

APPROVAL SHEET

Title of Dissertation: *Enhanced input and enriched context to improve the acquisition of the Spanish grammatical gender assignment and agreement: A preliminary study*

Name of Candidate: Mireia Toda Cosi
Master's in Intercultural Communication, 2019

Dissertation and Abstract Approved: Ana María Schwartz Caballero
Dr. Ana María Schwartz Caballero
Associate Professor of Spanish and Second
Language Education, Spanish Area Coordinator
Modern Languages, Linguistics, and
Intercultural Communication (MLLI)

Date Approved: 19 July 2019

ABSTRACT

ENHANCED INPUT AND ENRICHED CONTEXT TO IMPROVE THE ACQUISITION OF THE SPANISH GRAMMATICAL GENDER ASSIGNMENT AND AGREEMENT: A PRELIMINARY STUDY

Mireia Toda Cosi, Master's in Intercultural Communication, 2019

Directed by Dr. Ana María Schwartz-Caballero, MLLI

Research in Second Language Acquisition has diagnosed extensively how the acquisition of a language's grammatical gender system poses problems to learners of a second language while it is acquired effortlessly by native speakers. In addition, the main issue that has been addressed by the literature concerns assignment alone, thus not taking into consideration the full scope of the grammatical gender system. Research on types of learning has shown that it is possible to learn incidentally, namely without noticing, which allows for more efficiency time-wise. However, several factors modulate both the acquisition of a language's grammatical gender learning and types of learning, the main factor being proficiency.

With the goal to facilitate the acquisition of the Spanish grammatical gender for learners, this study puts forward a pedagogical approach that consists of enhanced and enriched input. Four treatment groups were created in a crossed design with two groups with enhanced input (bold font) and two without it. One group with enhanced and one unenhanced had enriched context (article-noun-adjective combination) and one did not (article-noun combination). Participants were exposed to a PowerPoint presentation that presented 30 novel nouns that were controlled by grammatical gender (masculine or

feminine) and noun ending (canonical, non-canonical, and exceptions) combined with 40 novel adjectives (in the article-noun-adjective condition, of which 20 were variable and 20 were invariable).

Results showed that treatments did not render significant differences and that proficiency was the best predictor for accuracy in gender assignment and agreement. Masculine nouns showed significantly higher accuracy rates with respect to feminine nouns and nouns with a canonical ending showed higher accuracy rates with respect to non-canonical and exceptional endings. Results in the delayed post-test show that there is a significant improvement in performance for canonical noun endings, while exceptional endings decrease significantly and non-canonical endings remain the same. No differences are observed between tests regarding gender of the nouns.

These results provide tentative evidence for Ullman's (2001) Declarative/Procedural model since results would suggest that the unmarked noun endings were proceduralized, but not other features. It is also argued that these learners show improvement depending on their proficiency, thus being capable of ultimate attainment. The results suggest that the errors made by learners are the result of features being reassembled in the learners' internal linguistic system.

ENHANCED INPUT AND ENRICHED CONTEXT
TO IMPROVE THE ACQUISITION OF THE
SPANISH GRAMMATICAL GENDER
ASSIGNMENT AND AGREEMENT:
A PRELIMINARY STUDY

by

Mireia Toda Cosi

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, Baltimore County in partial fulfillment
of the requirements for the degree of
Master of Arts in Intercultural Communication
2019

ACKNOWLEDGMENTS

I would like to extend my gratitude to my supervisor, Dr. Schwartz-Caballero, for her patience and support. I am also particularly grateful for the guidance Dr. Long provided me with, to the extent of being almost like a second supervisor and mentor. Many thanks too to the rest of the committee, Dr. Lambert-Bretiere and Dr. Oskoz, for their patience and support throughout.

My thankfulness also goes to the 201 Spanish instructors at UMBC, Mílvia, Mary-Ely, and especially Catalina and Sean, who let me come to their classrooms and gather participants. Also thanks to Elisabeth and María for their moral support throughout the process. Along with them, I would also like to thank all the students who agreed to participate in the study and made this research possible. Regards to UMBC's CIRTL for their input on the data layout. I would also like to thank Erika's advice and orientation with regard to general recommendation with regard to statistics.

Moltíssimes gràcies també a tots els de casa, als amics i a la família sobretot, per ser-hi des de la distància i ajudar-me a aclarir les idees i mantenir el rumb i el seny, per bé que ben poc me'n quedi. No sé pas on seria sense vosaltres, de ben segur no tan lluny. I la meva més sincera gratitud a la meva parella, qui de tanta ajuda com m'ha brindat bé es mereixeria co-autoria d'aquest treball, gràcies per les hores buscant imatges, fent de conillet d'índies, consultor, editor, ... I tantes coses més, però sobretot per creure en mi i ajudar-me a mantenir la calma quan ja l'havia perduda tota.

Thanks to you all for making this possible

Moltes gràcies a tots per haver fet que això fos possible

TABLE OF CONTENTS

ACKNOWLEDGMENTS	ii
TABLE OF CONTENTS.....	iii
LIST OF ABBREVIATIONS.....	vi
CHAPTER I: INTRODUCTION	1
The Spanish Grammatical Gender System.....	2
CHAPTER II: THE ACQUISITION OF GRAMMATICAL GENDER	9
<i>Grammatical Gender</i>	15
<i>Factors in the Acquisition of Grammatical Gender</i>	16
CHAPTER III: TYPES OF INSTRUCTION AND LEARNING	27
<i>Implicit and Explicit Learning</i>	27
<i>Intentional and Incidental Learning</i>	31
<i>Incidental and Intentional Learning of Vocabulary</i>	35
<i>Incidental Learning of Grammatical Features</i>	42
<i>Teaching of Grammatical Gender Systems</i>	47
CHAPTER IV: METHODOLOGY	50
<i>Participants</i>	53
<i>Research Design</i>	56
<i>Materials and procedure</i>	59
<i>Consent form</i>	60
<i>Lextale and linguistic background questionnaire</i>	60
<i>Treatment and answer sheet</i>	61
<i>Form recognition test</i>	65
<i>Translation test</i>	66
<i>Gender concordance test</i>	67
<i>Debriefing</i>	69
<i>Data coding</i>	70
CHAPTER V: STATISTICAL RESULTS	72

<i>RQ1: What differences are there between groups regarding the gender of nouns combined with concordance?</i>	73
<i>RQ2: What differences are there between groups regarding noun ending combined with concordance?</i>	75
<i>RQ3: What differences are there between accuracy of concordance in new versus old pairings of nouns and adjectives?</i>	80
<i>RQ4: What differences in performance are there between concordance, noun ending, gender and treatment?</i>	82
<i>RQ5: What differences are there between groups regarding the gender of nouns and the immediate and delayed post-test in assignment?</i>	84
<i>RQ6: What differences are there between groups regarding the gender of nouns and the immediate and delayed post-test in agreement?</i>	86
<i>RQ7: What differences are there between groups regarding noun ending and the immediate and delayed post-test in assignment?</i>	88
<i>RQ8: What differences are there between groups regarding noun ending and the immediate and delayed post-test in agreement?</i>	92
<i>RQ9: What differences are there between the learning of adjectives and nouns in the ANAB, ANA and control groups?</i>	97
<i>RQ10: What differences are there between the learning of adjectives depending on their frequency in the ANAB, ANA and control groups?</i>	98
<i>Correlations Between Dependent and Independent Variables</i>	99
CHAPTER VI: DISCUSSION	102
<i>Differences Among Treatment Groups.....</i>	102
<i>Differences Among the Grammatical Gender of the Nouns.....</i>	107
<i>Differences Between Concordance</i>	108
<i>Differences Among Noun Endings</i>	109
<i>Differences Between New and Old Noun-Adjective Pairings</i>	111
<i>Correlation of Variables</i>	111
<i>Learning of Adjectives and Nouns and Frequency Effects</i>	112
<i>Ultimate Attainment</i>	114
Chapter VII: Conclusions, Limitations and Future Research	116
<i>Limitations and future research</i>	118
APPENDIX A.....	122
APPENDIX B.....	124
APPENDIX C.....	125

<i>Consent Form</i>	126
<i>Lextale and Linguistic Background Questionnaire</i>	128
<i>Form Recognition Test (ANAB and ANA Groups)</i>	133
<i>Form Recognition Test (ANB and AN Groups)</i>	134
<i>Multiple Choice Translation Test (ANAB, ANA, and Control Groups)</i>	135
<i>Multiple Choice Translation Test (ANB and AN)</i>	140
<i>Multiple Choice Concordance Test</i>	145
<i>Debriefing Questionnaire</i>	149
<i>Delayed Post-Test for Known Nouns</i>	150
<i>Delayed Post-Test for Treatment Nouns</i>	154
<i>Delayed Post-Test of Unknown Nouns</i>	158
APPENDIX D.....	161
<i>RQ1: What differences are there between groups regarding the gender of nouns combined with concordance?</i>	161
<i>RQ2: What differences are there between groups regarding noun ending combined with concordance?</i>	165
<i>RQ3: What differences are there between accuracy of concordance in new versus old pairings of nouns and adjectives?</i>	169
<i>RQ4: What differences in performance are there between concordance, noun ending, gender and treatment?</i>	173
<i>RQ5: What differences are there between groups regarding the gender of nouns and the immediate and delayed post-test in assignment?</i>	176
<i>RQ6: What differences are there between groups regarding the gender of nouns and the immediate and delayed post-test in agreement?</i>	178
<i>RQ7: What differences are there between groups regarding noun ending and the immediate and delayed post-test in assignment?</i>	180
<i>RQ8: What differences are there between groups regarding noun ending and the immediate and delayed post-test in agreement?</i>	183
<i>Correlations</i>	185
<i>Data coding</i>	186
REFERENCES	189

LIST OF ABBREVIATIONS

AN	Article-Noun
ANA	Article-Noun-Adjective
ANAB	Article-Noun-Adjective with Bold enhancement
ANB	Article-Noun with Bold enhancement
DP	Declarative/Procedural model
FDH	Fundamental Difference Hypothesis
FFFH	Failed Functional Features Hypothesis
FL	Foreign Language
FRH	Feature Reassembly Hypothesis
FTFA	Full Transfer / Full Access
GJT	Grammaticality Judgment Task
HS	Heritage Speaker
ILH	Involvement Load Hypothesis
L1	First Language
L2	Second Language
L3	Third Language
MSIH	Missing Surface Inflection Hypothesis
NNS	Non-Native Speaker
NS	Native Speaker
SLA	Second Language Acquisition
TL	Target Language

CHAPTER I: INTRODUCTION

The field of Second Language Acquisition (SLA) has addressed a wide range of questions, from processing to pedagogy. One of the aspects that has gathered quite some evidence and a breadth of literature is the acquisition of grammatical gender, how it is acquired, the possibility of ultimate attainment and its processing. Research has shown differences in processing between native speakers and second language learners, however, as is the case in behavioral data, proficiency level seems to have strong modulating effects. Nonetheless, currently there seems to be consensus in that non-native speakers can eventually perform and behave like native speakers.

While most research has focused on gender assignment (attributing a gender to a noun, which is usually displayed through a determiner), few studies have explored gender agreement (making the noun agree with other features such as adjectives, possessive articles, demonstratives, pronouns, etc.); and even when doing so the testing materials might have given away the gender of the words, thus biasing the results (e.g. Montrul, Foote, & Perpiñán, 2008). Regardless, results show decreased comparative accuracy in agreement than in assignment. The corpus of research regarding the acquisition of grammatical gender is however extensive and, to some extent, contradictory with regard to the abilities of learners.

Overview of The Study

The majority of studies researching learning of grammatical gender used participants who departed from a clean slate and complete unfamiliarity with the targeted feature in the language of the study; nevertheless, to the best of my knowledge, no study has addressed

how the teaching of new items that abide by an already known feature, grammatical gender, can facilitate the acquisition thereof. This is the goal of this study, to explore the differences in assignment and agreement depending on the gender of nouns, their ending, and pairing across stimuli in Spanish.

Despite most studies diagnosing the problems in the acquisition of grammatical gender, some have also attempted to look into the teaching thereof: implicitly, explicitly, intentionally and/or incidentally. The results in that regard are inconclusive and do not provide sufficient evidence to disregard any instructional method. It is however hard to dismiss the advantages incidental learning, namely learning without awareness or intention, would have since it would allow for learners to acquire the feature while learning another aspect of the language at the same time; for this reason, this study has opted to use incidental learning to explore its success when teaching grammatical gender of novel nouns in Spanish. Despite the overarching idea of grammatical gender being similar across languages, namely that words are randomly assigned a gender, each language has its own specific dynamics, e.g. what elements the nouns show agreement with, the grammatical genders available for assignment, etc. This study has used Spanish as the target language, for this reason, a description of the gender system of the language is provided below.

The Spanish Grammatical Gender System

In Spanish gender is binary, meaning that it has masculine and feminine gender (as per opposition to other systems that differentiate neuter/non-neuter like Dutch or Danish or systems with more genders, e.g. German or Russian which add neuter). Not only that, but Spanish is considered to be a so-called transparent gender system, namely it has a highly regular system that makes the gender of nouns predictable; in contrast to opaque systems

such as that of Dutch or Danish in which gender is considered mainly arbitrary, or semi-opaque systems in which regularities are more present but still make guessing complicated, such as German or French. The “rule of thumb” taught to most Spanish L2 learners is that nouns ending in *-a* are feminine and those ending in *-o* are masculine. And that is true, to a certain extent. Based on the 18th century *Diccionario de la lengua española* by the Spanish Royal Academy, Teschner & Russell (1984) reported that 96.30% of nouns ending in *-a* were feminine while 99.8% of nouns ending in *-o* are masculine. This, however, is not the full picture. According to their (somewhat outdated) study, the typical endings for feminine nouns are *-a* and *-d*, for masculine they are *-i*, *-o*, *-r* and *-e* (in the latter, the majority of words as masculine, but the most frequent words are feminine), and some endings are indeterminate *-z*, *-n*, and *-s*. For the purpose of this study, I focused on the most regular endings (*-o*, *-a*), non-canonicals (*-consonant*, *-e*), and exceptions, see Table 1 for examples and an overview.

