliable cue condition, grammatical condition, and ungrammatical condition across the two proficiency groups and the control group*

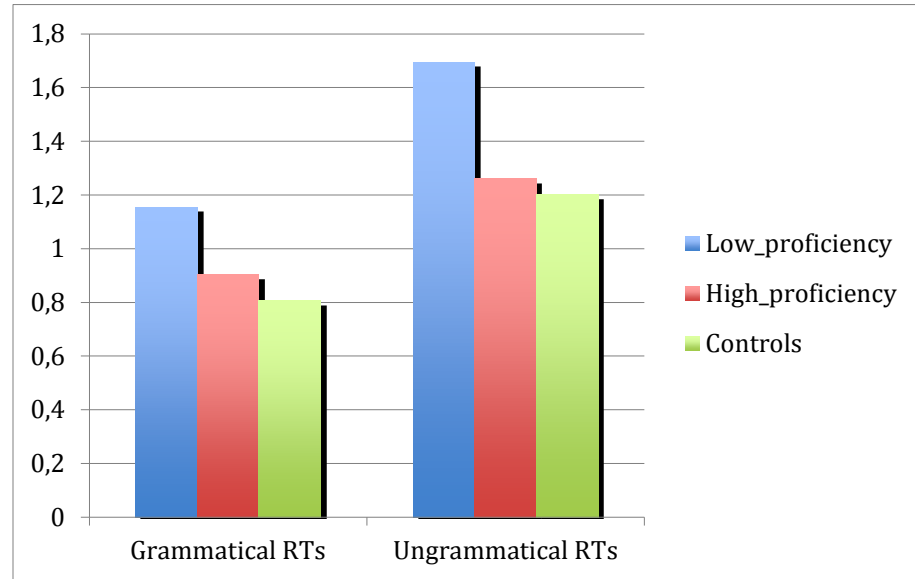
	<i>Mean RTs</i>		
	Low proficiency	High proficiency	Controls
Semantic	1.380116	1.068312	1.009030
Lexical	1.324136	0.993908	0.990175
Reliable	1.303367	0.995254	0.977809
Unreliable	1.494842	1.146679	0.986571
Grammatical	1.155087	0.905548	0.806824
Ungrammatical	1.694924	1.260203	1.201058

The data in Table 16 seem to indicate that the low proficiency group reacted slower than the high proficiency group and the control group, while the latter two groups did not behave differently from each other. Crucially, all groups exhibited a robust grammaticality effect, which indicates that they processed gender agreement violations. Furthermore, RTs on semantic gender type and on lexical gender type appear to be similar across all groups, while the RTs for the unreliable cue condition in the low proficiency group seem to be longer than the RTs for the reliable cue condition. I conducted a series of statistical tests in order to establish whether these preliminary observations were statistically significant.

Statistical analyses. First, in order to find out whether the differences between RTs on grammatical and ungrammatical condition were significantly reliable, I

conducted a repeated measures ANOVA with a 2 (Grammaticality) x 3 (Group) factorial design. Figure 28 presents mean RTs for both conditions across the three groups.

Figure 28. *Group mean RTs in milliseconds for grammatical vs. ungrammatical gender conditions across the two proficiency groups and the control group*



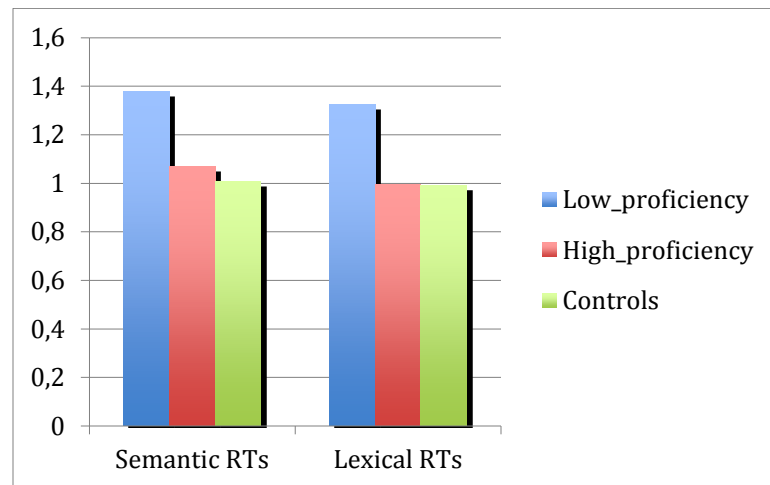
The test revealed a significant main effect for Grammaticality, $F(1,48)=114.884$, $p=.000$, and paired samples t-tests showed that all three groups were affected by Grammaticality: low proficiency group - $t(17)=-6.618$, $p=.000$, high proficiency group - $t(15)=-8.066$, $p=.000$, and the control group, $t(16)=-5.451$, $p=.000$.

The ANOVA also revealed a significant main effect for Group, $F(2,48)=10.483$, $p=.000$, and post hoc tests (Bonferroni) showed that the low proficiency group was significantly slower than the high proficiency group, $p=.004$ and the control group, $p=.000$, but that the high proficiency group was as fast as the control group, $p=1.000$.

The next step was to find out whether the two within subjects variables – Gender Type (semantic and lexical) and the Cue Type (reliable cue and unreliable cue) - affected group RTs. I conducted a repeated-measures ANOVAs with a 2 (Gender type) x 2 (Cue

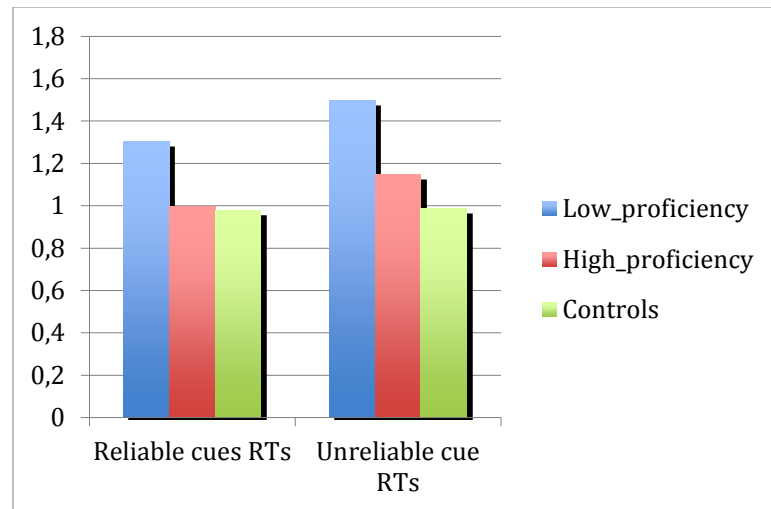
Type) x 3 (Group) factorial design. Graphic presentation of group means is provided in Fig. 29 for Gender Type and in Fig. 30 for Cue Type.

Figure 29. *Group mean RTs in milliseconds for semantic gender vs. lexical gender conditions across the two proficiency groups and the control group*



The ANOVA revealed no significant effect for Gender Type, $F(1,48)=.930$, $p=.340$, nor a significant interaction between Gender Type and Group, $F(2,48)=0.728$, $p=.488$, which means that the participants did not discriminate between semantic vs. lexical gender nouns.

Figure 30. *Group mean RTs in milliseconds for reliable cue vs. unreliable cue conditions across the two proficiency groups and the control group*



Although the bars in Fig. 30 seem to indicate differences between reliable vs. unreliable conditions for the two experimental groups, these differences did not reach significance: the ANOVA did not reveal a significant main effect for Cue Type, $F(1,48)=1.653$, $p=.205$ or a significant interaction between Cue Type and Group, $F(2,48)=.667$, $p=.518$.

However, the test did reveal a significant main effect for Group, $F(1,48)=12.401$, $p=.000$. Post hoc tests (Dunnett T3) revealed that low proficiency learners were significantly slower than the high proficiency group, $p=.001$, and than the control group, $p=.000$, but they failed to establish a statistically reliable difference between the high proficiency group and the control group, $p=1.000$.

To summarize the main findings from Experiment 1: Age of Exposure did not modulate accuracy on neither of the tasks, but Proficiency did. The L1 English low proficiency group was as accurate as the low proficiency L1 Russian group, but they both were less accurate than the two high proficiency groups and the control group. Crucially, the high proficiency L1 English group and the high proficiency L1 Russian group were as accurate as the control group. Agreement Type (D/N vs. N/A agreement) did not affect

accuracy, while Congruency affected the low proficiency L1 Russian group's accuracy in the GJT and it also affected both high and low proficiency L1 Russian groups' accuracy on the PMT.

Dissimilar results were obtained for RTs for AoE and Proficiency, although the results for the linguistic variables Type of Agreement and Congruency were largely the same. More specifically, early AoE did not lead to shorter RTs, as one might have expected; in fact, late AoE group (L1 English) reacted faster than the early AoE group (L1 Russian). Grammaticality affected all groups, both for Age of Exposure and Proficiency. Type of Agreement (D/N vs. N/A) affected neither the AoE groups nor the proficiency groups. Interestingly, while Congruency affected low proficiency L1 Russian group's accuracy, it affected high proficiency L1 Russian group's RTs in the GJT.

On the PMT, however, both low and high proficiency L1 Russian groups were affected by Congruency for RTs, as for Accuracy. Crucially, unlike for the results for Accuracy, both L1 English groups reacted significantly faster than their respective L1 Russian proficiency groups. That is, the low proficiency L1 English group was faster than the low proficiency L1 Russian group, and the high proficiency L1 English group was faster than the high proficiency L1 Russian group. Moreover, the high proficiency L1 English group was as fast as the control group, while the high proficiency L1 Russian group was slower than both of them.