Table 1
The Spanish Grammatical Gender

Type of noun ending	Masculine		Feminine	
	Ending	Example	Ending	Example
Canonical (C)	-o	libro (<i>book</i>)	-a	cama (<i>bed</i>)
Non-canonical (NC)	-consonant, -e	árbol (<i>tree</i>)	-consonant, -e	llave (<i>key</i>)
Exception (E)	-a	mapa (<i>map</i>)	-o	mano (<i>hand</i>)

Within these distinctions there are some categories worth noticing as reported in Teschner & Russell (1984). One is that of gender-ambivalent nouns, which are names that do not change regardless of biological gender, they usually refer to occupations and the majority of them end in *-ista*, a common suffix for jobs, e.g. *el tenista* “the tennis player_{MASC}” and *la tenista* “the tennis player_{FEM}”. This study used four of these words: *soprano* “soprano”, *modelo* “modelo”, *fantasma* “ghost” and the previous example.

Despite the caveat that both genders could be assigned to some of them, learners were only exposed to pictures and gender marks referring to one of the two genders, for this reason, learners were measured with regard to the gender they were taught during the treatment (see Chapter IV, Methodology). Another category worth mentioning is gender-ambiguous names, those that “change meaning when changing gender” (p. 116). Of these, *el cura* “the priest_{MASC}” was used, which in its feminine *la cura* it stands for “the cure”, thus, considering that learners were primed to the difference in meanings during exposure to the treatment, learners’ responses were measured against what they were exposed to.

Languages with a grammatical gender system are not only complex inasmuch as nouns are believed to have a gender (assignment, which tends to be shown in determiners) but also inasmuch as they need to *agree* with other features, such as demonstratives, pronouns, possessive articles, adjectives, modifiers, indefinite articles, etc. Spanish is no exception to this phenomenon, thus it does not suffice that one knows the determiner that goes along with a word, but rather needs to make it agree with other items in the sentence, as illustrated in (1), and they can be in the direct immediacy of the noun or placed at further distance. However, there are some elements that do *not* show gender agreement, from this category only invariable adjectives were chosen for this study. In Spanish invariable adjectives do not show agreement and the majority end in -consonant (*infantil* “infantile/childish”), *-e* (*enorme* “enormous”), or *-a* (*agrícola* “agricultural”), their use is illustrated in (2).

- (1) C1. el libro abierto
 the-MASC book-MASC open-MASC
 ‘the open book’
 C2. la casa abierta
 the-FEM house-FEM open-FEM
 ‘the open house’
 NC1. el árbol pequeño

		the-MASC	tree-MASC	tree-MASC
			‘the small tree’	
NC2.	la	the-FEM	llave key-FEM	pequeña small-FEM
			‘the small key’	
E1.	el	the-MASC	mapa map-MASC	cerrado close-MASC
			‘the closed map’	
E2.	la	the-FEM	mano hand-FEM	cerrada small-FEM
			‘the closed hand’	
(2)	C1.	el	libro the-MASC book-MASC	infantil childish-MASC
			‘the children’s book’	
	C2.	la	casa the-FEM house-FEM	infantil childish-FEM
			‘the children’s house’	
	NC1.	el	árbol the-MASC tree-MASC	verde green-MASC
			‘the green tree’	
	NC2.	la	llave the-FEM key-FEM	verde green-FEM
			‘the green key’	
	E1.	el	mapa the-MASC map-MASC	azul blue-MASC
			‘the blue map’	
	E2.	la	mano the-FEM hand-FEM	azul blue-FEM
			‘the blue hand’	

Having examined Spanish gender system, being aware of its irregularities, and having seen the struggles of L2 learners, and even heritage speakers in certain contexts (Montrul et al., 2008), when assigning gender to nouns, one might wonder what happens with neologisms or loaned nouns from other languages when they are processed by native speakers. Do they assign a gender by default? Does it depend upon the source language’s gender system? Studies such as Clegg’s (2010) have explored such issues. Using English words, they had Spanish speakers assign them a gender and they found high rates of

consensus and confirmatory data of the tendencies described above, thus confirming previous studies such as Bull's (1965). These data seem to indicate that native speakers develop an intuition for the gender of the words, even unknown words, which makes way for a follow-up question, can non-native speakers develop such intuition as well? In the current analysis the ability to generalize is only observed in the treatment words being paired with different adjectives, but other phases of this study also allow to observe whether non-native speakers can develop the ability to generalize the knowledge to known and novel words by being exposed to the gender system in more depth, see Chapter IV, Methodology.

Methodology

In order to address the current inquiry, this study proposes to expose learners to two kinds of stimuli (1) a combination of an article with a noun and (2) an article with a noun and an adjective. Each of these combinations is then divided into two different conditions, one in which the targeted items are enhanced with bold font and the other one which is not. The expectation is that learners will learn incidentally and implicitly, without noticing, but benefit from their attention being drawn towards the agreement by means of bold font while being told to only learn the vocabulary. In addition, larger exposure to agreement connections is expected to enhance the participants' accuracy in assignment and agreement.

Since the experiment took place in a classroom setting and the enhancing proposed, use of bold font, is cost-efficient and easy to implement, it was hypothesized that positive findings would facilitate learners' acquisition of grammatical gender, provide further evidence pro/against ultimate attainment (namely whether learners can acquire native-like

competence and processing in a language), and provide some insight into how learners learn their ability to generalize knowledge. In addition, were findings to be generalizable to unknown words, it could be argued that learners have the potential to develop an intuition similar to that of native speakers, who show high rates of agreement even in novel nouns.

Research Questions

The study aimed to answer the following research questions with regard to Spanish:

What is the difference in accuracy in gender assignment and agreement in learners depending on:

- 1 the grammatical gender (masculine or feminine) of the word and treatment group?
- 2 noun-ending (canonical, non-canonical, exceptions)?
- 3 old versus new pairings of nouns and adjectives?
- 4 time, do results hold two weeks later? If so, how are they (dis)similar?

What differences are there in learning depending on:

- 5 nouns versus adjectives.
- 6 frequency of exposure to the adjectives.

Summary

This chapter presented an overview of the study under consideration and an overview of the Spanish grammatical system. Chapter II presents a review of the literature regarding the acquisition of grammatical gender, focusing on attainment, the factors that affect accuracy among learners, and an emphasis on the acquisition of the Spanish grammatical

gender. In Chapter III types of learning are addressed, first covering studies that looked into explicit and implicit learning and then moving on to intentional and incidental learning. A miscellanea of other methodologies is briefly covered in order to show the inconclusiveness of said studies and the difficulty they pose with respect to being implemented. In both, Chapter II and Chapter III, several factors that influence learning have been discussed. Chapter IV describes the methodology that has been used, followed by Chapter V where the results are presented, Chapter VI with the discussion, and Chapter VII presents the conclusions, the study's limitations and future research.

CHAPTER II: THE ACQUISITION OF GRAMMATICAL GENDER

A longstanding question in second language acquisition (SLA) has been that of ultimate attainment, namely whether second or foreign language learners are capable of attaining the same level of proficiency and perform like native speakers (NS) of a given language, especially late learners. In this study learners are referred to as second language learners regardless of it being their second or further language. Such debate applies as well to the domain of grammatical gender, which is the object of study of this work. The literature¹ is split on the argument of ultimate attainment being available to late learners or not; however, research has provided a wealth of theories that nuance both positions. Given the variety of approaches I will just address the most widespread positions: the Failed Functional Features Hypothesis (FFFH) by Hawkins & Chan (1997; Smith & Tsimpli, 1995), the Full Transfer/Full Access Hypothesis (FTFA) by Schwartz & Sprouse (1996), the Feature Reassembly Hypothesis (FRH) by Lardiere (2008, 2009), and the Declarative/Procedural Model (DP) by Ullman (2001). In addition, there are some derived theories that build upon the aforementioned four. Considering that the domain of inquiry of this research is the acquisition of a language's grammatical gender system, the general theories will be presented and articles supporting them will show their applicability regarding the learnability of the Spanish grammatical gender system.

The Fundamental Difference Hypothesis (FDH) posits that the process of acquiring the first (L1) and the second language (L2) are fundamentally different on the basis that

¹ The literature cited regarding ultimate attainment is further developed when discussing the factors affecting learning, for this reason only the main outcome is to be addressed in the current section.

acquiring a first language departs from a clean slate while a second language is unavoidably dependent on the structure of the first language that was learnt. Recent research though has shown that there are differences not only between the acquisition of an L1 and an L2, but even between learning an L1 in a complete immersion environment versus doing so as a heritage speaker (HS), as shown by Montrul et al.'s (2008). In their study they found evidence that heritage speakers show differences from both native speakers and L2 learners, outperforming L2 learners orally but not in writing and only matching native speakers in oral performance. One could then argue that the developmental path of a bilingual individual also differs from that of a monolingual and that bilinguals also follow a significantly different developmental path or that their development is uneven.

Building on the FDH, the FFFH posits that adults who did not activate a certain uninterpretable feature before the critical period will not be able to access it, meaning that they will not be able to access those features that do not have an explicit contribution to meaning but that are reflected at a morphosyntactic level —e.g. gender agreement. Such learners are not perceived as capable of attaining native levels and, if they do attain those level behaviorally, there is an assumed “non-native underlying competence” (Hawkins, 2005; as cited in Blom, Polisenska, & Weerman, 2007; p. 298). Thus, the FFFH posits that there is an ongoing representational issue in late learners. It is assumed that syntactic representation and of the lexicon are divided and that learners who attain native-like performance behaviorally resort to associative lexical learning as a compensatory strategy. Recently, the authors have redefined the theory in response to evidence of native-like processing by late learners in several domains; they now lean towards it applying only to

uninterpretable features, purely grammatical issues (e.g. Hawkins & Casillas, 2008).

Regarding acquisition of grammatical gender, the theory is summarized by Hopp (2012):

[both the lexical gender learning and] the Representational Deficit hypothesis argue that lexical aspects of gender are implicated in L2 inflectional variability with syntactic gender agreement. [...] For the Representational Deficit hypothesis, associative lexical learning acts as a compensatory strategy to memorize exceptional, e.g. non-default, co-occurrence of gender marked forms and nouns because the L1 morphosyntax restricts availability of syntactic feature checking in adult L2 acquisition (p. 38).

Most of the evidence in support of the FFFH comes from studies in processing arguing that L2 learners process the target language (TL) differently than their L1 counterparts and that that is due to a faulty mental representation of their grammar. Several studies have looked into the acquisition of Spanish grammatical gender, the most cited are those by Lew-Williams (2009) and Lew-Williams & Fernald (2010), in both cases they found that L2 speakers of Spanish were significantly worse than L1 children and adults at using gender predictively in an eye-tracking experiment, thus concluding that they had failed to acquire the gender feature. However, as will be discussed, in later studies they interpret similar results differently in light of the advances in the field (Grüter, Lew-Williams, & Fernald, 2012).

In contrast, the Full Transfer/Full Access theory, posits that the entire system of the first language (L1) is transferred as the base the second language (L2) is going to be built upon. Once it has been transferred, the parameters of the L1 are adjusted to accommodate those of the L2 and can be adjusted to process language like a native speaker. In an eye-tracking experiment Dussias, Valdés Kroff, Guzzardo Tamargo, & Gerfen (2013) found evidence that L1 English-L2 Spanish speakers can process the Spanish grammatical gender

like native speakers if learners are of sufficient proficiency. Dussias et al. (2013) argue for this position:

across numerous studies, the accuracy rate for gender assignment and gender agreement is quite high, most often ranging between 80 and 90%. [...] Even those studies that argue most strongly for fundamental representational differences between native and nonnative speakers report accuracy rates above the 75-80% mark” (p. 355)

Building on the FTFA, the Missing Surface Inflection Hypothesis (MSIH) assumes that there is access along with a simultaneous mapping problem, which explains the emergence of errors. Learners have access to the syntactic feature of the language but have trouble mapping it onto the lexical realizations of it. It does, however, consider that it is within the learner’s capacity to acquire the linguistic feature. This would explain the path learners follow to attainment despite making errors along the way. Support for this theory comes from Lew-Williams. Revisiting his previous results in a new study with Grüter et al. (2012) they found that in offline comprehension L2 learners of Spanish performed at ceiling, showed errors in elicitation, and could not take as much advantage of gender cues in online processing. Thus concluding that it was an issue with the local representation of grammatical gender. Having measured written comprehension and oral production errors in L2 learners, HSs, and NSs Alarcón (2011) concludes that, at advanced proficiency, learners do have gender in underlying grammars and that the problem is with the “surface manifestations of the abstract features of gender” (p. 332).

Lardiere (2008, 2009) further hypothesizes that one of the most significant challenges for late learners is resetting the parameters of an already existing feature in their L1 rather than acquiring a new feature altogether, which she formulates into the Feature Reassembly Hypothesis (FRH). Support for the FRH is found in studies regarding different languages,

such as Swahili (Spinner, 2013), Italian (Santoro, 2012), and Spanish (Gabriele, Fiorentino, & Alemán Bañón, 2013). Behavioral data from Spinner (2013) shows that learners improve with proficiency, which she argues reflects features being reassembled. Santoro (2012) observed English learners of Italian and found that L2 learners could use determiner phrases, were able to acquire to use pronouns correctly, and that uniform errors, namely systematically assigning the wrong gender to a word, showed a deficiency in the assembly of the gender feature rather than a misrepresentation of the system, thus furthering the case for FRH. With regard to Spanish, Gabriele et al. (2013) on an ERP study² on processing conclude that the FRH is the best account to explain why English-speaking learners of Spanish show qualitatively similar processing to that of NSs.

The last theory that will be addressed is the Declarative/Procedural model (DP). According to Ullman (2001) different types of memory, declarative and procedural, realize different kinds of knowledge. Declarative memory would represent explicit knowledge and conscious retrieval of information, while procedural memory is implicit and reflects unconscious automatized processes. On that basis lexical knowledge, irregularities, and lexically based linguistic features, would rely on declarative memory; while grammatical features would rely on implicit procedural processes. Ullman posits that children's learning process relies on procedural knowledge while memorization, arguably the technique used by most late learners, relies on declarative memory, thus accounting for the differences we might observe in the acquisitional path between the two. It is then a difference in processing that we are facing rather than any deficit. His theory does not rule out the possibility of

² Event-related potential (ERP) is a technique that allows to measure the electrical impulses occurring in our brains and that reflected, among other things, how we process language.

turning declarative knowledge into procedural, it only states that the pathways to declarative memory remain stronger than those of procedural memory in adults versus children. In favor of the possibility of knowledge going from declarative to procedural and thus displaying native-like effects Sagarra & Herschensohn (2021) found that English learners of Spanish could qualitatively perform like NSs, although it could also be interpreted as aligning with the FRH. The same applies to Morgan-Short, Sanz, Steinhauer, & Ullman (2010), although their findings are a better advocate for the DP model. In their research they use explicit and implicit teaching and early exposure to implicit learning seems to yield a P600³ effect after training while being first exposed to explicit instruction led to N400⁴. N400 effects are attributed to declarative memories, while P600s are believed to reflect procedural memory, thus not only proving that L2 learners have access to procedural memory, but also that first exposing learners to implicit methods aids in the process.

In summary, most research now supports that L2 learners can attain native-like proficiency, nevertheless, data can be interpreted differently in support of different theories. Some theories like the MSIH and FRH could be defended by the same data, as they only differ in the process, namely the MSIH diagnoses a specific point on time while the FRH addresses development. Other theories like the DP look at where and how the information is stored which, arguably, could be read as the data being stored in the procedural memory being the one looking into syntactic representations. In conclusion, the theories are neither clear cut nor do they necessarily exclude one another, what is clear

³ Language-related event-related potential (ERP) that is thought to be the reaction to either reading or hearing ungrammatical language or syntactic anomalies. It shows itself as a negative wave.

⁴ Negative deflection in an ERP at about 400 milliseconds after perceiving a semantic violation.

though is that the trend is to consider that ultimate attainment is possible among L2 learners.

Grammatical Gender

Having described the most salient theories in SLA regarding the possibility of ultimate attainment and thus native-like performance, we should address their applicability to the feature under observation in this study: grammatical gender acquisition. Grammatical gender concerns nouns which have a gender assigned and that show agreement with other features of the language, such as, for instance, possessive, demonstrative or indefinite articles, adjectives, pronouns, or verbs. Grammatical gender can be observed through assignment, the noun is considered to be of a specific gender, or in agreement, the noun is to match another linguistic feature that complements it, e.g. adjectives. Assignment is usually ascribed to the use of definite articles, since most vocabulary is taught along with them, while agreement is reserved to other features. As Chini (1995) explains, grammatical gender is a lexical category half way between syntax and semantics and is realized in inflectional and derivational morphology. Nouns realize gender in the lexical and derivative domain, while the gender of adjectives is realized through inflectional morphology. Chini (1995) continues, on pointing out that in the early stages of acquisition adults tend to base gender on noun ending as a “Regola di Base” (*Rule of Thumb*, Chini, 1991; p. 679). It is then through formal features that gender is first approached rather than through biological gender. There are studies as well that point at learners resorting to a default gender, in cases like Spanish it would be the masculine since that is what is used to refer to abstractions and neutral forms (e.g. *ello*, *lo*, etc.; Corbett, 1991; p. 214), but in systems such as that of Dutch where the dichotomy is neuter versus non-neuter the default

is non-neuter (Corbett, 1991). Even though research has found learners to ascribe to that pattern (e.g. Martínez-Gibson, 2011), results are mixed, since other studies found the opposite (e.g. de Martino, Bracco, Postiglione, & Laudanna, 2017). Gender systems are also considered to have different levels of opacity. Languages such as Spanish or Italian are considered to be transparent, since the association of *-o* to the masculine and *-a* to the feminine cover the majority of words; while systems such as that of German, French, or Polish are considered more opaque, having a series of patterns that are not as predictable; while languages such as Norwegian, Dutch, or Danish are the most opaque, with gender assignment providing little to no formal cues (Corbett, 1991; Jedynak & Pytlarz, 2012).

Factors in the Acquisition of Grammatical Gender

The acquisition of grammatical gender has extensively been addressed by the literature, looking at native language, bilingual and foreign language acquisition in a variety of languages, such as German (Hopp, 2012; 2016), Swahili (Spinner, 2013), Dutch (Blom et al., 2007; Lemhöfer, Schriefers, and Hanique, 2010; Blom, Polisenska, & Unsworth, 2008; Cornips & Hulk, 2008; Brouwer, Sprenger, and Unsworth, 2017), Spanish (Morgan-Short et al., 2010; Dussias et al., 2013; Grüter et al., 2012), French (Foucart & Frenck-Mestre, 2011), and Italian (Santoro, 2012). The literature has extensively covered the phenomenon, from ERP or eye-tracking⁵ for processing to behavioral studies with results both in support and against the aforementioned theories. The consensus previous literature has reached is that the acquisition of grammatical gender is not determined by a sole factor, but rather is a complex construct affected by several variables such as proficiency (e.g. Martínez-

⁵ Eye-tracking studies present visual stimuli to participants and their eye movements are recorded in order to gain insight into where the attention is directed or how the information is processed depending on where they put their attention or how fast it shifts.

Gibson, 2010; Dussias et al., 2013; Hopp, 2016; Hopp & Lemmerth, 2016), working memory (e.g. Sagarra & Herschensohn, 2012; McDonough & Trofimovich, 2018), quantity and quality of input (i.e. Cornips & Hulk, 2008; Blom et al., 2008; Siegelman & Arnon, 2015; Rodina & Westergaard), transparency of the gender system (e.g. Vigliocco, Vinson, Paganelli, & Dworzynski, 2005; Caffarra, Siyanova-Cahnturia, Pesciarelli, Vespignani, & Cacciari, 2015; De Martin et al., 2017), animacy (e.g. Vigliocco et al., 2005), type of instruction (e.g. Tokowicz & MacWhinney, 2005; Foote, 2008; Morgan-Short et al., 2010; Lemhöfer et al., 2010), L1 influence (e.g. Jedynak & Pytlarz, 2012; Spinner, 2013) and age of acquisition (e.g. Lew-Williams & Fernald, 2007; Blom et al., 2007; Montrul et al., 2008; Lew-Williams, 2009; Lew-Williams & Fernald, 2010). Other aspects that have been reported to affect performance or the perceived attainment of the learners are the type of data, written or aural (e.g. Montrul et al., 2008). Results also vary depending on the measurement being online (event related potentials —ERP—, eye-tracking, moving window experiments) or offline (written production data, grammaticality judgment tasks). In summary, the design and variables accounted for can easily tip the scales. Most studies address several factors at a time, for this reason articles were classified according to the most relevant factor and other variables under observation are mentioned to a lesser extent.

Regarding L1 influence, Sabourin and Stowe (2008) make the strongest claim for it. They tested Dutch NSs, German learners of Dutch and Romance learners of Dutch. Based on linguistic proximity, the expectation was that German speakers would be able to process Dutch in a similar way to Dutch NSs but that Romance speakers would only be able to perform similarly on verbs. They looked for a P600 effect regarding dependencies on the

verbal domain and processing of gender agreement and found that NSs showed the effect in both cases and so did German speakers, albeit to a lesser extent, while Romance speakers failed to show a P600 in gender agreement. The outcome supports the theory that similar neural processing is possible in L1 and L2, but that L1-L2 dissimilar lexically-driven constructions do not render similar neural processing. Lemhöfer et al. (2010) showed also that gender assignment was more consistent in gender-congruent nouns between German and Dutch while assignment in gender incongruent words was more unstable. Jedynak & Pytlarz (2012) showed that Polish speakers at various levels of German would show transfer to English (a genderless language) when forced to determine the gender of a word in English. Transfer came from both languages and what would modulate it was the proficiency level the students had, thus showing that gender assignment was primed by the L1 and L2 and that the extent relied on linguistic competence. Nevertheless, the results of the study are questionable with regard to transfer since English is a genderless language. Interestingly, Spinner (2013) observed L1 speakers of English who learnt Swahili as a Foreign Language (FL). In this case, she had expected to find a facilitative effect for number over gender, since English does display the feature number but not gender. Unexpectedly, the results showed that number posed a larger issue than gender and she argues that might be due to either the knowledge of other languages such as Spanish facilitating acquisition, the different placement of the particles, or the way number is encoded in Swahili, which requires of an entirely new suffix that encodes gender and number simultaneously. In both studies, the results depended largely on proficiency as well, decreasing the number of errors as linguistic competence increased.

Proficiency is shown to be one of the most productive factors in predicting the acquisition of grammatical gender. In his 2012 study, Hopp found that proficiency, along with length of residence and length of exposure were the best predictors for behavioral data on gender. Hopp created two groups, one with consistent gender assignment and one without it. He concluded that the consistent group had assigned the correct label to each word and thus linked them to the “appropriate abstract gender class features” (p. 51) and that they could exploit that link rapidly in real-time processing. His hypothesis was that gender is an informative cue in L2 learners of high proficiency if they have correctly classified the nouns to the appropriate class. Hopp (2016) tested 34 L1 English speakers at the upper-intermediate level of German to see if they could predict gender facilitatively (online) and correctly assign and make nouns agree (offline). Performance both before and after training showed a correlation with proficiency. His results indicate that after training, learners were capable of using gender prediction facilitatively, although that only applied to those who made a consistent use of gender assignment. In another study Hopp & Lemmerth (2016) tested L1 Russian speakers with an upper-intermediate level of German against German native speakers and proficiency seemed again to account for most of the variability in the results. The design consisted in words that were either congruent (same gender in German and Russian) or incongruent (different genders in both languages). Participants were tested in production and comprehension⁶. Predictive gender processing was found in advanced learners regarding gender marking in articles and adjectives (even though marking works differently in both languages) and it was “irrespective of lexical congruency or syntactic overlap between L1 and L2 gender marking” (p. 20). The high-

⁶ Eye-tracking task in which they were shown 4 pictures and had to look at the target noun as soon as they identified it.

intermediate group only managed to predict the gender on adjectives, suggesting that “syntactically incongruent realizations of gender attenuate L2 prediction (p. 21); while on articles it depended upon gender congruency across the languages. Further evidence is provided by Dussias et al. (2013) who investigated a group of 18 English-speaking learners of Spanish, 15 Italian learners of Spanish and Spanish controls showing that Italian-Spanish bilinguals would use gender anticipatorily in an eye-tracking experiment, but only for feminine nouns —arguably such result might be due to the fact of the complete overlap of the feminine article *la* in both languages versus the masculine being realized as *el* (Spanish) and *il* or *lo* (Italian). Results supported again proficiency, since native-like gender processing arises among more advanced learners but it is missing at lower levels. Sagarra & Herschensohn (2012) further the case for proficiency along with animacy, age of acquisition (AoA), and working memory. They used a self-paced reading moving window followed by a GJT to test the performance of English-speaking beginner and intermediate learners of Spanish. Against Spinner’s (2013) study on Swahili, both sets of learners processed number more easily than gender and only intermediate learners showed qualitatively similar reactions to monolinguals, thus showing pointing at the relevance of proficiency; in addition, “larger working memory span positively correlated with sensitivity to gender agreement” (p. 620). Regarding animacy, learners showed longer reaction times (RTs) for animate over inanimate nouns, which the authors attribute to two words competing in the lexicon and having to ponder both genders for the noun instead of only having one option⁷. Montrul et al. (2008) also performed a study that looked into age of arrival (AoA), proficiency, word ending, and animacy, theirs however looked into

⁷ That is because animate nouns were also those with biological gender, e.g. *doctor*_{Masc} and *doctora*_{Fem} versus *mesa*_{Fem} and **mesa*_{Masc}.

offline measurements in production. They tested 72 English-speaking learners of Spanish, 69 Spanish heritage speakers and 22 monolingual Spanish NS. In the written task errors, accuracy was higher for masculine words over feminine in the first experiment and determiners over adjectives and canonical word endings in both experiments. Against their expectations, L2 learners outperformed heritage speakers in both settings and in the second experiment L2 learners performed better on feminine nouns than masculine. Such results challenge the assumption that AoA determines ultimate attainment, since L2 learners would outperform HSs who have learnt the language from birth; nevertheless, HSs tend to be exposed to their language orally rather than in written form and thus Montrul et al. (2008) carried out a similar spoken task. In this case, HSs outperformed L2 learners, proficiency was significant, accuracy for masculine nouns was again higher, as well as it was for animate nouns, consonant-ending nouns versus *-e* ending nouns, and masculine nouns ending *-a*. Together these results support that proficiency is key, but also that noun ending and animacy play a role. The role of the animacy of the degree of animacy in nouns is up for debate, as it seems to slow down processing, but seems facilitative when it comes to production. In summary, the research that take into account proficiency, seem to find clear results regarding its correlation with accuracy in gender assignment and agreement; nevertheless, other variables also play a role.

When discussing the effects of proficiency, Montrul et al. (2008) also addressed the effects of word ending, showing that it affected the success of learners in assigning the right gender, with canonical endings being facilitative. These effects have been researched more in L1 processing, where Martino et al. (2017) showed that processing was facilitated by canonical endings in Italian (*-o* masculine and *-a* feminine) while opaque endings (*-e*)

slowed the process. Not only that, but no advantage was found for masculine or feminine in opaque endings, pointing at a lack of activation of a gender default option. Caffarra et al.'s (2015) findings also support NSs processing of formal cues to detect gender during online processing. In their case they tested the words within sentences that violated gender agreement with nouns that were exceptions (*-a* masculine and *-o* feminine) and canonical, showing a P600 effect in both cases. In conclusion, research has looked mostly into NS processing of word ending and it seems a successful cue in processing (Martino et al., 2017; Caffarra et al., 2015) but also it seems to predict accuracy in L2 learners and HSs (Montrul et al., 2008).