Results from the Experiment 2 showed that Gender Type (lexical-semantic vs. lexical-syntactic) did not affect accuracy or results, but that Cue Type affected accuracy for all groups, but not RTs. With respect to proficiency, high proficiency learners were as accurate as the controls, and both of these groups were more accurate than the low

proficiency groups. A discussion and interpretation of these findings will be provided in the next chapter.

CHAPTER 5: Discussion

Two experiments were conducted to address the research questions raised in this dissertation. The findings show that later age of exposure to the syntactic gender features (in the L2) does not lead to compromised surface gender agreement: L1 English-L2 Spanish speakers were as accurate as the L1 Russian L2 Spanish speakers. Moreover, based on the combined results of the two experiments of this dissertation, high proficiency L1 English learners of Spanish behaved target-like on virtually all tasks of the two experiments not only in terms of accuracy, but also in terms of reaction times.

This finding has three implications. The first one is that phonological transfer may not permanently impair the acquisition of morphological features, at least for comprehension. If the opposite were the case, L1 English learners of Spanish would not be able to perform target-like on surface gender agreement. Admittedly, this is an arguable question, because reading tasks and exposure to the orthographic form may lead to different results than listening tasks and exposure to the acoustic form of the word. Nevertheless, as Levelt et al. (1999) suggest, and as it was mentioned in the Literature Review chapter, both the orthographic and acoustic form of the word will lead to the phonological decoding, which will subsequently activate the syntactic and semantic features of this word. The second implication is that L2 learners are capable of building the correct lexical representation of the noun, because high proficiency L2 learners were target-like both in terms of accuracy and RTs, which shows that they had linked L2 lemmas to the correct gender nodes and were able to access the nodes and use the lexical information for the purposes of surface gender agreement.

The third and the most crucial implication is that the syntactic component of gender – the abstract syntactic gender features – is acquirable after the critical period. This finding lends support to the Full Transfer Full Access Hypothesis (Schwartz & Sprouse, 1996) and Feature Reassembly Hypothesis (Lardiere, 2008), according to which features that were not instantiated in the L1 can be acquired/reassembled in the L2. On the other hand, this finding is not compatible with hypotheses such as the Failed Formal Features Hypothesis (Hawkins & Chan, 1997) or the Representational Deficit Hypothesis (Hawkins, 2009), which claim that L2 learners cannot reach target-like gender concord due to a permanent syntactic impairment, and which found support in such studies as Hawkins (1998) and Franceschina (2000). I will discuss these studies in the following paragraphs and will report on methodological and theoretical differences that could account for the different findings in those studies and the present dissertation.

Hawkins (1998) collected production data (three minute description of an animated film) from twenty L1 English highly proficient L2 French speakers and analyzed their performance on gender agreement. Based on the results, he made a conclusion that gender representation on the part of English native speakers is not target-like even at advanced stages of L2 acquisition. Two issues can be brought up with respect to such a statement.

First of all, it may be methodologically not appropriate to test syntactic representations through a production task, because such tasks are prone to confounding factors such as affective filter and processing constraints (Prévost & White, 2000; Alarcón, 2011). For example, L2 learners in Alarcón (2011) performed target-like on a comprehension task, but not on a production task. The author argues that this finding is

compatible with the Missing Surface Inflection Hypothesis (Prévost & White, 2000), which predicts gender errors in production. She suggests that there is “a divergence between their acquisition of abstract features, as indicated by their high accuracy scores on the comprehension task, and their continuing difficulties with surface manifestations, as revealed by their significantly lower scores on the production task” (p.344).

Second, the study design in Hawkins (1998) did not include a group of participants with gender in their L1 or a control group. If one were to determine whether it is only L1-instantiated features that are acquired in L2, then it would be relevant and informative to provide data both from gendered and ungendered L1 learners of L2.

Such a methodological design (comprehension tasks instead of production tasks, comparison of gendered L1 with an ungendered L1 as well as with a group of native speakers) was employed in this dissertation, and this appears to be one of the reasons why the findings from Hawkins (1998) are different from the findings in this dissertation. The crucial distinction between Hawkins (1998), on the one hand, and the studies in this dissertation, on the other hand, is based on the theoretical distinction between syntactic features and their morphological spell-out. It has been claimed in this dissertation that it is theoretically imprecise to ascribe *morphological* variability to impairment at the level of *syntactic* features.

Another study that explicitly ascribes errors in surface gender agreement to a syntactic impairment is Franceschina (2001). The author compared informal conversation recordings from two highly proficient L1 Italian L2 Spanish with those from two highly proficient L1 English L2 Spanish speakers. Before I discuss her results, I need to note here that the two Italian participants had lived 48-50 years in the Spanish-speaking

country whereas one of the L1 English participants had lived there for 24 years and the other one only for 7 years. The author argues that this participant still had been considerably exposed to the L2, as she is married to an Argentinean and uses Spanish at work. She does not mention, however, whether the Italian participants were married to native speakers of Spanish. All in all, 48 years of residence are several times longer than 7 years, and comparing L2 speakers with such different linguistic backgrounds may be methodologically not appropriate, especially given that gender acquisition is partly a lexical process, which requires constant reinforcement through exposure and practice.

Another potential methodological limitation, as it has been suggested for Hawkins (1998), is the use of a production task to tap into abstract syntactic representations.

In fact, the major theoretical pursuit in Franceschina (2001) was to challenge Lardiere's (2000) idea that it is a competence at the level of morphology, not at the level of syntax that causes morphological errors. The author argued that if Lardiere's proposal were correct, Italian native speakers should be target-like on gender, because it is morphologically realized identically in Italian and Spanish, but they should have problems marking plural nouns, because number is marked differently in Spanish and Italian. The results showed that both Italian and English L1 were target-like as far as number was concerned, but that English L1 speakers were less accurate on gender marking than Italian L1 and the controls.

The author considers these findings to be counterevidence to the morphological explanation and suggests that they support the idea that it is the presence or absence of syntactic features in one's L1 that plays a crucial role in feature acquisition. However, it

is not completely clear how the finding that Italian L1 speakers perform like controls on Spanish number disprove Lardiere's hypothesis about mapping morphological forms to syntactic features (2000). Rather than doing so, it appears to show that even in case a feature present in both L1 and L2 is morphologically manifested differently (number in L1 Italian L2 Spanish), it can be acquired in the L2. Moreover, Lardiere puts forward her hypothesis to account for morphological errors, and the L1 Italian speakers behaved target-like on number, and so there are no errors to account for.

Furthermore, Franceschina (2001) argues the L1 English learners make persistent gender errors such as *una problema* 'a.F problem-M', while L1 Italian do not exhibit such errors. She interprets this observation as evidence for an inability to acquire syntactic features in the L2. However, this error can alternatively be interpreted as a gender assignment error. This has been the case in Grüter et al. (2012) and Hopp (2012), where the vast majority (e.g., 10 times as many gender assignment errors as gender agreement errors in Grüter et al., 2012) of gender errors were gender assignment, not gender agreement errors. It has also been shown to be the case in the Experiment 2 of this dissertation, where L1 English native speakers performed significantly less accurately on the nouns whose gender morphology does not follow the general *-a* for feminine/ *-o* for masculine pattern. The Spanish word *problema* 'problem' is a noun that contradicts this pattern, and hence its gender assignment is very challenging. Furthermore, it is obvious why the Italian participants did not make such an error – *problema* is a cognate with a congruent gender value in Italian and Spanish. The error rate in Franceschina (2001) was 8% for the L1 English participants, but again, it is not clear whether these errors were assignment or agreement errors.

This leads to the second major finding of this dissertation - surface gender agreement is in fact affected by the morphological complexity of the L2 gender system (cue unreliability). While cue reliability had a facilitative effect on all groups, this effect was very robust for the low proficiency L1 English group. This group scored 15.39 on average for reliable cue items out of the maximum score of 17 and 11.17 for the unreliable cue items out of the same maximum score.

Two observations have to be pointed out here. First, low proficiency L1 English learners of Spanish were more target-like on surface gender agreement for nouns with transparent gender morphemes (cue reliability). In other words, they performed significantly more accurately on surface gender agreement when the hindering effect of cue unreliability was eliminated. This indicates that even low proficiency learners have *syntactic* competence, and that it allows them to start acquiring *morphological* competence. Second, some low proficiency learners did not merely accept grammatical DPs such as *la casa* ‘the.F house.F’, but they also systematically rejected DPs such as *la luz* ‘the.F light.F’ and *la miel* ‘the.F honey.F’, although these are grammatical in the target language (Spanish). This indicates that the consonant endings such as *-z* and *-l* mark masculine gender in their grammars – **el luz* ‘the.M light.M’ and **el miel* ‘the.M honey.M’. This is a clear case of a gender assignment error, similar to those that L2 learners in Grüter et al. (2012) made. Both of these two observations lend support to Missing Surface Inflection Hypothesis (Haznedar and Schwartz, 1997; Prévost and White, 1999) as well as (some version of) the Separation Hypothesis (Beard, 1988, 1995; Lardiere, 2000).