The role of input has also been addressed, the main argument being that the success of children and NSs in their use of grammatical gender is having been exposed to (1) unsegmented aural input, especially from an early age, and (2) a variety of morphological realizations, adjectives, definite and indefinite articles, etc. The assumption in the first argument is that children use a distributional learning strategy linking gender-marked determiners and nouns. Siegelman & Arnon (2015) tested Hebrew⁸ speakers on an artificial language that had agreement with articles. They divided them in two conditions, a segmented-first or unsegmented-first with 22 participants in each. Words were arbitrarily assigned to two genders, without there being any facilitative phonological cues. Their results showed that participants in the segmented-first condition outperformed the other group significantly when selecting the correct article, but no effects were found with regard to choosing the correct noun labels. In addition, the study worked with two articles, *se* and *fo*, and there seem to have been some L1 effects, as there were higher accuracy rates with

⁸ Hebrew has grammatical gender, but it is not realized through articles.

the article *se*, which the authors argue might be due to the similarity with the Hebrew demonstrative article *se* ('this'). They did a follow-up experiment in which words were assigned to a gender based on semantic reasons and having modified the article *se* for *si* to make it less similar. The results did not hold, indicating that unsegmented learning only helped when nouns had had a gender arbitrarily assigned, not when there the article carried semantic information. Cornips & Hulk (2008) also looked at the acquisition of determiner phrases in different types of Dutch bilingual children and found that early age of onset along with lengthy and intensive input were the best predictors, while the influence of input quality was harder to prove. In addition, and in relation to the role of the L1, they conclude that their data supports facilitative effects only in those cases in which the L1's grammatical gender is similar, otherwise children having a grammatical system in their L1 or not did not affect their success. In conclusion, Cornips & Hulk (2008) argue for the need of extensive input for children to perform like monolingual NSs, which Rodina & Westergaard (2017) provide further evidence for. Their study tested Norwegian monolingual children and bilingual children who spoke Norwegian and Russian in two scenarios, two Russian parents and Russian-Norwegian parents. Their findings suggest that the three groups perform similarly in Norwegian, the language in which they are immersed, but that in Russian those with two Russian parents, and thus arguably exposed to more input, outperform those in Norwegian-Russian dyads; despite the fact that the grammatical gender system of Russian is simpler than that of Norwegian. Further evidence comes from looking at high schoolers in the US. Martínez-Gibson (2010) observed first and second-generation Spanish heritage speakers against L2 learners of Spanish. Her results show that the quantity of errors in use of gender and number is lower for the first-generation HSs

(2%), followed by second generation HSs (5.5%) and L2 learners (14%). Arguably, first-generation HSs are exposed to more Spanish than second-generation HSs, thus explaining the results. Together, these studies, show support for the role of input quantity in the acquisition of grammatical gender, even over quality of input (Cornips & Hulk, 2008).

Age of Acquisition is another factor that has been determinant, even children with a vocabulary of 500 words already show processing advantages using gender as a predictive cue, a characteristic found also in Spanish-speaking adults (Lew-Williams & Fernald, 2007). Thus, learnability of grammatical gender is available at an early age, even though other gender systems might require more time, like that of Dutch, in which children display unstable representations until the age of nine (Blom et al., 2008). In Blom et al. (2008) they found that L2 adult learners seem to rely on lexical representations based on input without extrapolating from it while children use grammar-based representations and thus manage to extrapolate their knowledge to the entire system. They base these findings on the observation of Dutch monolingual children and Moroccan children and adults with L2 Dutch. Each group was divided in two levels of proficiency to assure comparability. In doing so, all group overgeneralize the non-neuter gender with determiners in line with the argument that learners overgeneralize the default gender. Nevertheless, in adjectives, children only generalize one suffix, while adults use both adjectival forms mistakenly. In addition, proficiency played a role again. Lew-Williams & Fernald (2010) also showed that adult L2 learners of Spanish did not perform like L1 adults nor children. In their experiment they observed whether participants could use gender predictively in a looking-while-listening paradigm, L2 learners were not able to anticipate the word through gender cues. In order to eliminate the possibility of the degree of familiarity with the words, a second

experiment was carried out in which four novel words with canonical endings were taught along with definite articles. In that case, L1 and L2 adults both managed to use the cues predictively. However, when a new group was trained on the same words with indefinite articles and were exposed to the same paradigm with definite articles, L2 adults did not display the same ability, thus suggesting that adults rely on lexical representations that they cannot generalize to the entire grammatical system, namely they learn in chunks. In a similar study, Lew-Williams (2009) found the same lack of predictive processing, however, with words with biological gender (*la niña* ‘the girl’, *el niño* ‘the boy’), L2 learners performed similarly to NSs. In contrast, Foote (2011) looked into Spanish NSs, what she calls early bilinguals (heritage speakers), and late bilinguals (started learning Spanish in school after age 10). Her findings show that in a word-by-word sentence reading task there were no differences in the behavioral data nor in the processing of subject verb number agreement; in noun-adjective gender agreement the late bilinguals did not show the same sensitivity to gender agreement violations until the third time frame. In addition, the sensitivity effects were also lessened in all groups when items were not adjacent. In summary, previous research indicated that age of acquisition affects the learnability and processing of grammatical gender, showing that at least in eye-tracking paradigms and behavioral data L2 learners do not have the same advantages as L1 speakers; nevertheless, we find that errors decrease with proficiency (Cornips & Hulk, 2008) and that L2 learners show sensitivity to gender agreement violations similar to that of native and heritage speakers (Foote, 2011).

The last variable that I will be discussing is implicit and explicit learning, which is external to the learner. The literature addressing type of learning regarding grammatical

gender is meager, however, Morgan-Short et al. (2010) explored the results of teaching an artificial language implicitly and explicitly. Participants partook in three sessions, they got assessed after the first training and the third session. Based on their behavioral, data they were divided between low and high proficiency, at low proficiency agreement and assignment violations rendered an N400 effect in the group with implicit training. At higher proficiency levels noun-article agreement violations elicited N400 effects in explicit and implicit groups, while “noun-article agreement violations elicited P600s for both groups” (p. 155). The presence of N400s suggests that learners rely on lexical/semantic (declarative) processes in opposition to NSs, who rely on P600s (procedural memory), for noun-adjective agreement; while for article-noun agreement they rely on procedural memory. What this study suggests is that explicit teaching renders faster results that show in declarative memory but that implicit teaching facilitates storing in the procedural memory in the longer term. No further studies were found that addressed the issue within the domain of inquiry, however, implicit incidental learning will be discussed in more depth in the coming section.

CHAPTER III: TYPES OF INSTRUCTION AND LEARNING

In this section, studies addressing the success of implicit and explicit learning on the one hand and incidental and intentional learning on the other are reviewed. Studies showing support for implicit and explicit learning will first be presented, followed by those supporting incidental and intentional learning, and research supporting multimodality (namely both methods together). In order to aid the understanding of such variability in the field, the factors modulating learning effects will be addressed in the end.

Implicit and Explicit Learning

The issue of how humans learn has long been debated and a strong case has been made for implicit learning on the ground that that is how children learn. One of the main experts on explicit and implicit learning is DeKeyser, whose studies are the basis for most of this review. As pointed out by DeKeyser (2008) even at the early stages of the field in 1976, Arthur Reber identified the core of the issue being “lack of consciousness of the structure being learned” (p. 314). This is the base upon which DeKeyser operationalizes implicit learning as “learning without awareness of what is being learned” (p. 314), for example, learning the past tense by without having the structure specifically pointed out or taught. Explicit learning is, thus, the opposite, being aware of what is being learned, e.g. direct instruction of the past tense.

Reviewing previous research, DeKeyser concludes that results in implicit learning seem to come from concrete and contiguous elements in learning and that it does not necessarily reflect the acquisition of abstract features; he also adds that explicit instruction seems to render better results when it comes to knowledge of explicit rules (Ellis, 1993; as

cited in DeKeyser, 2008) or that, within explicit teaching, its deductive mode (regular instruction) renders better results than the inductive mode (learners' discovering the rules) as reported by Robinson (1996; as cited in DeKeyser, 2008). These studies, as informative as they are, were carried out with artificial language grammars, thus one is left wondering about their generalizability. In contrast, some studies seem to report that other factors such as awareness of the feature being learnt are at the core of the issue (Leow, 1998; Rosa & O'Neill, 1999; as cited in DeKeyser, 2008). Similar observations have been made about classroom studies with natural languages. Research addressing the dichotomy between implicit/explicit either makes questionable distinctions between the types of teaching (e.g. VanPatten & Oikkenon, 1996; as cited in DeKeyser, 2008) or shows the superiority of the explicit groups. However, as pointed out by DeKeyser, it is noteworthy that implicit knowledge is thought to be accessed when participants carry out tasks under time constraints, it might well be the case that these learners still accessed their explicit knowledge. Not only that, but that the training phases in these studies were not long enough for learners to acquire implicit knowledge, as "the development of explicit declarative knowledge into fully implicit, automatized procedural knowledge takes more time than any of these studies allowed for" (p. 326). He carries on arguing that it may be necessary to have a middle stage of explicit knowledge before it can become implicit, for which he gathers further evidence in Suzuki & DeKeyser (2017). There are other models as well, such as connectionist models, which defend that learning relies on statistical associations between input and output patterns and that knowledge formed that way is implicit. This view is supported by McDonough & Trofimovich (2016) although they themselves state

that previous research has required explicit instruction to find patterns in order to achieve said results.

Those who advocate for implicit learning tend to do so in tasks such as extensive reading and assume that the already known vocabulary and features is to aid the acquisition of the new aspects. This idea capitalizes on Stephen Krashen's Monitor Model (1982) and his theory of the $i + 1$, in which i stands for the current knowledge of the learner and "+ 1" represents the next step of learning to be acquired by exposure to input alone, so that learners infer the new features aided by their original knowledge. Later on, it has been hypothesized that noticing of the feature is necessary, albeit not necessarily understanding and being able to explain the rule or knowledge. Later, Long (1991) based focus on form on that observation. Through instruction, one can enhance necessary subsequent noticing.

Hulstijn (1995; as cited in DeKeyser, 2008) hypothesizes that a variety of factors determine the effectiveness of explicit learning, among others "complexity, universal grammar status, subset-superset relationships, scope and reliability of the rule, semantic redundancy and the possibility of item learning" (p. 332). Evidence of that is found in DeKeyser (1995), who found that clear-cut morphological rules would benefit from explicit-deductive learning of an artificial language while fuzzy prototypical rules, e.g. English past tense, would benefit from implicit-inductive instruction. More supportive evidence is provided by Williams (2005), who looked at the implicit, and arguably inductive, acquisition of \pm animate when learners were told to only learn articles that stood for close/far.⁹ The stimuli used two made-up particles for near and far that also

⁹ 66% of the participants were L1 English speakers and the language of the experiment was English with the only difference that these new articles were introduced.

encompassed animacy. Near was represented by *gi* (animate) and *ro* (inanimate) while far was represented by *ul* (animate) and *ne* (inanimate). They listened to sentences, indicated the meaning of the words, then repeated them out loud and formed a mental representation in the meantime. They were then tested with items that appeared during the treatment but also using the articles with other nouns. His results show not only that they had learned the \pm animate differentiation in the articles implicitly, but also that they were able to generalize. The results might have been partially affected by L1s spoken by participants, as there was a correlation between having a gendered L1 and performance. The ability to generalize did emerge even when they saw each noun associated with only one article, e.g. *gi (near) dog* appeared, but *ul (far) dog* did not. Another relevant factor DeKeyser pinpoints is proximity, as seemingly learning distant agreement seemed to be more taxing on memory (Ellis & Schmidt, 1997; as cited in DeKeyser, 2008).

Evidence of the effects of time on learning is further provided by Mirzaii (2012). He tested two groups, one in an implicit condition, 20-minute extensive learning sessions, and one in an explicit condition, teaching followed by deep-level cognitive processing activities. The results suggest that in the immediate post-test there were no significant differences and in the delayed post-test there also were not for the recall of collocations, however, vocabulary and meaning recall differed, showing an advantage for explicit learning. This would provide support against DeKeyser's suggestion that more time might lead to better results with implicit learning, unless the 18 20-minute sessions in the span of 5 weeks were not enough exposure.

Other studies have put evidence forward regarding the general heightened effectiveness of explicit teaching. In Tipurita & Jean (2014), after being taught two canonical endings

for each French gender explicitly, learners showed improvement in gender assignment even in a delayed post-test, not only that, but they were capable of generalizing regular assignment rules to novel words. Explicit attention has also shown positive effects in Wust & Roche (2015), who directed participants' attention to noun endings (form-focused) and found significant improvements.

It seems then that there is not a clearly superior across the board way of learning between explicit and implicit, but it rather depends on a myriad of factors. In addition, research trying to address the dichotomy is strongly biased by time constraints, which tend to favor explicit instruction; and fail to account for many factors such as awareness. Overall, DeKeyser's (2008) conclusion is that explicit instruction seems more favorable, but there have been imposed time constraints favoring it, and that when implicit learning has been successful there has been a focus on form, not only on meaning, which has facilitated the noticing of patterns. Nevertheless, as seen in the previous section, when balanced for time, even with lack of focus on form, implicit learning proves successful in acquisition of artificial grammar (Morgan-Short et al., 2010).

Intentional and Incidental Learning

Intentional and incidental learning represent two sides of the same coin. Intentional learning stands for a conscious effort to learn a specific target, the deliberate determination to commit words or structures to memory, while incidental learning represents an effortless "picking up" of forms without having that goal (Hulstijn, 2008). These, however, are sometimes mistaken with implicit and explicit learning. While it is correct to assume that incidental learning requires learning to also be implicit, intentional learning can be both. A learner who has been told to read a text for meaning but realizes that there is a pattern, e.g.

adverbs denoting future and verbs with a specific conjugated form co-occurring might actively start paying attention and learning said feature. Similarly, a student who is explicitly taught the mechanisms of the future tense and asked to memorize them would also be learning intentionally. However, a learner who might or might not have picked up on the co-occurrence of the adverbs and the morpheme denoting future but is not actively trying to memorize and retain said knowledge would be learning incidentally, without an intent of learning. In conclusion, as summarized by Hulstijn (2008; p. 361), “attention is deliberately directed to committing new information to memory in the case of intentional learning, whereas the involvement of attention is not deliberately geared toward an articulated learning goal in the case of incidental learning”. As Hulstijn (2008) points out, incidental learning has been addressed regarding the acquisition of grammatical features, while intentional learning has been neglected. Nevertheless, in the domain of vocabulary, several designs have aimed to address the different (dis)advantages of both.

Schmidt (1995; as cited in DeKeyser, 2008; p. 357-8) puts forward three definitions of incidental learning (1) “learning without the intent to learn”, (2) “learning of one stimulus aspect while paying attention to another stimulus aspect”, and (3) “the learning of formal features through a focus of attention on semantic features”. In the current study, incidental learning is operationalized by combining Hulstijn (2008) and Gass’s (1999; as cited in Hulstijn, 2008) definition, thus incidental learning happens as (1) a by-product of another task and (2) without any (cognizant) intention to learn the targeted feature.

As reported by Hulstijn (2008), several factors are believed to affect both types of learning, some of them are the density of new words (e.g. Holley, 1973), frequency (e.g. Rott, 1999), oral input (e.g. Ellis, 1995), glossing and/or inferencing (Cobb, 1997) and

dictionary use (Fischer, 1994), etc., which will be discussed more extensively below. All these factors have in common that their effects are believed to depend upon depth of processing (Craik & Lockhart, 1972), which has been more elaborated and formulated in the form of the Involvement Load Hypothesis (ILH) by Laufer and Hulstijn (2001). They theorize that involvement is composed of three factors (1) *need*, which reflects the motivation to understand a given word, (2) *search*, the cognitive component involvement that stands for how the meaning of the word are reached, and (3) *evaluation*, deciding whether the information applies to the given context. Based on these factors, the authors designed a point system to quantify the degree of involvement that has received empirical support (e.g. Ellis & He, 1999; Ansarin & Bayazidi, 2016). This system has now been contested by Hu & Nassaji (2016), who propose a more fine-grained approach by looking into motivation, noticing (attention and awareness of the word/feature), retrieval (absence or presence of retrieval as well as passivity thereof), generation (of the word), and retention. They found that such model had better explanatory power for the tasks under analysis. Despite the differences in the theories, it appears that cognitive involvement is an uncontested *sine qua non* condition for any type of learning.

As with explicit and implicit learning, results seem again mixed as far as the success of incidental and intentional learning to the extent of researchers challenging the existence of incidental learning as we know it. These researchers consider the term impracticable due to the reliance it has on the internal processes of learners and they propose alternatives such as “induced vocabulary salience”¹⁰, which capitalizes on external intervention (Bruton,

¹⁰ Bruton et al. (2011; p. 765) “where either specific vocabulary items in the linguistic discourse input, or the expected linguistic output, are intentionally made more salient or prominent for the language learner by other parties, especially teachers and/or materials”.

García López, and Esquiliche Mesa, 2011). However interesting, this approach does not allow for insight into learners' internal processes. Some caveats in research on incidental learning might be the cause for it, as Hulstijn (2008) wisely points out the two main issues are (1) the possibility that the targeted words or features are already known by our participants and (2) whether immediate post-tests suffice or delayed post-test should also be in place. To that regard, he suggests a twin approach in which both native speakers (NSs) and L2 learners are to be tested. The same design should be used on both, but NSs would be administered pseudo-words (low-ecological value but high reliability) and L2 learners real words (high ecological validity but low reliability), in this manner effects might be compensated. In the current study an alternative was chosen, to test learners who were not to participate in the study but belonged to the same level on their word knowledge under the assumption that learners at the same level would have a similar knowledge of the targeted words.

In conclusion and as reviewed by Choo, Lin & Pandian (2012), learning relies more on selective attention and elaborated processing, while intention does not play such a deterministic role. Considering the caveats in the discipline and the difficulty to access the minds of learners, it is easy to agree with the idea that it is hard to determine the effectiveness of intentional versus incidental learning, however, both notions are necessary. Albeit according to Presson, MacWhinney, & Tokowicz (2014) explicit grammar instruction improves performance efficiently as reported on meta-analyses (Norris & Ortega, 2000; Spada & Tomita, 2010), these results might not reflect the actual productiveness of explicit instruction, but rather the need to direct the learners' attention to the features to be learnt (Ellis, Loewen, & Erlam, 2006; Sanz & Morgan-Short, 2004).

However, I agree with Choo et al.'s (2012) conclusion that incidental and intentional learning should be looked at in combination with explicit and implicit learning. As mentioned, incidental/intentional literature has focused on the acquisition of vocabulary, as vocabulary is a tangential part of this project, a brief overview of the literature is provided followed by a more detailed description of the studies observing acquisition of grammatical features.

Incidental and Intentional Learning of Vocabulary

Most of the literature supporting either incidental or intentional learning provides evidence in favor of incidental learning. Shahpari, Shamshiri, & Rashidi (2014) found non-significant advantages for incidental learning, however, their sample was small (7 individuals per group) and there was no mention of control for previously known vocabulary. More robust findings are posited by Ahmad (2012), who does find significant differences in favor of the incidental group versus the intentional. A total of 50 words were tested, half of them taught incidentally (extensive reading) and half intentionally (crosswords, synonyms, antonyms and word substitution). The level of the words is not addressed beyond them being qualified as “easy”. And no distinction is shown among the degrees of cognitive involvement each task requires or how many words were taught by each methodology, thus not providing information with regard to many possible confounding factors.

Other studies exploring the success of incidental methods have precisely addressed how different methods and variables can affect the success of incidental learning. Teng (2016) shows that incidental learning is possible, and, in his sample, it depended upon frequency. He found that 14 exposures were necessary for recognition while 18 exposures were needed

for production. For the study they used pseudo-words that took the place of commonplace English words (e.g. *warm*) in a 300-word graded reader in which 36 were pseudo-words. Effects though decreased with time, going from about 10 words to 3 retained in memory after three months from last exposure for the higher frequency items (21-22 exposures). Strengthening the case for frequency in incidental learning, Pellicer-Sánchez (2015) showed that non-words were read significantly faster after 4 exposures and that regular reading rates appeared after 8 exposures. Longer reading times per word also aided recall later. Similar results are found in Godfroid, Ahn, Choi, Ballard, Cui, Johnston, Lee, Sarkar, & Yoon (2017), who approached the issue through an eye-tracking study in which they evaluated the effects of the time each participant spent on each item and the effects of their frequency. They used a mix of English NS and NNS with different backgrounds at a Midwestern university but also excluded participants who might speak Dari, Farsi, or any language with several cognates with such languages, as the stimuli consisted of English texts with target words substituted for Dari words. The experiment lasted for two days, in the end their results showed that time spent on a given target predicted the acquisition of meaning on a recall and a recognition post-test, while frequency alone helped predict the acquisition of form. Therefore, suggesting that frequency alone does not suffice, but rather needs to be supplemented with deliberate focus to acquire both meaning and form. However, Ansarin & Bayazidi (2016) argue that total encounters exhibited larger effects than elaboration of meaning; elaboration of meaning rendered better recall in multiple-choice and fill-in-the-blank than sentence-making exercises.

Along the lines of deliberate attention, Ellis & He (1999) explored the role of modified input and output with regard to incidental acquisition of word meaning. Based on Long's

(1981) Interaction Hypothesis and Krashen's (1985) Input Hypothesis they created three settings (1) premodified input, students were given directions to find an object in a picture of a room with some description of the object; (2) interactionally modified input, the teacher provided formulaic questions to request clarification and then learning could ask for more information to the teacher, the teacher would however always provide the same definition; and (3) negotiated output, the teacher gave them pictures of the objects and said the words for which they wrote the names down and then had to create directions to find the objects in the apartment. In support of Long's Interaction Hypothesis and the Involvement Load Hypothesis, negotiated input, which required more interaction and had higher cognitive demands, had the best results. Inasmuch as that proves the role of output, it remains to be explored whether learning was implicit and intentional, since the instructor provided the names for the objects in the pictures before the output task in the negotiated output treatment. The authors however sustain that the focus of the task was in learning how to give directions and that the learning of the words is a "possible natural outcome of performing the task" (p. 291).

Only one study, by Ahmadi (2017), was found clearly advocating for the effectiveness of intentional learning over incidental learning. Learners were exposed to intensive reading (implicit) that entailed being given definitions, meaning-focused extensive reading (they read for meaning) and form-focused extensive reading (participants were told to keep a log of new words they learnt and to think of a synonym and use it in a sentence). Results show more effectiveness for intensive reading after 8 sessions. The author himself points out at the fact that proficiency might have factored in, as insufficient competence might make

guessing inefficient. Thus, cognizant of the body of literature, results seem to indicate other factors having affected the outcome.

After having compared the success of both and concluding that there is no clear superior method; a breadth of studies supports multimodality, namely combining incidental and intentional methods. Zandieh & Jafarigohar (2012) used 184 learners of English at an English school who were divided in two treatments, those in the incidental learning group were not told there would be a test in the end and those in the intentional learning condition were. They used hyper-glosses, which show the target vocabulary providing a link to other contexts and uses of the words for learners to infer vocabulary. The immediate post-test showed an advantage for the explicit learners, while the incidental group showed better retention in the delayed post-test four weeks later. Bordag, Kirschenbaum, Rogahn, and Tschirner's (2017) worked with an incidental setting in which participants were not told to learn and both read passages while in the intentional setting they were given definitions and arguably their findings could be said to support intentional learning. Bordag et al. (2017) compared learning depending on orthotactic probability (possible orthographic combinations of letters). They would then read sentences in a self-paced reading paradigm. L1 speakers of German were tested in both conditions with pseudo-words and L2 learners of German were tested with unknown German words (DeKeyser's suggested twin approach). Low orthographic probability (the likelihood of clusters of letters appearing together in a language) favored the incidental learning in L1 speakers, but high orthographic probability favored intentional L2 learning. A possible explanation for these results is that L1 learners have a better grasp on the orthographical possibilities of the language, thus causing low orthographic probability words to be more salient; while L2

learners require more intention when learning and familiar patterns are more amenable to acquire, thus suggesting that the system is still forming and effects of the size of the mental lexicon. Teng (2015) also points out the decisive role of vocabulary size, in order for incidental learning to occur. In his study he compares a 100% incidental learning group who is exposed to extensive learning to a mixed methods group in which they read extensively but also are subject to pushed-output activities. Neither group was warned that there would be a test afterwards. Both methods resulted in gains, but the combined methodology rendered significantly better results; however, it could be argued that the ILH or frequency effects can account for the results, as both groups read the same books the same amount of times and the mixed methods group performed follow-up activities, while the extensive reading group did not get further “involved” nor did it get more exposure to the words. In addition, they used English pseudo-words, which diminishes the ecological validity in favor of guaranteeing words are unknown, but some words were dangerously similar to English words, e.g. “pitful”, as the authors themselves acknowledge. Xie, Zou, Lau, Wang, & Wong (2016) also provide evidence for the success of incidental learning over intentional, showing that intentional learning only surpasses incidental learning when the content of the text is of interest to the learner, thus one should bear in mind that results might be affected by the interest or relevance of the content presented to the participants.

We have seen that results change or can be challenged on the basis of (un)accounted factors. Huckin & Coady (1999) provide a good overview of reasons for the unresolved issue of the superiority of incidental or intentional learning, such as the lack of understanding of the mechanism, the vocabulary necessary for guessing in each case and proficiency, frequency of exposure, efficacy of guessing strategies, teaching of explicit

techniques for guessing, type of texts and interest thereof, and kind of input enhancement (e.g. glosses as facilitative mean in Varol & Erçetin, 2016). In addition, the current review has shown that other factors can also play a role, such as delivery mode (reading versus listening; e.g. Malone, 2018), consciousness/awareness (Bell & Collins, 2009), the type of task (e.g. Hu & Nassaji, 2016), the operationalization of intentional and incidental (see Schmidt's, 1995, definitions), the informativeness level of the context (e.g. Zandieh & Jafarigohar, 2012), orthotactics (e.g. Bordag et al., 2017), involvement (e.g. Hu & Nassaji, 2016), timing (how long is spent on the task/words; e.g. Mirzaii, 2012), working memory (e.g. Varol & Erçetin, 2016, which showed it mediated reading comprehension), chunk-strength (namely bigrams or trigrams controlled by frequency, e.g. Robinson, 2005¹¹), IQ (e.g. Robinson, 2005, where it negatively correlated with implicit learning), etc.

A great example of the interplay of several factors in incidental learning is provided by Malone (2018) in which he attempts to shed some light on the effects of frequency, aural enhancement and working memory. In his study he observes incidental learning of 32 words in four stories by L2 learners of English where he controlled for frequency (two versus four exposures), working memory, and aural enhancement. He found that at two exposures participants already performed above chance in form recognition and form-meaning tests and that there was significantly more learning at four exposures. He only found aural enhancement effects at two exposures, arguing that the lack of effects at four exposures reflected processing similar to familiar words in line with (Pellicer-Sánchez, 2016); nevertheless, this did not cancel frequency effects. In form-meaning connections

¹¹ One could argue that this operationalization falls within the definition of frequency, however, it is a ratio of both aspects, thus creating a new category.

aural enhancement effects are present across frequency levels, thus improving performance; as Malone (2018) puts it “simultaneous aural information appears to enhance learning of semantic information in new words” (p. 20). Regarding working memory, Malone found no or weak effects with regard to the groups without aural enhancement; however, and in line with previous research, he found a correlation between form-recognition and working memory in the aural enhancement groups but weaker relations with form-meaning. Thus, working memory seemed to affect recognition, but not semantic connections for new words.