I would like to emphasize here that while the line of reasoning provided in Rohrbacher (1994), Vainikka and Young-Sholten (1994, 1996), and Eubank (1993/1994), is partially consistent with the findings of this dissertation – the presence of syntactic features helps acquire morphological competence (production and comprehension of overt morphological forms), this line of reasoning is not *completely* correct, because the presence of syntactic features may not guarantee morphological competence.

Lardiere (2000) brings up a relevant example from Beard (1995). Russian case features are such that nominative case is assigned to the subject of a finite clause, dative case is assigned to the subject of an infinitive, and genitive case is assigned to the subject of an NP. However, merely mapping genitive case to the subject of an NP does not predict the correct surface production of the morphological forms, because the spell-out of the forms varies depending on noun class and number/stem phonology. Lardiere (2000, p. 124) thus concludes that, “I suspect that it is among this increasingly complex “outer”-layer mappings from morphology to PF that we are likely to find the greatest vulnerability to “fossilization” and “critical period” effects.” The results from Experiment 1 of this dissertation support this hypothesis.

With respect to the acquisition of the lexical-syntactic gender, Lemhofer et al. (2008) concluded that if the L1 does not have gender, there is no gender system at all, and if it appears, it is very unstable. Such a pessimistic prognosis for L2 gender learning may not hold true, because high proficiency L1 English L2 Spanish learners in Experiment 2 of this dissertation were capable of behaving target-like on morphologically unreliable items and low proficiency learners seem to have linked L2 lemmas to distinct gender classes. These classes may not be target-like, but they abide by

the laws of the learners' developing grammars. The only distinction between the lexical representation of native speakers and second language learners is that the former have a stable long-practiced link between the lexical-phonological/lexical-morphological form of the word and the lemma of the noun, while the latter are in the process of establishing these links based on the morphological characteristics of the specific gender systems.

Therefore, the ability to map syntactic features to their morphological forms, then classify nouns into categories and subsequently use this newly acquired knowledge for surface gender agreement is present in L2 learners, even if they do not have gender in L1.

In fact, having gender in the L1 does not facilitate the acquisition of L2 gender, at least for Russian L1 Spanish L2 learners, and this is another major finding of this dissertation. The results from Experiment 1 showed that the early AoE (Age of Exposure) group (L1 Russian speakers) did not outperform the late AoE group (L1 English). Moreover, whereas high proficiency English speakers were as fast as the controls in their reaction times, the high proficiency L1 Russian group reacted significantly slower than the control group and the high proficiency English L1 group. Thus, one should not expect that if the L1 has gender, L2 gender will be acquired more easily than if the L1 does not have gender. This is so because although the abstract gender features transfer from L1 to L2 and thus are available in the L2 from the onset of the L2 acquisition, the L2 lexical-morphological instantiation of these features has to be learnt for the surface gender agreement to be correct, and it seems to be more challenging to relearn a new lexical-syntactic feature, as is the case for L1 Russian L2 Spanish speakers, than to learn a completely new one, as is the case for L1 English L2 Spanish speakers.

This finding is in line with that obtained in Tokowicz and MacWhinney (2005), where it was determined that structures absent in the L1 are easier to process in the L2 (gender for L1 English L2 Spanish) than structures that are present in the L1 but are manifested differently (number for L1 English L2 Spanish). Similarly, Sabourin and Stowe (2008) found out that L1 German L2 Dutch showed target-like cortical reactivity on gender agreement tasks whereas L1 Romance L2 Dutch did not, despite the fact that both German and Romance have gender. This result is accounted for by the fact that Dutch and German are very similar in terms of gender assignment, whereas Romance and German (like Russian and Spanish) are not, although the authors of the study did not specifically test participants' behavior on L1/L2 congruent vs. incongruent gender items.

In Experiment 1 of this dissertation, the gender agreement tasks were specifically controlled with respect to Russian/Spanish congruency of the lexical gender values. Such a manipulation of research items led to another major finding of this dissertation - surface gender agreement is affected by L1/L2 congruency. The results from Experiment 1 showed that the early age of acquisition group performed surface gender agreement more accurately and faster on L1/L2 congruent gender items than on incongruent gender items. As for proficiency groups, while high proficiency L1 Russian group's accuracy was not affected by (in)congruency, low proficiency L1 Russian group's accuracy was.

Remarkably, although the high proficiency group was not affected by congruency in terms of accuracy, it was affected by it in terms of RTs, and although the low proficiency group was affected by congruency in terms of accuracy, it was not affected by it in terms of RTs. This result seems to reveal that the low proficiency learners have not incorporated L2 gender values into the lemmas, and that they use the available L1

gender values for surface gender agreement in L2, as is predicted by Jiang's (2000) psycholinguistic model of lexical acquisition. As for the high proficiency L1 Russian participants, they are in the process of linking the L2 nouns to the L2 gender nodes in the syntactic lexicon, and they retrieve this newly acquired lexical-syntactic gender information for the purposes of surface gender agreement.

This finding is also compatible with Revised Hierarchical Model (Kroll and Stewart, 1994), according to which low proficiency L2 speakers retrieve lexical items through a translation strategy, and higher proficiency L2 speakers directly access the concept without having to resort to the L1 lexeme. More specifically, when low proficiency L2 speakers are exposed to an L2 lexeme, they do not access the lexical-semantic meaning of the word directly but rather retrieve the L1 lexeme through which they subsequently access the lexical-semantic meaning. This is referred to as the 'lexeme route.' High proficiency L2 speakers, on the other hand, activate the lexical-semantic representation directly from the L2 lexeme. This is called the 'concept mediation route.' The lexeme route is cognitively less costly and thus faster than the concept mediation route.

Although the Revised Hierarchical Model does not make predictions specifically about the lexical-syntactic representation, Levelt et al. (1999) predict that the L1 lexeme, once activated, sends activation to the L1 lemma before the lexical-semantic representation is activated. Table 17 captures the stages of lexical access in L1 Russian L2 Spanish learners and Table 18 captures these stages in L1 English L2 Spanish learners. Thus, both the fact that the low proficiency learners were more accurate but not faster on congruent items and the fact that the high-proficiency L1 Russian speakers were

equally accurate on congruent and incongruent items but were slower in general are compatible with Revised Hierarchical Model, with the addition of the lemma stage adopted from Levelt et al. (1999). The low proficiency Russian speakers use a transfer strategy that leads them to quickly activate the inappropriate L1 lemmas, whereas high proficiency Russian speakers directly access the appropriate L2 lemma, but this direct route to the L2 lemma leads them to longer reaction times.

Table 17. *Stages of lexical access in low proficiency L1 Russian L2 Spanish learners*



L2 lexeme	L1 lexeme	L1 lemma	Lexical-semantic representation
<i>Casa</i> ('house' in Spanish)	<i>Dom</i> ('house' in Russian)	MASCULINE	

Table 18. *Stages of lexical access in high proficiency L1 Russian L2 Spanish learners*

L2 lexeme	L2 lemma	Lexical-semantic representation
<i>Casa</i> ('house' in Spanish)	FEMININE	

The reason why the links between the L2 lexeme and the concept are weaker than the links between the L1 lexeme and the concept is that the L1 links have been long established and practiced throughout a lifetime. Findings from Grüter (2012) emphasize the importance of establishing strong associations between the lexeme and the lemma. In this study, L2 learners were not able to use gender on determiners as cues to predict the gender of familiar nouns as effectively as L1 learners, but they indeed were able to do so for nonce words. Although the links between the lexemes of familiar nouns and their

lemmas should be present in L2 learners, they are undoubtedly more stable in native speakers. In other words, the more often a speaker activates the lemma, the more stable the link between the lemma and the lexeme becomes. Native speakers access the lemma much more often than (even proficient) L2 speakers, which is why access to the lemma becomes reinforced and highly automatic. This can perfectly account for the fact that L2 speakers were able to use gender as a cue for nonce words as effectively as L1 speakers, because native speakers were not disadvantaged by the less automatic lexical access.

The Revised Hierarchical Model may account for the slower RTs for high proficiency group but not for the low proficiency group in the sense that the concept mediation route takes longer than the lexeme route. However, it cannot account for the fact that L1 Russian speakers were in general slower than L1 English native speakers. Such a group effect may indicate interference from the L1 gender values for the L1 Russian groups. That is, high proficiency L1 Russian participants need more time to inhibit the L1 lemma content (Green, 1998) to retrieve the L2 gender value and use it for agreement purposes.

According to Levelt et al. (1999), lexical features compete for selection. Any lexeme (e.g., *casa* ‘house’) will activate a number of other lexemes with similar phonological shape (e.g., *cama* ‘bed’, *queso* ‘cheese’) and all these lexemes will subsequently activate their relevant lemmas, which will finally lead to activation of the lexical-semantic network. Then, each L2 lexeme will activate L2 lemmas, if these are present, and will also activate the L1 lexeme, as suggested by the Revised Hierarchical Model. The L1 lexeme will subsequently activate L1 lemmas with all of the features that it hosts. In order for the correct lemma and the correct lexical-semantic representation to

be selected, all of the inappropriate lemmas and lexical-semantic representations must be inhibited. Since Russian lemmas have gender features and English lemmas do not, lexical selection for L1 Russian L2 Spanish learners requires more inhibition, and hence is more costly than lexical selection for L1 English L2 Spanish learners. This can account for the slower RTs in Russian L1 groups.