In summary, the literature concerning the acquisition of vocabulary intentionally or incidentally is convoluted, and it seems to be so for a variety of reasons: (1) the issue of novel vocabulary and its ecological validity, using pseudo-words ensures lack of knowledge but reduces validity while real vocabulary in the L2 ensures reliability but risks learners already knowing the word, which leads us to (2), verifying the knowledge of the vocabulary before and after the test. A pre-test might give away the goal of the study and prime participants, but asking in *a posteriori* is not a good solution as they might not recall information properly. (3) Lack of comparability of the different studies (see Malone, 2018, p. 5-6, for an overview of differences in some studies). And (4) the influence of the aforementioned factors; it is thus necessary to control or account for such factors or at least take them into consideration when analyzing the outcome of vocabulary learning. In conclusion, it seems undeniable as Ramos (2015) points out, that learners do learn incidentally from informative contexts and that such phenomenon is enhanced in reading with the use of multimodal glosses.

Incidental Learning of Grammatical Features

As pointed out by Hulstijn (2008), the acquisition of grammatical features has been mainly explored from the perspective of incidental learning. His rationale for this is that incidental “in principle, can apply to abstract as well as to factual declarative knowledge, whereas intentional appears to be applicable to factual knowledge only” (p. 357). Thus, grammatical features requiring certain degree of abstraction are the only suitable candidates for incidental learning in this dichotomy. It is probably for this reason that most studies have focused on the effects of incidental learning. This section provides an overview of such studies and research that has attempted to compare both types of learning.

Looking into the L3 acquisition of a semi-artificial language called Japlish, Grey, Williams, & Rebuschat (2014) found that grammar could indeed be learnt incidentally, since many learners succeeded in acquiring the correct word-order rules, but fewer succeeded in case-marking. As per a debriefing questionnaire at the end of the treatment, rule knowledge seemed to play a role in improving performance, but lack of explicit knowledge of the rule did not impede improvement in performance. Their delayed post-test, two weeks later without further exposure to the stimuli, showed that effects not only held, but improved with time. This finding is in line with previous literature in which “improvement [was found] only after a delay (Ellis et al., 2006; Mackey, 1999; Morgan-Short & Bowden, 2006) or in general from immediate to delayed testing (Spada & Tomita, 2010)” (p. 637). Further evidence is put forward by Loewen (2005) after observing 17 hours of form-focused instruction. During the sessions he focused on focus-on-form episodes in which there was negotiation of meaning and accounted for whether learners incorporated the newly acquired forms to their repertoire or not. He found that learners

recalled 60% of the forms a day after the focus on form episode and 50% of the times two weeks later; thus showing the success of incidental learning. There are however two caveats regarding this study (1) it is incidental learning inasmuch as the form addressed was not the focus of the lesson, but learners did receive explicit information on it during the negotiation of meaning, and (2) given the naturalistic environment and spontaneity of the episodes and that each learners' test was tailored to the individual episodes they were involved in, it was impossible to assert whether the episodes reflected mistakes (occasional lapses in language use) or errors (actual blanks in linguistic knowledge). A debriefing questionnaire after the post-test might have shed some partial light onto the issue, although such measure is also of questionable reliability as it depends upon the learners' perception. Lee (2002) also found successful incidental learning, in this case of Spanish future tense morphology via extensive learning. In his study, he controlled for frequency of the forms (6, 10, or 16 exposures), the type of task (no orientation given, meaning, or form oriented) and the presence of cues (namely co-occurrence of time adverbs that hinted at future). He tested the results immediately after reading, two weeks, and a month later. Regarding frequency he found that six exposures sufficed but were not optimal, but that it had no effects in long-term retrieval. The presence of cues (adverbs) also improved the outcome and orientation to meaning outperformed orientation to form. His results indicate, however, that focus on form might hinder comprehension to some extent and vice versa and that learners could be -form+ meaning or +form-meaning, namely that they did not recognize the forms but selected the correct meanings and vice versa, thus casting doubt on the need for meaning before acquiring form. For this reason, he asks for future research to disentangle the factors for such differences. Finally, Bell & Collins (2009) provide

evidence of incidental learning in grammatical gender acquisition, much like Morgan et al. (2010) did for implicit learning¹². Their study, however, has more ecological validity with regard to choosing a natural language. In contrast to Williams' (2005) results showing the ability to generalize implicitly learnt features, Bell & Collins' results show that, despite improvement in article-noun pairs involved in the treatment, learners do not seem to be able to extend the pattern to novel nouns with the same noun endings. They argue that this is related to awareness, which none of their participants seemed to have. In line with Schmidt (1990) they suggest that "noticing a linguistic feature may lead to item learning, but in order to learn the system of the linguistic feature, it was necessary to understand the feature" (p. 287), namely its inner workings. Their results thus would disprove Schmidt's claims inasmuch as noticing being required for item learning, but might provide further evidence for the need of awareness to acquire a linguistic a feature. Despite the ecological validity of the design with a natural language, the lack of control over learners' proficiency and exposure make their results harder to interpret. In addition, they used canonical French endings, which is a straight-forward pattern, but learners already had knowledge of French, which possibly exposed them to other noun endings thus revealing the entire gender system and aiding with the feeling of lack of consistency and inexistence of patterns, while learners who are new to French might have been able to develop sensitivity if departing from a clean slate.

In another study by Bordag, Kirschenbaum, Opitz, Rogahn, & Tschirner (2016) that used again a twin approach with German, they researched the factors affecting the success

¹² Considering the several tests learners were subjected to, it cannot be known whether learning was intentional or incidental.

of incidental learning of transitivity and verb irregularities. They found that frequency affects the outcome, with three occurrences being necessary for learning, that the syntactic complexity of the text aided performance (tentatively, this might be due to higher involvement and could be argued posits evidence for the ILH), and that morphological markedness also favors acquisition among L2 speakers. However, there were some differences depending on the feature, while transitivity was equally learned by both groups, verb (ir)regularity showed differences. Their results suggest that “whereas the L2 learners can infer the (ir)regularity status of novel irregular verbs and perceive their regularization as a violation, native German speakers perceive any novel irregular form as violation and seem to classify all new verbs as regular irrespective of evidence that they receive through input” (p. 471). This shows that the learning paths are different in each group and that not only contextual cues, but also syntactic complexity and frequency play a role in incidental learning and that not all features work in the same manner.

Research looking into different learning modes also shows evidence of success in incidental learning of grammar. Denhovska & Serratrice (2017) tried to teach L1 English speakers who did not know Russian the rules of gender agreement of Russian. In their study, those in the explicit learning condition were explained the explicit rules, while those in the incidental group were shown the sentences transliterated into the Latin alphabet along with their English translation to be more engaging. Both treatments were administered for the same amount of time, 15 minutes. Only three participants reported awareness of the patterns in agreement and they all had higher accuracy in their answers. Overall accuracy in a Grammaticality Judgement Task (GJT) showed no differences between groups, although those in the explicit learning condition did outperform those in

the incidental learning condition. There were also no statistical differences in the performance of trained versus untrained items, thus indicating that participants could generalize the patterns, there was however a trend towards a slight worse performance in untrained items in the implicit learning group in the GJT, however in accuracy of production the explicit group also outperformed the implicit group. Reaction times in the GJT were also lower in the explicit learning group. These results suggest that incidental learning allows for receptive knowledge but not productive and that working memory aids in incidental learning, which would be in line with learning needing to be first receptive to then become productive; so, in line with DeKeyser (2008), implicit modes of learning might require more time to reach the same levels of success.

A really interesting effort was made by Tammenga-Helmantel, Arends, & Carrinus (2014), who gathered data from 981 Dutch students in lower secondary education learning German, English or Spanish. They compared the acquisition of degrees of comparison in the three languages in order to account for complexity, the Spanish system being considered of higher difficulty, and compared performance in explicit-deductive, explicit-inductive, implicit, and incidental groups. The groups were randomly assigned and the instructors were given a booklet on how to proceed. However, the results were inconclusive, and the researchers concluded that all treatments rendered improvement with respect to no instruction but that none of them was clearly superior to the rest and that complexity of the feature did not affect performance. These inconclusive results just further the case that it is really complex to operationalize different kinds of learning and control for the variety of factors that modulate learning.

Teaching of Grammatical Gender Systems

One might ask why I have not addressed methodologies for acquisition of the nouns' grammatical gender along with vocabulary. The reason is that the literature is just as inconclusive as it is for incidental learning, however, the methodologies proposed would be more difficult, less optimal, more expensive, and realistically not easily implemented even if learning were significant. Some of the methodologies that have been proposed are: color coding (visual enhancement; e.g. Arzt & Kost, 2016; inconclusive), images (visual enhancement; e.g. de Oliveira Santos, 2015; the image group outperformed the rest, however, it was passive recall and meaning became hindered when gender was to be acquired as well); mnemonic techniques such as the adaptation of the keyword method (visual and phonetic reliant enhancement; e.g. Desrochers, Gélinas, De Roy, & Wieland, 1989; mnemonic techniques worked, but learning gender brought down recall); type of feedback (e.g. Lyster, 2010; 5th grade immersion classes benefited more from prompts while university-level learners benefited equally from prompts and recasts); exposure to L2 input outside the class (e.g. Matteini, 2014; where external input played a larger role than classroom exposure); gendered actors (visual enhancement; e.g. Arzt & Kost, 2016; inconclusive); and prose, rhyme or melody (aural enhancement; Bebout & Belke, 2017; rhyme and melody combined worked best). Other methods such for visual input enhancement seem more sensible and realistic methods available to instructors and publishing houses, see Malone (2018). Lee & Huang (2008) for instance found that “learners exposed to enhanced texts outperformed learners who read unenhanced texts by a very small size effect” (p. 322) and, again that meaning could have been negatively impacted by attention being directed to other features, e.g. form, due to the enhancement,

thus providing even further evidence for the competition between processing form and meaning.

In summary, were it to be successful, enhanced input that facilitated incidental learning of grammatical gender and vocabulary could significantly take weight off the learners' effort to learn. Research so far has not been consistent in its results, thus calling for it to be furthered explored with more control over the aforementioned factors that affect learning. Another front to tackle is testing types of learning in more ecologically valid settings and languages. Given the scope of this project, not as many factors as desired were controlled for; however, it was my goal to provide an ecologically valid setting by conducting classroom research and using a natural language. By using incidental learning in the treatment the goal is to be as efficient in the teaching and to pose further evidence of the effects of incidental learning. In many previous cases, the researched linguistic feature was novel to learners, who had to fully acquire it from the beginning. This means that learners were completely new to the entire linguistic system and might have had to make a different effort than learners acquiring novel items with an already known binomial feature. Except for Bordag et al. (2016) and Bell & Collins (2009), to the best of my knowledge, no study seems to have explored incidental learning of an already known feature in novel items, which is what I proposed. Bell & Collins (2009) had little control over the level of French and exposure of their participants, which suggests that, were those factors to be controlled, they might have had positive findings. Bordag et al.'s (2016) had better control over confounding variables and they found that markedness of the features played a positive role in acquisition. Also, the corpus of research has concerned itself with assignment more than agreement and to trained items rather than learners' ability to generalize the patterns

learned. Not only that, but when looking into generalizability only novel items were tested, never observing whether the performance of already known items that were not part of the study changed. The combination of these factors makes this research a meaningful attempt at furthering our knowledge of the acquisition of grammatical gender.

CHAPTER IV: METHODOLOGY

This chapter describes the methodology used to assess the efficacy of four different pedagogical approaches for teaching the grammatical gender of novel words in Spanish. A brief introduction to the study is provided leading to the research questions addressed by this project and the hypotheses tested statistically. The next section describes the demographics of the study participants followed by the research design, the materials and procedures (subdivided in the different tests the participants were administered) and finally, how the data was coded.

As seen in the previous chapters, the literature regarding acquisition of grammatical gender provides inconclusive results, it rarely addresses consistency of use of the grammatical gender system, and the behavioral data rarely accounts for both gender assignment and agreement or concordance (from here on and for clarity, when referring to assignment and agreement together as a category, the term concordance is used). Research also shows that even when there is correct and consistent assignment, the performance in agreement does not follow. In addition, studies (i.e. Arzt & Kost, 2016) that have looked into teaching grammatical gender more effectively use methods that require color-coding and really specific, and at times unrealistic pictures which would, arguably, be costly and inefficient to implement by publishing houses and instructors. These studies have also been inconclusive, showing that more pedagogical approaches should be explored. The current research tries to take a step in the direction of finding a simpler more efficient way to teach grammatical gender in novel vocabulary.

Ideally, given the time constraints in a language course, implicit ways of learning would be the most efficient way to try to enhance learning without sacrificing time of instruction. With this goal in mind the four treatments for this study (1) enhanced article-adjective-noun group with bold font (ANAB); (2) article-adjective-noun group without enhancement (ANA); (3) enhanced with bold font article-noun (ANB); and (4) article-noun group without any enhancement (AN)– were created. This way it could be observed whether exposing learners to more gender-rich input would help them improve their accuracy in gender assignment and agreement. Additionally, half of the noun-adjective pairings were undone and those adjectives were repaired with other nouns to observe if effects only held for the original stimuli or if learners were able to generalize their knowledge. In this manner it could be observed whether learners learnt in chunks, namely learnt the article-noun combination or the article-noun-adjective combination as if they were an undividable fixated string instead of interiorizing the noun's gender.

To the best of my knowledge, no long-term effects have been observed regarding the acquisition of grammatical gender, for this reason, this study administered a delayed post-test two weeks later to observe the participants' performance with regard to the target forms, and known and unknown vocabulary. In the target forms one could observe whether effects remained the same, increased, or decreased over time; while the questionnaires with known and unknown vocabulary would speak to whether learners were capable of developing a certain degree of intuition regarding gender assignment and agreement that might resemble that of NSs (Lew-Williams & Fernald, 2009, 2010).

The questions that guided this research are the ones posited below. In order to test them each was transformed into a statistically testable hypothesis, as seen in the results' section in Chapter V. The predictions follow in the paragraph below:

Performance of the treatment by the features of the nouns in the immediate post-test:

- RQ1:** What differences are there between groups regarding the gender (masculine and feminine) of nouns combined with concordance (assignment and agreement)?
- RQ2:** What differences are there between groups regarding noun ending (canonical, non-canonical, and exceptions) combined with concordance (assignment and agreement)?
- RQ3:** What differences are there between accuracy of concordance (assignment and agreement) in new versus old pairings of nouns and adjectives?
- RQ4:** What differences in performance are there between concordance, noun ending, gender and treatment?

Performance of the treatment by gender and ending of the nouns in the delayed post-test:

- RQ5:** What differences are there between groups regarding the gender of nouns and the immediate and delayed post-test in assignment?
- RQ6:** What differences are there between groups regarding the gender of nouns and the immediate and delayed post-test in agreement?
- RQ7:** What differences are there between groups regarding noun ending and the immediate and delayed post-test in assignment?
- RQ8:** What differences are there between groups regarding noun ending and the immediate and delayed post-test in agreement?

Intentional learning of nouns and incidental learning of adjectives:

- RQ9:** What differences are there between the learning of adjectives and nouns in the ANAB, ANA and control groups?
- RQ10:** What differences are there between the learning of adjectives depending on their frequency in the ANAB, ANA and control groups?

My predictions were that the article-noun-adjective group enhanced with bold font would outperform the rest in concordance. The next best performing group was expected to be the article-noun-adjective group without enhancement since it is expected that more exposure to agreement through the presence of adjectives should reinforce learning of the grammatical gender, followed by the enhanced article-noun and the article-noun group without any enhancement. The same results from the immediate post-test were expected to hold in the delayed post-test, although performance was expected to be somewhat less accurate. Regarding acquisition of nouns and adjectives both article-noun groups were expected to underperform in the acquisition of adjectives in comparison to the article-noun-adjective groups. Within the acquisition of adjectives it was expected that the enhanced article-noun-adjective group would outperform the article-noun adjective group without enhancement and that adjectives that appeared with a higher frequency would present higher accuracy rates. It was expected that participants would use the masculine as a default, thus accuracy would be higher in masculine than in feminine nouns. Nouns with canonical endings would have higher accuracy across the board, followed by non-canonical endings and exceptions. Proficiency was expected to play the largest role in modulating accuracy.

Participants

In order to address these questions, college students were recruited from Spanish 201 classes (intermediate level I) at a mid-Atlantic public US university. Each of the five recruited classrooms were assigned to a different treatment group. In compensation for their participation the subjects were offered snacks and extra credit. Testing took place early December during their classroom hours, thus groups were random inasmuch as each

class was assigned a treatment randomly. In order to round up the group totals, more participants were recruited from other Spanish 201 classes at the same institution, those participants had different instructors. In total a fifth of the participants (28 total) was tested outside of the classroom in groups of 2-6 in a small office and received additional extra credit and snacks. The first group was exposed to sequences composed of an article, a noun, and an adjective that were enhanced with bold font, the second group was exposed to the same sequence without any enhancement, the third one was exposed to article noun sequences enhanced with bold font, the fourth group was exposed to the same condition without any enhancement, and the fifth group was the control group, which was exposed to no treatment whatsoever. Testing took place in two sessions, the treatment session and a delayed post-test two weeks after, in the classroom for the main groups and outside the classroom for the extra recruits from other classes.

I tested 141 adult subjects of which 13 had to be removed from the sample because they missed the treatment session or did not complete the test, thus leaving it at 128 eligible candidates (71 male and 57 female). Extreme outliers were found in the answer sheet given during treatment: one individual in the ANAB group, one in ANB, and two in AN. There were three extreme outliers for languages spoken in the ANB group. Regarding the variable practice hours there were 1 in the ANAB group, one in the ANA, four in the ANB, one in the AN, and one in the control group. The full linguistic background data is reported in Table 2. Despite the fact that it is worthwhile noticing that there were certain outliers that might have slanted the outcome, none of them were withdrawn from the sample as they were reflective of what one would find in a regular classroom and it was one of the goals to provide an ecologically valid and realistic setting. Regarding learners' performance in

the treatment forms there were fourteen extreme outliers, one in ANAB, five in ANA, four in ANB, three in AN, and one in the control group. In this case extreme outliers were withdrawn from the sample because the statistical analyses required it. After removing them, the final count of participants was 109. Of the final sample 20 were in Group 1, 24 to Group 2, 25 to Group 3, 22 to Group 4, and 18 to the control group. Mean proficiency was 6.7 in a scale of -60 to 60 in the Spanish version of the Lextale (Izura, Cuetos, & Brysbaert, 2014), a vocabulary knowledge test that has been proved to correlate with more comprehensive proficiency measures such as DELE (Diploma de Español Lengua Extranjera, the official examination for Spanish). The correlations, its predictive value of proficiency, and time efficiency are the reasons this test was employed. None of the subjects reported having a disability. Most participants were monolingual (77), 22 spoke an additional language, 9 spoke two languages, and an individual spoke up to 4 languages. No data was gathered regarding their proficiency in these languages. A 62.4 % of participants reported not practicing Spanish outside of the classroom and 37.6% did practice it, of which the majority reported practicing for an hour weekly. Subjects reported knowing 2.25% of the words in the experiment beforehand, versus 80.9% afterwards. The average accuracy (86.76%, range of 25-100%) on their answer sheet during treatment shows the participants were engaged since they correctly responded to the comprehension questions on the PowerPoint treatment.

Table 2
Language background and demographic data

Group	Age	Gender		Prof.	AS	FR	D	0	1	2	4	Other languages		
		M	F									Practice	Pre-K	Post-K
ANAB	21.19	12	9	8	86.34%	34.56%	0	9	5	6	1	.95	4.8%	85.7%
ANA	22.19	14	11	4.92	84.86%	22.49%	0	19	5	1	0	1.36	4%	60%

ANB	21.14	11	11	4.95	88.79%	47.94%	0	20	1	1	0	.36	0%	95.5%
AN	21.72	9	9	6.39	86.76%	50.37%	0	12	6	0	0	.61	0%	83.3%
C	20.45	14	6	10	NA	NA	0	14	5	1	0	1.3	NA	NA
Total	21.37	60	49	6.7	86.76%	37.24%	0	77	22	9	1	.91	2.25%	80.9%

Note. ANAB=article+noun+adjective in bold font, Group ANA=article+noun+adjective, ANB=adjective+noun in bold font, AN=article+noun, and C=control group. M=male, F=female. Prof.=proficiency score. AS=answer sheet score FR=form recognition test. D=disability. Practice=practice hours per week. Pre-K=percentage of participants who knew all the words, Post-K=percentage of participants who knew all the words. NA=Non-applicable.

Research Design

A group of 30 students taking a Spanish 201 class who were not going to participate in the experiment were given a preliminary survey on their word knowledge in order to make sure that the treatment words were novel to the subjects. These participants were also administered a set of words that they were assumed to know and another set of unknown word which were later used to observe whether the subjects in the study had developed an intuition similar to that of native speakers and were able to generalize the knowledge acquired through the treatment to words that did not belong to the training. In order to test the hypotheses, 30 nouns and 40 adjectives were selected from this preliminary survey.¹³ Following Montrul et al. (2008), half the nouns (15) were masculine and half were feminine. Each gender had 5 words that had a canonical ending, 5 with a non-canonical ending, and 5 were exceptions (see Table 3) for examples and Appendix A for the full list). Among the exceptions, 4 words had biological gender and were animate. These biological nouns are invariable, meaning that both the masculine and feminine counterparts have the same form, e.g. *el malabarista* ‘the (male) juggler’ and *la malabarista* ‘the (female) juggler’. However, even native speakers will not use gender predictively after being primed

¹³ The original design was targeting students taking 102 Spanish and had twice as many target words. After piloting the materials and seeing that participants were unable to carry out the tasks, it was decided to go a level higher and to cut down the number of words to half. By doing that, participants were able to fulfill the tasks in the second round of piloting.

with nouns with wrongfully assigned gender (Hopp, 2012). During the treatment, learners were exposed to *el malabarista* with a male in the picture, thus priming them for the masculine version. For this reason, it was presumed that wrongful assignment or agreement with these nouns would depend upon learners using noun ending as a cue rather than assuming the agent was of a gender they were not taught. Additionally, these nouns with biological gender were gender biased, since they were a soprano, a juggler, a model, and a priest, which arguably are jobs that tend to be associated with people of a specific gender performing them. All nouns were shown 5 times, learners first saw the noun alone, then the noun in two different sentences and then in two immediate comprehension questions that repeated the sentences (see the Materials and Procedure section for a more in detail explanation). In each of the two sentences each noun was paired with a different adjective, in the first sentence a variable adjective and an invariable adjective in the second one. All elements were read aloud, which meant that in total learners were exposed to each item 10 times total, 5 visually and 5 aurally (see Malone, 2018, about the effects of stimuli being presented in writing and aurally simultaneously).

Table 3
Type of nouns

	Masculine	Feminine
Canonical	teclado (5)	sonrisa (5)
Non-canonical	hogar (5)	cumbre (5)
Exceptions	fantasma (5)	soprano (5)

Note. In bold and italics the type of ending, in between parentheses the amount of items per category.

Of the adjectives, half (20) were variable and half were invariable (see Table 4) for examples and the Appendix A for the full list), of each category, each half had a frequency of 2 or 4 times in writing. Those adjectives with a frequency of 2 appeared matched to one

noun and adjectives with a frequency of 4 were matched to two different nouns. Again, as with the nouns, adjectives were read aloud, for this reason, and following Malone (2018), it is to be assumed students were exposed to adjectives 4 and 8 times respectively. To the best of my knowledge no study has addressed the use of invariable adjectives by learners, probably due to their invariability and lack of agreement; nevertheless, in my own experience as an instructor, learners do force agreement onto adjectives or deprive them of it, that is why I decided to include them in my design, to observe their use by learners.

Table 4
Type of adjectives

	Variable	Invariable
Frequency of 2	<i>caro/a</i> (10)	<i>enorme</i> (10)
Frequency of 4	<i>barato/a</i> (10)	<i>impresionante</i> (10)

Note. In bold and italics the type of ending, in between parentheses the amount of items per category.

As mentioned, nouns and adjectives were matched for the treatment. However, for the immediate and delayed post-test, half of the noun-adjective combinations were scrambled and rematched in order to create new pairs and assess the impact of chunk learning (e.g. Siegelman & Arnon, 2015). In the sentences created to frame the nouns and adjectives, the article-noun-adjective were out at the beginning of the sentences in order to be systematic. Also, articles, nouns, and adjectives were put adjacent to each other as often as possible, however, at times, there was a need for using *ser* or *estar* ‘to be’ in between the nouns and adjectives in order to make the sentences more natural. Trying to make the articles, nouns and adjectives adjacent was motivated by studies showing that when distance increases between the elements that need to agree processing slows down and target-like production and processing decreases (e.g. Foote, 2011).

Materials and procedure

All participants followed the same procedure. They were administered a proficiency and linguistic background questionnaire, followed by the treatment, a form recognition test, a translation test, a multiple-choice gender assignment and agreement test, and a debriefing questionnaire, all in their classrooms. The entire process took about 45 minutes and all instructions were given in English to ensure clarity. After the treatment, they were all given the form recognition test at the same time, after that, they were instructed to raise their hand as soon as they were done so the researcher could give them the next part. This prevented their following the wrong order or spending too much time reviewing their answers. Testing started at the beginning of the class to ensure that there would be enough time to complete the experiment. Participants were told that they would be learning vocabulary in order to hide the goal of the study and prevent them from learning intentionally instead of incidentally. All forms and tests can be found in Appendix C. Two weeks later the subjects were administered the same gender assignment and agreement test along with two other versions, one with known and one with unknown nouns and adjectives. The full PowerPoint presentations used in the treatment can be found in Appendix B. The following materials were used in the study and administered in this order:

1. Consent form.
2. Linguistic background formulary and LexTale proficiency test.
3. Treatment and answer sheet.
4. Form recognition test.
5. Multiple choice translation test.
6. Gender concordance test.

7. Debriefing form.

Consent form

The Institutional Review Board approved the experiments' protocol. Participants were told of the benefits and risks of the research and that they could withdraw at any point. All participants were given a copy of the form.

Lextale and linguistic background questionnaire

As mentioned, the proficiency measure used was the Spanish version of the Lextale (Izura et al., 2014). The test measures vocabulary size and has proven successful in distinguishing different levels of proficiency, even among native speakers. The test has proven successful and reliable across languages in its adaptations in German, Dutch, French, and English, its original. The test presents a list of nouns to learners and they need to discriminate if they are words or non-words in Spanish in this case. This proficiency is measured in a range between -60 and 60, thus negative scores are possible. The instructions were given in English and the linguistic background data required in the original test was modified to include other relevant variables such as learning disabilities, their native language, other languages participants spoke besides English, Spanish practice outside the classroom, and for how long they had received Spanish instruction. Self-rated proficiency was not added, as perceived proficiency was not deemed relevant and the Lextale was already providing a standardized proficiency measure to compare students. Collecting the students' grades from their instructors was considered, however, they all had different instructors and they might not grade equally, thus an independent measure that held them all to the same standard was deemed more appropriate. The full questionnaire, namely the Lextale and the linguistic background questions, can be found in Appendix C.

Treatment and answer sheet

The treatment consisted of a PowerPoint with 101 slides that lasted 20 minutes approximately that was projected on a screen for all the students while the audio played on speakers (the treatment slides for each treatment can be found in Appendix B. All participants were given the instructions on Figure 1. The presentation then continued to a trial item so participants could understand the dynamics of the experiment. Feedback on the trial item was provided to ensure comprehension of the dynamics but no feedback was given during the experiment to prevent participants from fixing their answers or getting delayed and distracted. After the trial item participants were asked if they understood what was expected of them and the experiment moved on. In the groups exposed to the article-noun-adjective combination each item was shown to participants for 12.5 seconds and they were exposed to the comprehension slide for 6.5 seconds. In the article and noun groups participants were exposed to each item for 11.5 seconds and had 6 seconds to provide an answer. Times were stipulated after seeing how much time participants needed during the pilot and averaging it out.