In fact, it is remarkable that L1 Russian L2 Spanish learners switch from using L1 gender values to using L2 gender values for the purposes of gender agreement so quickly given that they have to carry out two tasks simultaneously - they must learn to inhibit the L1 lemma and at the same time they must perform a morphological analysis of the L2 lexeme in order to acquire the L2 gender morphology, as it is reflected in Table 19. Table 20, on the hand, shows that an L1 English learner does not have to inhibit L1 lemma, because it is void of gender values, or, using Levelt et al. (1999) terminology, the L1 lemma does not automatically activate gender nodes in the lexicon.

Table 19. *The lexical-morphological tasks for an L1 Russian L2 Spanish learner*

Exposure to an L2 lexeme	
TASK SET 1:	TASK SET 2:
INHIBIT: L1 lexeme	PROCESS MORPHOLOGY
INHIBIT: L1 lemma	ASSIGN A GENDER VALUE
Concept	

Table 20. *The lexical-morphological tasks for an L1 English L2 Spanish learner*

Exposure to an L2 lexeme	
TASK SET 1:	TASK SET 2:
INHIBIT: L1 lexeme	PROCESS MORPHOLOGY
	ASSIGN A GENDER VALUE
Concept	

Such successful and in fact very rapid acquisition of the L2 morphology (some high proficiency L1 Russian participants had learned Spanish only for three years) *despite already having gender values in the L1* and the integration of the L2 gender values into the lexicon is only possible if the L2 learners start attending to the L2 gender morphemes right from the very beginning of the L2 acquisition, and the reason why they use the L1 gender values is that they have not yet abstracted the rules of the L2 morphological gender system.

What could be a motivation for attending to the morphemes in the input even when gender values already exist in the L1 lexicon? I suggest here that it is the abstract syntactic features that motivate such a search process both for genderless and gendered L1. It should be noted here that L1 Russian learners of Spanish do not resort to specific L1 morphemes for L2 gender agreement. Russian and Spanish are morphophonologically congruent on feminine gender value (both typically have *-a* ending), but are incongruent

on masculine gender value exponent (while *-o* ending typically marks masculine in Spanish, it marks neuter in Russian). Despite such a morphophonological congruency for feminine gender nouns, neither high nor low proficiency Russian L1 Spanish L2 learners in this dissertation were shown to have a preference for feminine gender; they perform equally across the two gender classes. This finding demonstrates that Russian L1 Spanish L2 do not rely on L1 gender morphological forms when establishing gender values in their L2. If that were the case, they would perform better on feminine items, because they have *-a* ending, which is almost unequivocally feminine in Russian, whereas *-o* ending is a neuter gender ending in Russian. Thus, L1 Russian L2 Spanish learners start acquiring L2 morphology right from the start, but because they still lack reliable knowledge at the beginning stages of L2 acquisition, and since the abstract syntactic features urge them to perform agreement, they use the available L1 lexical gender values.

Thus, the abstract syntactic gender agreement is active in low and high proficiency L1 Russian learners of Spanish. Low proficiency learners behave significantly more accurately on congruent than on incongruent items, which means that once a gender value is available, they are capable of performing surface gender agreement. Moreover, despite longer reaction times, high proficiency L1 Russian learners of Spanish perform as accurately as the control group both on congruent and incongruent items. This finding is remarkable because it allows teasing apart the syntactic and the lexical component of gender, and it clearly indicates that syntactic gender agreement is present and active in Russian L1 Spanish L2 speakers even at the early stages of L2

acquisition, and that it is the lexical component that the interlanguage grammar initially lacks.

The finding that Russian L1 learners of Spanish perform syntactic agreement in the L2 but resort to the L1 lexical values is in line with the Functional Convergence Hypothesis (Sanchez, 2003). According to this theory, a functional projection may have different sets of features in the L1 and L2. It may include a feature in the L1 (evidentiality in Quechua), but not include this feature in the L2 (Spanish). Conversely, it may include a feature in the L2 (aspect in Spanish), but not in the L1 (Quechua). Bilinguals have been shown to use L2 morphemes that mark an L2 feature (aspect) to “express” an L1 feature (evidentiality) that is absent in the L2 (Sanchez, 2003). In this case, L2 morphemes that mark a specific value for an L2 feature are mapped on to a non-congruent L1 feature value. This could be referred to as “feature convergence”, namely, the mapping of L1 feature values onto L2 morphemes .

Another research finding made in this dissertation concerns the idea of modularity of linguistic components such as syntax and lexical semantics – while they interact, they are not equivalent. I base this claim on the fact that, although the semantic lexicon is shared between the two languages of a bilingual individual, and thus lexical-semantic gender (e.g. *woman* as conceptually being feminine) are available in the L1 and in the L2, these values do not trigger gender agreement, at least on comprehension tasks such as the ones used in this dissertation. Neither of the two experimental groups (low proficiency and high proficiency L1 English groups) performed more accurately or reacted faster on nouns with lexical-semantic gender compared to the nouns with lexical-syntactic gender.

While this finding could be caused by a ceiling effect for the high proficiency group, such an explanation could not be used for the low proficiency group.

This finding may serve as evidence in favor of one of the assumptions of the syntactic theory adopted here – only lexical-syntactic features, but not the lexical-semantic features, are available for the computational mechanism. Hence, L2 learners whose L1 does not have gender will need to incorporate gender values in the L2 lemmas (lexical-syntactic gender) regardless of whether the noun has a lexical-semantic gender value.

The last but not least finding of this dissertation is that gender agreement is not more target-like between more frequent DP-types (definite article followed by a noun) than on less frequent DP-types (noun followed by an adjective). Frequency of input as defined in Mariscal (2008) does not affect the acquisition of surface gender agreement, regardless of the L1 background. None of the four experimental groups (low proficiency L1 English and Russian, high proficiency L1 English and Russian groups) or the control group showed more accurate surface gender agreement on definite article/noun sequences, which are considered to be more frequent in the input (Mariscal, 2008), than on noun/adjective sequences, which are considered to be less frequent in the input. That is, L2 learners distribute lexical-syntactic gender values among both determiners and adjectives.

This finding seems to challenge the idea that input, as defined in Mariscal (2008), is the most crucial factor for language acquisition, as is maintained by the proponents of Connectionism (Ellis, 1998; Mariscal, 2008). If it were true, and input indeed was the only factor responsible for language acquisition, agreement between the definite article

and the noun should in fact be more target-like than the agreement between the noun and the adjective. Again, this was not the case in this dissertation.

This finding also seems to support the idea that L2 learners represent the DP as a syntactic unit all of the elements of which share features and feature values. Thus, while acquisition of gender is partially a word-learning process, as suggested in Unsworth (2008), it fundamentally depends on syntactic representations, because, as it was suggested above, the lexical features of nouns are acquired through access to abstract syntactic features that motivate acquisition of morphological competence and the subsequent categorization of nouns into lexical gender values.

As Audring (2008) points out, despite the common observation that the gender of nouns (the *lexical-syntactic gender*, as per the terminology adopted in this dissertation) determines gender of determiners and modifiers, the link can in fact be in the opposite direction, so that agreement determines the assignment. Based on an elaborate linguistic analysis of multiple languages all over the world, Audring (2008) concludes that it is the presence of agreement in a language that makes gender assignment available. In languages that gradually lose the formal exponent of agreement, assignment classes also get lost. Therefore, because there is abstract syntactic gender agreement, L2 learners attend to the morphemes on the elements of the DP and subsequently build a lexical-syntactic representation of the noun. That is, the syntactic component is primary, and the lexical component is secondary in gender agreement, although it may appear to be the opposite.

In conclusion, six main findings were made in this dissertation. First, L2 learners whose L1 does not have gender (L1 English) are not disadvantaged compared to the L2

learners whose L1 has gender (Russian). Abstract syntactic features may indeed be acquirable in the L2, and the studies that claim the opposite were conducted on surface gender agreement, and thus cannot be used as evidence against the availability of abstract syntactic features in the L2.⁵

Second, the decreased accuracy on surface gender agreement in those studies (and in general) may in fact be indicative of a lack of lexical and morphological knowledge, because such knowledge needs to be learnt (for an L1 that lacks gender) or relearnt (for an L1 that has gender) and practiced before it eventually becomes automatized.

Third, L2 learners whose L1 lacks gender (L1 English) are capable of behaving target-like on gender agreement. Thus, the findings from the two experiments of this dissertation supported the results obtained in the studies proposing acquirability of gender after the critical period.

Fourth, having gender in L1 does not necessarily facilitate the acquisition of gender in L2. While it is true that Russian L1 learners of Spanish activated the abstract syntactic gender features before the critical period (in their L1), they had to relearn the specific lexical gender values and the concrete morphological gender markers of Spanish nouns in their L2, and the relearning process is fraught with the consequences of lexical interference from L1. In fact, the results of Experiment 1 indicate that relearning lexical knowledge is more challenging than acquiring entirely new lexical knowledge.

Fifth, lexical-semantic gender values are not appropriate for the syntactic operation gender agreement; hence, lexical-syntactic values have to be acquired for both

⁵ Certainly, it has to be admitted that tapping into abstract linguistic representations is challenging, and very careful methodological designs have to be developed to address the question. Moreover, while such carefully thought-out methodologies could potentially reveal that abstract features are in fact *available* after the critical period, no behavioral methodology should be able to allow one to conclude that they are *not acquirable*.

nouns with and without lexical-semantic gender, which supports the notion of modularity of syntax.

Last but not least, the claim that agreement between definite articles and nouns is more easily acquired than agreement between nouns and adjectives because the former are more frequent in the input than the latter was not supported in this dissertation. Thus, the higher frequency of occurrence of D/N sequences in the input compared to the lower frequency of the N/A sequences is not a crucial factor for gender acquisition in the L2.

Future research on gender agreement could proceed in different directions. A question of the age of acquisition could be investigated from a different perspective – a group of child L1 learners of a gendered language could be compared with a group of adult L2 of this language. However, as it has been emphasized, if the results indicate that gender agreement is compromised in the adults, this should not be considered as evidence against acquirability of *abstract syntactic features*, unless a unique methodology is developed that taps specifically into the abstract syntactic representations and eliminates all confounding factors.

An interesting direction to take with respect to gender research would be to explore the effect of L1 phonological transfer on the acquisition of gender agreement. For example, acquisition of gender morphemes in Spanish may be contingent upon phonological contrast between *-a* and *-o*. Spanish is a language that does not exhibit vowel reduction, and the *-a* and *-o* vowels could represent a cue for acquiring gender. Since English has vowel reduction, L1 English learners of Spanish may find it more challenging to acquire the *-a/-o* contrast and hence the distinction between masculine

and feminine gender morphemes compared to L1 learners whose L1 is similar to Spanish in terms of vowel reduction (e.g. Finnish).

Exploring the issue of implicit and explicit knowledge and the role of instruction for the two types of knowledge could also provide new insights into the topic of gender acquisition.

A word should be said about potential limitations of the study. First of all, the high proficiency group in both experiments consisted of 16 participants, and such a sample may not be fully representative of the target population. Second, the control group was not represented by monolingual Spanish speakers, but rather by bilinguals whose first language was Spanish, and who also were highly proficient in English. They spoke English on a daily basis at work, which some may argue could have affected their *processing* of gender feature. However, there is no reason to believe that their gender *representation* was compromised by their speaking a second language given their results with respect to accuracy.

Conclusion

It was reported in the Introduction chapter of this dissertation that there are two main questions on the research agenda of the SLA field. The first question is the broader SLA question and concerns the issue of whether language is acquired through access to an innate linguistic knowledge triggered by the input or simply through exposure to the language. The second question is a question within the generative SLA field and concerns

the issue of UG availability in the L2, after the critical period postulated by these theories has ended.

With respect to the first question, the findings of this dissertation are compatible with the idea of modularity of syntax and the notion of preexisting syntactic knowledge provided by the UG. This is so because the L2 learners tested in this dissertation were able to attend to semantically meaningless gender morphemes while disregarding lexical-semantic cues (sex), to categorize nouns in gender classes based on these morphemes relatively early in the L2 acquisition process, and to perform gender agreement like native speakers regardless of whether it was the more frequent agreement type (between the definite article and the noun) or the less frequent agreement type (between the noun and the adjective).

With respect to the second question, the finding that the L2 learners in this dissertation were able to exhibit target-like surface gender agreement challenges the syntactic impairment hypotheses such as Failed Formal Features Hypothesis (Hawkins & Chan, 1997) and more recent representational deficit hypotheses (Hawkins 2000, 2003, 2009; Hawkins & Liszka 2003; Tsimpli 2003), and lends support to the no deficit hypotheses such as Full Access Full Transfer Hypothesis (Schwartz & Sprouse, 1996), Functional Convergence (Sánchez, 2003) and Feature Reassembly (Lardiere, 2008). More specifically, this finding indicates that L2 learners are capable of acquiring three types of linguistic competence – lexical competence, morphological competence and syntactic competence, which further supports the idea that the abstract syntactic features provided by the UG are present and active in their grammars.

Moreover, the finding that the low proficiency L2 learners were clearly affected by the complexity of Spanish gender morphology supports the Separation Hypothesis (Beard, 1995; Lardiere, 2000) and the Missing Surface Inflection Hypothesis (Haznedar & Schwartz, 1997; Prévost & White, 1999).

Furthermore, the congruency effect observed in low proficiency L1 Russian L2 Spanish speakers is in line with the Revised Hierarchical Model (Kroll and Stewart, 1994) and the L2 lemma blocking idea suggested by Jiang (2000). The fact that L1 Russian participants in general were slower than the L1 English participants is compatible with the Inhibitory Control model (Green, 1998), according to which potential competitors (such as the irrelevant L1 lexical-syntactic gender features) that are activated in the course of lexical retrieval have to be inhibited, which may cause longer reaction times.

Finally, the findings overall support the WEAVER ++ model (Levelt et al., 1999) by showing that the lexicon is not a mere storage of phonological forms of the word but rather of complex sets of lexical-semantic, lexical-syntactic, and formal (lexical-phonological and lexical-morphological) features. Crucially, as it is suggested by the generative linguistic theory (Hauser et al., 2003), only lexical-syntactic and formal (lexical-phonological and lexical-morphological) features, but not lexical-semantic features, affect the syntactic operation gender agreement.

All in all, the findings of this dissertation show that gender is acquirable in the L2. In fact, it is surprising that despite a number of potential hindering factors – lack of lexical/morphological knowledge or interference from L1 lexical knowledge, differences in L1/L2 phonology, possible shallow processing of morphological markers - L2 learners

of Spanish still perform target-like on surface gender agreement. Although one has to admit that behavioral data may not be sensitive enough to reveal slight differences between the control group and the experimental groups, these results are still relevant, and they add empirical data to the existing research on the L2 acquisition of lexical and syntactic features in general and on the acquisition of gender agreement in particular. They also highlight the idea that bilingual speakers are an optimal population for exploring interfaces between the different linguistic modules (Sanchez, 2015).

References

- Aguirre, C. (1995). *La adquisición de las categorías funcionales en el español*. Doctoral dissertation, Universidad Autónoma de Madrid.
- Alarcón, I. (2011). Spanish gender agreement under complete and incomplete acquisition: Early and late bilinguals' linguistic behavior within the noun phrase. *Bilingualism: Language and Cognition*, 14, 332-50.
- Andonova, E., D'Amico, S., Devescovi, A., & Bates, E. (2004). Gender and lexical access in Bulgarian. *Perception & Psychophysics*, 66, 496-507.
- Audring, J. (2008). Gender assignment and gender agreement: Evidence from pronominal gender languages. *Morphology* 18, 93-116.
- Austin, J., Blume, M., & Sánchez, L. (2015). *Bilingualism in the Spanish-speaking world: Linguistic and cognitive perspective*. Cambridge University Press.
- Badecker, W., Miozzo, M., & Zanuttini, R. (1995). The two stage model of lexical retrieval: Evidence from a case of anomia with selective preservation of grammatical gender. *Cognition*, 57, 193-216.
- Baker, M. (2013). *The Cambridge handbook of generative syntax*. M.den. Dikken (Ed.). New York, NY: Cambridge University Press.
- Bates, E., Devescovi, A., Pizzamiglio, L., D'Amico, S., & Hernandez, A. (1995). Gender and lexical access in Italian. *Perception & Psychophysics*, 57, 847-862.
- Beard, R. (1988). On the separation of derivation from morphology: Toward a lexeme/morpheme-based morphology. *Quaderni di Semantica*, 9, 3-59.
- Beard, R. (1995) *Lexeme-morpheme base morphology*. Albany, NY: SUNY Press.
- Bernstein, J. (1993). *Topics in the syntax of nominal structure across Romance*. (Unpublished doctoral dissertation). City University of New York, Graduate Center, New York.
- Bernstein, J. (2001). The DP hypothesis: Identifying clausal properties in the nominal domain. In M. Baltin & C. Collins (Eds.), *The Handbook of Contemporary Syntactic Theory* (pp.536-561). Oxford: Blackwell.
- Bialystok, E. (1979). Explicit and implicit judgments of L2 grammaticality. *Language Learning*, 29, 81-103.
- Biran, M., & Friedmann, N. (2012). The representation of lexical-syntactic information: Evidence from syntactic and lexical retrieval impairments in aphasia. *Cortex*, 48(9), 1103-1127. doi:10.1016/j.cortex.2011.05.024

- Blom, E., & Unsworth, S. (2010). *Experimental Methods in Language Acquisition Research*. Amsterdam: John Benjamins.
- Bock, K., & Levelt, W. (1994). Language production: Grammatical encoding. In M.A. Gernsbacher (Ed.), *Handbook of psycholinguistics*. San Diego, CA: Academic Press.
- Bordag D., Pechmann T. (2007). Factors influencing L2 gender processing. *Biling. (Camb. Engl.)*, 10, 299-314
- Bordag, D. (2004). Interaction of L1 and L2 systems at the level of grammatical encoding: Evidence from picture naming. In S. Foster-Cohen (Ed.), *EUROSLA Yearbook 4*, (pp.203-230). Amsterdam: John Benjamins.
- Bordag, D., Opitz, A., & Pechmann, T. (2006). Gender processing in first and second languages: The role of noun termination. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(5), 1090-1101.
- Boškovic, Z. (2008). What will you have, DP or NP? Proceedings from the NELS: 37th Annual Meeting of the North East Linguistic Society: 101-114. University of Massachusetts, Amherst, MA: GLSA (Graduate Linguistic Student Association).
- Bottari, P., Cipriani, P., & Chilosi, A. M. (1993). Protosyntactic Devices in the Acquisition of Italian Free Morphology. *Language Acquisition*, 3(4), 327-369.
- Bowles, M. (2011). Measuring implicit and explicit linguistic knowledge. *Studies in Second Language Acquisition*, 33, 247-271. doi:10.1017/S0272263110000756
- Bruhn de Garavito, J., & White, L. (2000). L2 acquisition of Spanish DPs: The status of grammatical features. In A. T. Pérez-Leroux & J. M. Liceras (Eds.), *The acquisition of Spanish morphosyntax: The L1/L2 connection*. Dordrecht: Kluwer.
- Cacciari, C., Carreiras, M., & Barbolini-Cionini, C. (1997). Anaphora resolution in Italian. *Journal of Memory and Language*, 37, 517-532.
- Cameron-Faulkner, T., Lieven, E. & Tomasello, M. (2003). A construction based analysis of child directed speech. *Cognitive Science*, 27(6), 843-873.
- Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology*, 14, 177-208.
- Caramazza, A., & Miozzo, M. (1997). The relation between syntactic and phonological knowledge in lexical access: Evidence from the 'tip-of-the-tongue' phenomenon. *Cognition*, 64, 309-343.
- Carroll, S. (1989). Second language acquisition and the computational paradigm. *Language Learning*, 39, 535-594.

- Carstens, V. (2000). Concord in minimalist theory. *Linguistics Inquiry*, 31, 319-355.
- Carstens, V. (2010). Implications of grammatical gender for the theory of uninterpretable features. In M. Putnam, (Ed.), *Exploring Crash Proof Grammars* (pp.31-57). Amsterdam: John Benjamins.
- Ceytlin, S. N. (2005). Kategorija roda v detskoj rechi. In A. V. Bondarenko, & S. A. Shubik (Eds.), *Problemy funkcionalnoj grammatiki* (pp.346-375). St Petersburg, Russia: Rossijskaja Akademija Nauk.
- Ceytlin, S. N. (2009) *Očerki po slovoobrazovaniju i formoobrazovaniju v detskoj rechi*. Moscow, Russia: Znak.
- Chomsky, N. (1959). Review of Skinner's *Verbal Behavior*. *Language*, 35, 26-58.
- Chomsky, N. (1995). Categories and transformations. In *The minimalist program* (pp.219-394). Cambridge, Mass.: MIT Press.
- Chomsky, N. (1995). *The minimalist program*. Cambridge, MA: MIT Press.
- Chomsky, N. (2000). Minimalist inquiries: The framework. In R. Martin, D. Michaels, & J. Uriagereka (Eds.), *Step by Step: Essays on Minimalist Syntax in Honor of Howard Lasnik* (pp.89-155). Cambridge, MA: The MIT Press. DOI: 10.1017/s0022226702271625
- Chomsky, N. (2001). Derivation by Phase. In M. Kenstowicz (Ed.), *Ken Hale. A Life in Language* (pp.1-52). Cambridge, MA: MIT Press.
- Chomsky, Noam. 2013. Problems of projection. *Lingua*, 130, 33-49
- Cinque, G. (1994). On the evidence for partial N-movement in the Romance DP. In G. Cinque (Ed.), *Paths towards Universal Grammar* (pp.85-110). Washington DC: Essays in honor of Richard S. Kayne, Georgetown University Press.
- Cinque, G. (2010). *Syntax of adjectives: A comparative study*. MIT Press: Cambridge, MA
- Clark, E. V. (1986). Acquisition of Romance, with special reference to French. In D. I. Slobin (Hrsg.), *The Crosslinguistic Study of Language Acquisition*. Vol.I., The data (687-782). Hillsdale, NJ: Erlbaum.
- Colomé, A., & Miozzo, M. (2010). Which words are activated during bilingual word production? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36, 96-109.
- Corbett, G. (1991). *Gender*. Cambridge University Press: Cambridge, MA
- Costa, A., Caramazza, A., & Sebastián-Gallés, N. (2000). The cognate facilitation effect: Implications for models of lexical access. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 26(5), 1283-1296.

- Costa, A., Kovacic, D., Fedorenko, E., & Caramazza, A. (2003a). The gender congruency effect and the selection of freestanding and bound morphemes: Evidence from Croatian. *Journal of Experimental Psychology: Learning Memory and Cognition*, 29, 1270-1282.
- Costa, A., Kovacic, D., Franck, J., & Caramazza, A. (2003b). On the autonomy of the grammatical gender systems of the two languages of a bilingual. *Bilingualism: Language and Cognition*, 6, 181-200.
- De Groot, A., & Nas, G. (1991). Lexical representation of cognates and noncognates in compound bilinguals. *Journal of Memory and Language*, 30, 90-123.
- DeKeyser, R. M. (2003). Implicit and explicit learning. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp.312-348). Oxford, UK: Blackwell.
- DeKeyser, R. M. (2007). Skill acquisition theory. In B. VanPatten & J. Williams (Eds.), *Theories in Second Language Acquisition: An introduction* (pp.97-112). Mahwah, NJ: Erlbaum.
- DeKeyser, R. M. (2009). Cognitive-psychological processes in second language learning. In M. H. Long & C. J. Doughty (Eds.), *The handbook of language teaching* (pp.119-138). Oxford, UK: Wiley-Blackwell.
- Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, 93, 283-321.
- Dewaele, J-M., & Véronique, D. (2001). Gender assignment and gender agreement in advanced French interlanguage: A cross-sectional study. *Bilingualism: Language and Cognition*, 4(3), 275-297.
- Dijkstra, A., & Van Heuven, W.J.B. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, 23, 175-197.
- Dijkstra, T. (2005). Bilingual visual word recognition and lexical access. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp.179-201). Oxford: Oxford University Press.
- Dussias, P., Valdés Kroff, J., Guzzardo Tamargo, R., & Gerfen, C. (2013). When gender and looking go hand in hand. *Studies in Second Language Acquisition*, 35(02), 353-387.
- Ellis, N. (1998). Emergentism, connectionism and language learning. *Language Learning*, 48(4), 631-664.
- Ellis, N. C. (2005). At the interface: Dynamic interactions of explicit and implicit language knowledge. *Studies in Second Language Acquisition*, 27, 305-352.
doi:10.1017/S027226310505014X

- Ellis, R. (2004). The definition and measurement of explicit knowledge. *Language Learning*, 54, 227-275.
- Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in Second Language Acquisition*, 27, 141-172.
- Ellis, R., Loewen, S., Elder, C., Erlam, R., Philp, J., & Reinders, H. (2009). *Implicit and explicit knowledge in second language learning, testing and teaching*. Tonawanda, NY: Multilingual Matters.
- Erlam, R. (2006). Elicited imitation as a measure of L2 implicit knowledge: An empirical validation study. *Applied Linguistics*, 27, 464-491. doi:10.1093/applin/aml001
- Eubank, L. (1994). On the transfer of parametric values in L2 level. *Language Acquisition*, 3, 183-208.
- Eubank, L. (1996). Negation in early German-English Interlanguage: more Valueless Features in the L2 initial state. *Second Language Research*, 12, 73-106.
- Fitch, W. T., M. D. Hauser, & N. Chomsky. 2005. The evolution of the language faculty: clarifications & implications. *Cognition*, 97, 179-210.
- Franceschina, F. (2001). Against an L2 morphological deficit as an explanation for the differences between native and non-native grammars. In S. H. Foster-Cohen & A. Nizgorodcew (Eds.), *EUROSLA Yearbook: Vol. 1* (pp.143-158). Amsterdam: Benjamins.
- Franceschina, F. (2005). *Fossilized second language grammars: The acquisition of grammatical gender*. Amsterdam/Philadelphia: J. Benjamins.
- Franks, S. & A. Pereltsvaig. (2004). Functional categories in the Nominal Domain. In O. Arnaudova, W. Browne, M. L. Rivero & D. Stojanovic (Eds.) *Proceedings of FASL 12. The Ottawa Meeting 2003* (pp.109-128). Ann Arbor, MI: University of Michigan Press.
- Garrett, M. F. (1976). Syntactic processes in sentence production. In R. J. Wales & E. Walker (Eds.), *New approaches to language mechanisms*. Amsterdam: North Holland.
- Garrett, M. F. (1992). Disorders of lexical selection. *Cognition*, 42, 143-180.
- Gess, R., & Herschensohn, J. (2001). Shifting the DP parameter: A study of Anglophone French L2ers. In J. Camps & C. R. Wiltshire (Eds.), *Romance syntax, semantics and their L2 acquisition* (pp.105-119). Amsterdam: Benjamins.
- Goad, H., White, L. & Steele, J. (2003) Missing inflection in L2 acquisition: Defective syntax or L1-constrained prosodic representations? *Canadian Journal of Linguistics*, 48, 243-263.

- Godfroid, A., Loewen, S., Jung, S., Park, J.-H., Gass, S., & Ellis, R. (2015). Timed and untimed grammaticality judgments measure distinct types of knowledge: Evidence from eye-movement patterns. *Studies in Second Language Acquisition*, 37(2).
- Granfeldt, J. (2000). The acquisition of the Determiner Phrase in bilingual and second language French. *Bilingualism: Language and Cognition*, 3, 263-80.
- Green, D. (1998). Mental control of the bilingual lexico- semantic system. *Bilingualism: Language and Cognition*, 1, 67-81.
- Grüter, T., Lew-Williams, C., & Fernald, A. (2012). Grammatical gender in L2: A production or a realtime processing problem? *Second Language Research*, 28, 191-215
- Guillelmon, D., & Grosjean, F. (2001). The gender marking effect in spoken word recognition: The case of bilinguals. *Memory & Cognition*, 29(3), 503-511.
- Harris, J. (1991). The exponence of gender in Spanish. *Linguistic Inquiry*, 22, 27-62.
- Hauser, M., Chomsky, N., & Fitch, T. 2002. The faculty of language: what is it, who has it, and how did it evolve. *Science*, 198, 1569-79
- Hawkins, R. (1998). "Explaining the difficulty of gender attribution for speakers of English", paper presented at *EUROSLA* 8, Paris.
- Hawkins, R. (2000). Persistent selective fossilization in second language acquisition and the optimal design of the language faculty. *Essex Research Reports in Linguistics*, 34, 75-90.
- Hawkins, R. (2003) 'Representational deficit' theories of adult SLA: evidence, counterevidence and implications. Invited plenary talk presented at EuroSLA, Edinburgh, September 2003.
- Hawkins, R. (2009). Statistical learning and innate knowledge in the development of second language proficiency: Evidence from the acquisition of gender concord. In A.G. Benati (Ed.) *Issues in second language proficiency* (pp.63-78). London: Continuum International Publishing.
- Hawkins, R. & Liszka, S. (2003) Locating the source of defective past tense marking in advanced L2 English speakers. In van Hout, R. et al. (Eds.) *The lexicon-syntax interface in second language acquisition* (pp.21-44). Amsterdam: John Benjamins.
- Hawkins, R., & Chan, C. (1997). The partial availability of Universal Grammar in second language acquisition: The 'failed functional features hypothesis'. *Second Language Research*, 13, 187-226.
- Hawkins, R., & Franceschina, F. (2004). Explaining the acquisition and non-acquisition of determiner-noun gender concord in French and Spanish. In P. Prevost & J. Paradis (Eds.), *The acquisition of French in different contexts. Focus on functional categories* (pp.175-207). Amsterdam: John Benjamins.

- Haznedar, B., & Schwartz, B.D. (1997). Are there Optional Infinitives in child L2 acquisition? In E. Hughes, M. Hughes & A. Greenhill (Eds.), *Proceedings of the 21st Annual Boston University Conference on Language Development*, 21 (pp.257-68). Somerville, MA: Cascadilla Press.
- Heinen, C. & Kadow, H. (1990). Acquisition of French. In J. M. Meisel (Ed.), *Two first languages: Early grammatical development in bilingual children* (pp.48-71). Dordrecht: Foris Publications
- Herbert, R. & Best, W. (2010). The role of noun syntax in spoken word production: Evidence from aphasia. *Cortex*, 46, 329-342.
- Hermans, D., Bongaerts, T., de Bot, K., Schreuder, R. (1998). Producing words in a foreign language: can speakers prevent interference from their first language? *Bilingualism: Language and Cognition*, 1(3), 213-230
- Hernández Pina, F. (1984). *Teorías psicosociolingüísticas y su aplicación a la adquisición del Español como lengua materna*. Madrid: Siglo:XXI.
- Hopp, H. (2012). Grammatical gender in adult L2 acquisition: Relations between lexical and syntactic variability. *Second Language Research*, 29(1), 33-56. DOI: 10.1177/0267658312461803
- Jackendoff, R. (2002). *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford:OUP. DOI: 10.2298/jfi0561213i
- Jackendoff, R. (2013). Constructions in the parallel architecture. In T. Hoffmann & G. Trousdale (Eds.) *The Oxford Handbook of Construction Grammar*. New York: OUP.
- Jiang, N. (2000). Lexical representation and development in a second language. *Applied Linguistics*, 21, 47-77.
- Johnson, E. (2005). Grammatical gender and early word recognition in Dutch. In A. Brugos, M. Clark-Cotton & S. Ha. (Eds.). *Proceedings of the 29th Annual Boston University Conference on Language Development (BUCLD 29)* (pp.320-30). Somerville: Cascadilla Press.
- Karmiloff-Smith, A. (1979). *A Functional Approach to Child Language: A Study of Determiners and Reference*. Cambridge: Cambridge University Press.
- Kroll, J. F., and Tokowicz, N. (2005). Models of bilingual representation and processing: Looking back and to the future. In J. F. Kroll and A. M. B. De Groot (Eds.) *Handbook of bilingualism: Psycholinguistic approaches* (pp.531-553). New York: Oxford University Press.
- Kroll, J.F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal*

- of Memory and Language*, 33, 149-174. DOI: 10.1006/jmla.1994.1008
- Kulke, F., & Blanken, G. (2001). Phonological and syntactic influences on semantic misnamings in aphasia. *Aphasiology*, 15, 3-15.
- La Heij, W., Hooglander, A., Kerling, R., & van der Velden, E. (1996). Nonverbal context effects in forward and backward translation: Evidence for concept mediation. *Journal of Memory and Language*, 35, 648-665.
- Lantolf, P., & S.L. Thorne. (2007). Sociocultural theory and second language learning. In B. Van Patten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp.693-701). N.J: Lawrence Erlbaum Associates, Inc.
- Lardiere, D. (2000). Mapping features to forms in second language acquisition. In J. Archibald (Ed.), *Second language acquisition and linguistic theory*, 102-129. Oxford: Blackwell.
- Lardiere, D. (2008). Feature-assembly in second language acquisition. In J. Liceras, H. Zobl, & H. Goodluck (Eds.) *The role of formal features in second language acquisition* (pp.106-140). Mahwah, NJ: Lawrence Erlbaum Associates.
- Lee, M., & Williams, J. (2001). Lexical access in spoken word production by bilinguals: Evidence from the semantic competitor priming paradigm. *Bilingualism: Language and Cognition*, 4, 233-248.
- Lemhofer, K., Spalek, K., & Schriefers, H. (2008). Cross-language effects of grammatical gender in bilingual word recognition and production. *Journal of Memory and Language*, 59, 312-330.
- Levelt, W. J. M. (1989). *Speaking. From intention to articulation*. Cambridge, MA: The MIT Press.
- Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22, 1-75.
- Lew-Williams, C., & Fernald, A. (2007). Young children learning Spanish make rapid use of grammatical gender in spoken word recognition. *Psychological Science*, 18, 193-198.
- Lidz, J., & Gagliardi, A. (2015). How Nature Meets Nurture: Universal Grammar and Statistical Learning. *Annual Review of Linguistics*, 1333-1353.
- Lleó, C. (1997). Filler syllables, proto-articles and early prosodic constraints in Spanish and German. In A. Sorace, C. Heycock, & R. Shillcock (Eds.), *Proceedings of GALA '97 Conference on Language Acquisition* (pp.251-56). Edinburgh: The University of Edinburgh.
- Lleó, C. (2001). The interface of phonology and syntax. The emergence of the article in the early acquisition of Spanish and German. In J. Weissenborn, & B. Höhle (Eds.), *Approaches to*

- Bootstrapping. Phonological, Lexical, Syntactic and Neuropsychological Aspects of Early Language Acquisition* (pp.23-44). Amsterdam: John Benjamins.
- López-Ornat, S. (2003). Learning earliest grammar: Evidence of grammar variations in speech before 22 months. In S. Montrul & F. Ordóñez (Eds.), *Linguistic theory and language development in Hispanic language*, (pp.254-74). Sommerville, MA: Cascadia Press
- MacWhinney, B., Leinbach, J., Taraban, R., & McDonald, J. L. (1989). Language learning: Cues or rules? *Journal of Memory and Language*, 28, 255-277.
- Marinis, T. (2010). Using on-line processing methods in language acquisition research. In E. Blom & S. Unsworth (Eds.), *Experimental methods in language acquisition research* (pp.139-162). Philadelphia, PA: John Benjamins Publishing Company.
- Mariscal, S. (2008). Early acquisition of gender agreement in the Spanish noun phrase: starting small. *Journal of Child Language*, 36, 143-171.
- Mazeika, E. (1973). A Comparison of the Grammar of a Monolingual and a Bilingual (Spanish-English) Child. Paper presented at Biennial Meeting of the Society for Research in Child Development, Philadelphia, PA.
- Meisel, J. (1997). The acquisition of the syntax of negation in French and German: Contrasting first and second language development. *Second Language Research*, 13(3), 227-263.
- Montrul, S., & Potowski, K. (2007). Command of gender agreement in school-age Spanish-English bilingual children. *International Journal of Bilingualism*, 11(3), 301-328.
- Montrul, S., Foote, R., & Perpinán, S. (2008). Gender Agreement in Adult Second Language Learners and Spanish Heritage Speakers: The Effects of Age and Context of Acquisition. *Language Learning*, 58(3), 503-553.
- Müller, N. (1994). Gender and number agreement within DP. In J. M. Meisel (Ed.), *Two first languages: Early grammatical development in bilingual children* (pp.53-88). Dordrecht: Foris.
- Oliphant, K. (1998). Acquisition of grammatical gender in Italian as a foreign language. *The Canadian Modern Language Review*, 54 (2), 239-262.
- Paolieri, D., Lotto, L., Morales, L., Bajo, T., Cubelli, R., and Job, R. (2010). Grammatical gender processing in romance languages: evidence from bare noun production in Italian and Spanish. *Eur. J. Cogn. Psychol.*, 22, 335-347
- Pérez-Pereira M (1991) The acquisition of gender: What Spanish children tell us. *Journal of Child Language*, 18, 571-90.
- Pesetsky, D. & Torrego, E. (2004). *The syntax of valuation and the interpretability of features*.

Ms. MIT/U.Mass Boston

- Pizzuto, E. & Caselli, M. C. (1992). The acquisition of Italian morphology: Implications for models of language development. *Journal of Child Language*, 19, 491-557
- Popova, M. I. (1958). Grammaticheskie elementy yazyka v rechi detei preddosh'kol nogo vozrasta [Grammatical elements of language and speech in preschool children]. *Voprosy Psikhol.*, 3, 106-117.
- Potter, M. C., So, K. F., von Eckhardt, B., & Feldman, L. B. (1984). Lexical and conceptual representation in beginning and more proficient bilinguals. *Journal of Verbal Learning and Verbal Behavior*, 23, 23-38.
- Prévost, P., & White, L. (2000) Missing surface inflection or impairment in second language acquisition? Evidence from tense and agreement. *Second Language Research*, 16, 103-33.
- Prévost, P., & White, L. (1999). Accounting for morphological variability in second language acquisition: Truncation or Missing Inflection? In M.-A. Friedemann and L. Rizzi (Eds.), *The acquisition of syntax* (pp.202-235). London: Longman publication. doi:10.1111/lang.12138.
- Putnam, M., & Sánchez, L. (2013). What's so incomplete about incomplete acquisition? – A prolegomenon to modeling heritage language grammars. *Linguistic Approaches to Bilingualism*, 3(4), 478-504. doi: 10.1075/lab.3.4.04put
- Rappaport, G. (2000). The Slavic Noun Phrase in comparative perspective. In G. Fowler (Ed.), *Comparative Slavic morphosyntax*. Bloomington: Slavica Publishers.
- Rebuschat, P. (2013). Measuring implicit and explicit knowledge in second language research. *Language Learning*, 63, 595-626. doi:10.1111/lang.12010
- Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, 42, 107-142.
- Rohrbacher, B. (1994). *The Germanic VO Languages and the Full Paradigm*. Doctoral dissertation, University of Massachusetts, Amherst.
- Sabourin, L., & Stowe, L. A. (2008). Second language processing: When are L1 and L2 processed similarly. *Second Language Research*, 24, 397-430.
- Salamoura, A., & Williams, J. N. (2007). The representation of grammatical gender in the bilingual lexicon: Evidence from Greek and German. *Bilingualism: Language and Cognition*, 10(3), 257-275.
- Sánchez, L. (2003). *Quechua-Spanish Bilingualism: Functional Interference and Convergence*. Amsterdam: John Benjamins. doi: 10.1075/lald.35

- Sánchez, L. (2004). Functional convergence in the tense, evidentiality and aspectual systems of Quechua-Spanish bilinguals. *Bilingualism: Language and Cognition*, 7, 147-162. doi: 10.1017/s136672890400149x
- Sánchez, L. (2015). Crosslinguistic influences, functional interference, feature reassembly and functional convergence in Quecha-Spanish Bilingualism. In S. Perpiñan and T. Judy (Eds.) *The Acquisition of Spanish as a Second Language: Data from Understudied Language Pairings* (pp.19-48). Amsterdam: John Benjamins.
- Schnell de Acedo, B. (1994). Early Morphological Development: The Acquisition of Articles in Spanish. In J. L. Sokolov, & Snow, C. E. (Eds.), *Handbook of Research in language development using CHILDES* (pp.210-253). Hillsdale, NJ [u.a.]: Erlbaum.
- Schriefers, H. (1993). Syntactic processes in the production of noun phrases. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 841-850
- Schwartz, B.D., & Sprouse, R.A. (1996). L2 cognitive states and the “full transfer/full access” model. *Second Language Research*, 12, 40-72. doi: 10.1177/026765839601200103
- Sekerina, A., Brooks, P. J., & Kempe, V. (2005, March). *Gender transparency facilitates noun selection in Russian*. Paper presented at the 18th Annual City University of New York Sentence Processing Conference, Tucson, Arizona.
- Snyder W., Senghas A., and Inman, K. (2001). Agreement morphology and the acquisition of noun-drop in Spanish. *Language Acquisition*, 9, 157-73.
- Socarrás, G. M. (2011). *First Language Acquisition in Spanish. A Minimalist Approach to Nominal Agreement*. London, New York: Continuum.
- Soler, M. R. (1984). Adquisición y utilización del artículo. In M. Siguán (Ed.), *Estudios sobre psicología del lenguaje infantil* (pp.139-165). Madrid: Pirámide.
- Spada, N., Shiu, J. L.-J., & Tomita, Y. (2015). Validating an elicited imitation task as a measure of implicit knowledge: Comparisons with other validation studies. *Language Learning*, 65, 723-751. doi:10.1111/lang.12129
- Suzuki, W., & DeKeyser, R. (2015). Comparing elicited imitation and word monitoring as measures of implicit knowledge. *Language Learning*, 1-36. doi: 10.1111/lang.12138
- Taraban, R., & Kempe, V. (1999). Gender processing in native and non-native Russian speakers. *Applied Psycholinguistics*, 20, 119-148.
- Taraban, R., & Roark, B. (1996). Competition in learning language-based categories. *Applied Psycholinguistics*, 17, 125-148.
- Tokowicz, N., & MacWhinney, B. (2005). Implicit and explicit measures of sensitivity to violations in second language grammar: An event related potential investigation. *Studies in Second Language Acquisition*, 2, 173-204.

- Tomasello, M. (2003). *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Cambridge, MA: Harvard University Press.
- Tomasello, M. & Brooks, P. (1999). Early syntactic development: A construction grammar approach. In M. Barret (Ed.), *The Development of Language* (pp.116-90). Hove: Psychology Press.
- Tsimpli, I. (2003) Features in language development. Invited plenary talk presented at EuroSLA, Edinburgh, September 2003.
- Unsworth, S. (2008). Age and input in the acquisition of grammatical gender in Dutch. *Second Language Research*, 24(3), 365-395.
- Unsworth, S., Argyri, F., Cornips, L., Hulk, A., Sorace, A., & Tsimpli, I. (2012). The role of age of onset and input in early child bilingualism in Greek and Dutch. *Applied Psycholinguistics*. doi:10.1017/S0142716412000574
- Vainikka, A., & Young-Scholten, M. (1994). Direct access to X⁰-theory: Evidence from Turkish and Korean adults learning German. In T. Hoekstra & B. D. Schwartz (Eds.), *Language Acquisition Studies in Generative grammar* (pp.265-316). Amsterdam: John Benjamins.
- Vainikka, A., & Young-Scholten, M. (1996). Partial transfer, not partial access. *Behavioral and Brain Sciences*, 19, 744-745.
- Vanpatten, B., & Benati, A. G. (2010). *Key terms in second language acquisition*. New York: Continuum International Publishing Group.
- Vigliocco, G., Antonini, T., & Garrett, M. F. (1997). Grammatical gender is on the tip of Italian tongues. *Psychological Science*, 8, 314-317.
- Vigliocco, G., Lauer, M., Damian, M. F. & Levelt, W. J. M. (2002). Semantic and syntactic forces in noun phrase production. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 28, 46-58.
- White, L., Valenzuela, E., Kozłowska-Macregor, M., & Leung, Y. I. (2004). Gender and number agreement in non-native Spanish. *Applied Psycholinguistics*, 25, 105-133.
- Zhang, R. (2015). Measuring university-level L2 learners' implicit and explicit linguistic knowledge. *Studies in Second Language Acquisition*, 37, 457-486. doi:10.1017/S0272263114000370

Appendices

Experiment 1

Grammaticality Judgment Task (GJT)

	Masculine items		Feminine items	
	Congruent	Incongruent	Congruent	Incongruent
1.	oceano	cuarto	pizarra	cerveza
2.	diario	bolígrafo	plaza	playa
3.	viento	ejercicio	comida	bicicleta
4.	frigorífico	gorro	montaña	novela
5.	lápiz	amor	historia	computadora
6.	color	imperio	lámpara	nevera
7.	experimento	rasgo	tortuga	alfombra
8.	barrio	dormitorio	ropa	piscina
9.	nivel	cuaderno	escultura	danza
10.	hablante	cigarillo	sonrisa	pelota
11.	televisor	papel	pizza	maleta
12.	azucar	peligro	telenovela	ducha

Picture Matching Task (PMT)

	Masculine items		Feminine items	
	Congruent	Incongruent	Congruent	Incongruent
1.	Ojo	Pescado	Blusa	Corbata
2.	Yogur	Periódico	Taza	Pulsera
3.	Ajo	Tenedor	Chaqueta	Cebolla
4.	Cinturon	Libro	Fresa	Naranja
5.	Tomate	Brazo	Pizza	Copa
6.	Queso	Pollo	Puerta	Manzana
7.	Armario	Libro	Servilleta	Corbata
8.	Oso	Coche	Cereza	Copa
9.	Pepino	Maíz	Camisa	Isla
10.	Azucar	Bolso	Sandalia	Casa

Experiment 2

	Lexical-semantic condition	Lexical-syntactic reliable cues condition	Lexical-syntactic reliable cues condition
--	----------------------------	---	---

1.	niña	casa	calle
2.	hermana	silla	leche
3.	abuela	guerra	llave
4.	novia	vista	sangre
5.	señora	falta	tarde
6.	prima	bolsa	muerte
7.	maestra	cama	mano
8.	hija	carta	piel
9.	tía	ducha	voz
10.	chica	lucha	ley
11.	amiga	playa	pared
12.	sobrina	cosa	carne
13.	mamá	camisa	suerte
14.	camarera	puerta	nieve
15.	cocinera	cabeza	nariz
16.	azafata	mesa	salud
17.	abogada	semana	clave
18.	enfermera	cerveza	miel
19.	vendedora	cena	luz