You will see a series of slides that will look like this:

- A new word will be introduced to you along with a picture and its translation.
- There will be two examples below the original accompanied by pictures and a translation, focus on learning the meaning of the word.
- Answer on your answer sheet.
- Answers won't appear on the screen. After both sentences you will see a new item



Figure 1. Instruction slide for the treatments.

The PowerPoint was voiced over by the author. The recording was made in a sound proof booth with a high-quality recording device. The items were blocked, namely all masculine canonical items were first, followed by feminine canonical, masculine non-canonical, feminine non-canonical, masculine exceptions, and feminine exceptions. Blocking is presumed to aid with learning as described by Bedout & Belke (2017) “A blocked presentation is necessary and sufficient for the acquisition of a complex grammatical paradigm” (p. 17).

There were four treatments, two in which they were exposed to sentences with an article, noun, and adjective combination and two in which they were exposed to article-noun combinations. Each of the two aforementioned combinations had a version with their items in bold font for enhancement and one without enhancement, see Table 5 for an example of all the versions. As previously stated, each noun was paired with a variable and with an invariable adjective. Some items were slightly modified across conditions to keep the sentences meaningful and as natural as possible as well as relevant to the pictures that appeared with them. All pictures were free of copyright.

Table 5
Sample stimuli

	<u>ANA</u>		<u>AN</u>	
	Variable	Invariable	1 st sentence	2 nd sentence
Bold	No hay basura en la calle pulcra	En la calle pobre la gente no tiene casas	No hay basura en la calle	La gente en la calle no tiene casas
Not-bold	No hay basura en la calle pulcra	En la calle pobre la gente no tiene casas	No hay basura en la calle	La gente en la calle no tiene casas

Note. Article+noun+adjective = ANA; Article+noun = AN

Once the treatment started, three pictures without text were displayed, then a noun appeared along with the translation below the upper-center picture, followed by a first sentence with a noun paired with a variable adjective in Spanish with the English translation below and then a second sentence with a noun paired with an invariable adjective also in Spanish with the translation below. Next, participants were asked “What picture goes with this sentence?”. Then they saw and heard the sentences and had to choose the correct picture out of a 4 picture display. The order in which the questions appeared on the picture display were randomized, namely they would not always be asked to find the picture for the sentence with the variable adjective first. The pictures on the display changed order as well. Both actions were to keep participants’ attention as suggested by subjects during the pilot. A sample of the setup is given in Figure 2 below.



La calle
The street



No hay basura en la calle pulcra
There is no trash on the neat street



En la calle pobre la gente no tiene
On the poor street people don't have

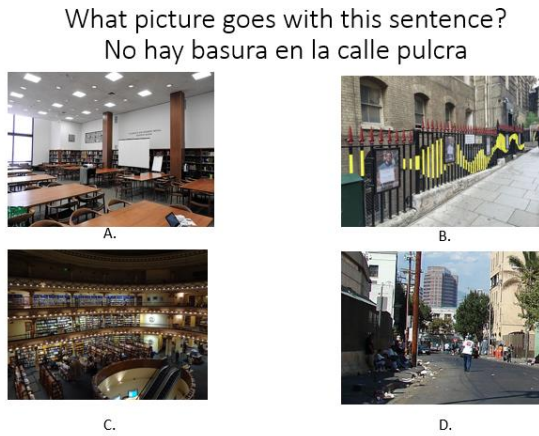
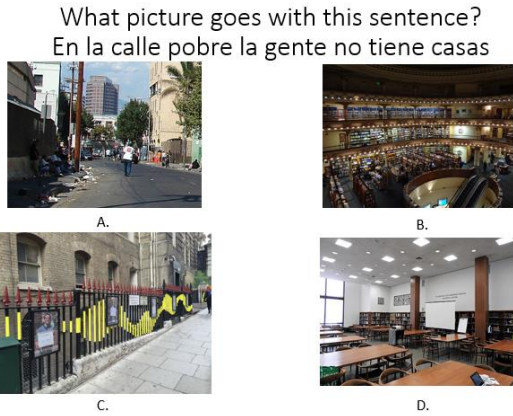


Figure 2. Samples from the treatment, exposure slide followed by the two comprehension slides.

Before the treatment started participants were given an answer sheet where they had to circle either A, B, C, or D. in order to answer to the questions asked in the prompt, e.g. Figure 2. The answer sheet was numbered, it had 60 questions that matched the questions they were being asked on the screen, which were reproduced so participants could easily catch up if they missed an item. Arguably, that might have exposed participants to the stimuli at least an additional time. For this reason, a question in the debriefing was added about whether they looked at the screen, the answer sheet or both, so such factor could be

accounted for in the analysis. In addition, the timing of the experiment made it unlikely that participants could go back and re-read the sentences several times.

The control group did not undergo any treatment whatsoever. They only were administered the other tests and forms, with the exception as well of the debriefing. This was to observe performance of participants who were not exposed to the treatment.

Form recognition test

Following the treatment, participants were given a form recognition test (as used by Malone, 2018; Hulstijn, Hollander, & Greidanus, 1996; Webb, 2007). This test is designed to check whether participants can recognize the vocabulary they were exposed to. For the groups in the noun-adjective combination they were given a list of 140 items, of which 70 had appeared in the treatment and 70 were unknown words. Subjects in the article-noun group were given a list of 60 words, of which 30 were nouns on the treatment and 30 were nouns unknown to them. In the pilot it was observed that subjects got confused and dubitative when the list contained other words they might know, for this reason efforts were made to use only words participants would not be familiar with. This part of the testing took about 1-2 minutes. Participants were not allowed to go back and look at their answer sheet.

As seen on Table 2, along with the rest of the participants' data, there is a notable difference in the form recognition between the performance of the article-noun-adjective groups (34.56% in the ANAB and 22.49% in the ANA group) and the article-noun groups (47.94% in the ANB and 50.37% in the AN group). Nevertheless, group differences did

not render statistical significance. The control group did not complete this test since they never underwent a treatment.

Translation test

The next test participants were administered was a translation test in order to make sure participants had learnt the meaning of the targeted nouns and adjectives. The control group was also administered this test, in order to observe what their knowledge of the experiments' nouns and adjectives was. Originally, participants were to be provided with the English sentences and translate them into Spanish, however, this task proved too challenging for subjects during the pilot. A translation from Spanish to English also proved too challenging, even when it required only the experiments' words to be actively translated instead of the whole sentence; for this reason, it was decided that a multiple-choice questionnaire would be more efficient. In addition, when designing the test, it was necessary to make it the most time efficient, which led to participants being provided a sentence in Spanish and having to choose between four English options. All the options they were given were semantically related in order to make the task more challenging and only the target words changed between questions to ensure that participants were not confused by other words. In the article-noun-adjective condition two options had the right noun and two options had the right adjective. For this reason, two answers showed learning of the noun and two questions revealed learning of the adjective; however, only one question showed that both had been learnt. In the article-noun group only one question was correct since they were not tested on their knowledge of the adjectives. Subjects were not given any feedback. Sample items are provided in Table 6.

Table 6

Sample translation items

Article+noun+adjective group	Article+noun group
Mi madre me dio el mensaje soso	Mi madre me dio el mensaje
a) My mom gave me the loving message	a) My mom gave me the package
b) <u>My mom gave me the dull message</u>	b) <u>My mom gave me the message</u>
c) My mom gave me the loving letter	c) My mom gave me the letter
d) My mom gave me the dull letter	d) My mom gave me the envelope
e) I don't know	e) I don't know

The test had a total of 40 questions in order to target the 30 nouns and 40 adjectives, 10 of the nouns repeated themselves. It was considered participants had learned the repeated nouns only if they chose the right translation in both questions where it appeared. The order of the questions was randomized and it was assured that those nouns that were repeated were far from one another.

Gender concordance test

Based on Montrul et al. (2008) and Alarcón (2011), participants had to decide on not only the article showing assignment, but also the adjective showing agreement. However, and as pointed out on the literature review, their design was embedded in a text and the articles, adjectives and pronouns participants had to choose from tended to have a cue, e.g. *la casa bonito/bonita*, where *la* is providing a gender cue besides the noun ending. In this study, participants were presented with 30 of the original sentences and 30 sentences in which the remaining nouns and adjectives were scrambled to form different pairs, still maintaining each noun paired with a variable and an invariable adjective. Balance between nouns was prioritized, namely 5 canonical, 5 non-canonical, and 5 exception nouns were re-paired with 10 invariable and 10 variable different adjectives. One can argue that in this manner too much attention was drawn towards gender assignment and, on that regard, a handful of participants were observed first selecting all the articles in the test and then all

the adjectives. However, in the debriefing, no participant reported having noticed any relevant patterns and guessed a variety of possibilities as the purpose of the study, none of which capitalized on grammatical gender. The four groups, including the control, took this test in order to observe differences in gender assignment and agreement across groups. There was a total of 60 items, with each of the 30 nouns from the design appearing twice, once with the adjective they were originally paired with in the treatment and once with a different adjective from the treatment in order to assess consistency in gender assignment and agreement, namely whether participants assigned the same gender to the noun in both occasions.

In order to not give cues to participants, an invariable version of the adjectives was created for variable adjectives, they ended either in consonant or *-e*, e.g. *blanco*_{masc} shows agreement as it can be *blanca*_{fem}, but the invariable adjective created was *blanc*_{inv}. A version showing agreement for the masculine and feminine genders was created as well for invariable adjectives, e.g. *verde*_{inv} also offered the options of *verde*_{masc} and *verda*_{fem}, see examples of the stimuli in Table 7.

Table 7
Sample items for the multiple-choice gender assignment and agreement

Type of adjective	Stimuli
Variable adjective	No hay basura en el / <u>la</u> calle peligroso / <u>peligrosa</u> / peligros
Invariable adjective	En el / <u>la</u> calle pobro / pobra / <u>pobre</u> la gente no tiene casas

Two additional versions of this test were also created and administered for known and unknown vocabulary that had not been part of the treatment as described in the design, thus resulting in three different questionnaires. All items were randomized so that all words belonging to a category would not be together and so that each occurrence of the word was

away from each other. During the pilot some learners realized words were repeated, but not always and just 2-5 words. In addition, they seemed confused by the original design, so besides the already existing sample items the target nouns were enhanced with bold font and additional spaces were added between the articles and the nouns and the nouns and the adjectives to make the nouns more salient. During the treatment session subjects were only given the test with treatment words due to time constraints and to avoid tiredness or boredom effects. Two weeks later the researcher came back to the classroom and administered a delayed post-test that was the exact same as the one provided during the treatment along with the tests with known and unknown words in order to observe long-term effects, whether a capacity to generalize the rest of the gender system had emerged and whether generalizability applied to known and/or unknown words. The Tests were counterbalanced, namely each third of the group started with a different test, e.g. (1) known, (2) treatment, and (3) unknown; while another third did (2), (3), and (1); and (3), (1), (2). In doing so the goal was to balance out tiredness so its effects would not always be present on the same test. This test is also found in Appendix C.

Debriefing

Lastly, all groups except the control group were administered a debriefing questionnaire, which can be found in Appendix C. The main goal of said questionnaire was to observe whether any participants had realized the target of the experiment and was intentionally paying attention to grammatical gender or if, alternatively, they had learnt incidentally. No participant became aware of the actual goal of the experiment. Knowing of the effects of frequency in vocabulary learning (Malone, 2018) and arguably the necessity of frequency in implicit learning (e.g. Morgan-Short et al., 2010), participants

were also asked whether they read the screen during treatment or both on the screen and the answer sheet. In general, the split was quite even, 45.5% looking at only either one versus 54.5% looking at both, see Table 8 for more details.

Table 8
Distribution of learners looking at the screen and/or answer sheet

Group	Screen or sheet	Both
Group 1	40%	60%
Group 2	44.4%	55.6%
Group 3	52%	48%
Group 4	45.8%	54.2%
Average	45.4%	54.5%

Data coding

The accuracy with regard to the answer sheet was calculated by adding up the total of correct answers and calculating the percentage. With regard to the form recognition test all the items learners identified were counted. Then those identified items that were not on the test were multiplied by two and the result was subtracted from the first total.

The data were originally coded numerically. Articles were coded as 1-*el* and 2-*la* while adjectives were labelled as 1-masculine, 2-femenine, and 3-invariable. Having done that, an equation on Excel interpreted the figures against the answer key to determine which one of the possible 11 cases matched the combination being evaluated (the full coding sheet can be found in Appendix D). Invariable adjectives forced to agree and variable adjectives turned into invariable and thus not showing agreement were considered correct when accounting for agreement. After doing so, each noun had a score calculated for assignment and one for agreement. If the feature under analysis (assignment or agreement) was consistent, the noun would receive a score of 1, otherwise the score was 0. Once that was done, accuracy percentages were calculated for each category under observation, e.g.

masculine and feminine words or canonical, non-canonical, and exemptions. Some of the data were missing because participants left an entire item blank or did not select either the article or adjective. In those cases, percentages were adjusted to exclude said items, e.g. for masculine and feminine the cells were added, multiplied by 100 and divided by the total amount of items (15), however, if two items in that category were missing it would have been divided by 13: $\Sigma_{\text{cells}} * 100 / 13$.

CHAPTER V: STATISTICAL RESULTS

The data were analyzed with a mixed-design ANOVA with the IBM Statistical Package for the Social Sciences (SPSS). The between-subjects independent variable was always the Treatment group, which had five levels (ANAB, ANA, ANB, AN, and Control). Within-group dependent variables are described by each hypothesis. Given the differences in the total amount of people per group conservative measurements were used.

Given the amount of data to be reported, non-significant interactions are not going to be elaborated on, thus statistical information is only provided for significant ones. In addition, in order to explore the direction of effects, Bonferroni corrections were performed on the interactions. If an interaction did not prove significant, data are not reported for it, however, the analyses were still performed and full tables along with their descriptions are provided in Appendix D. The tables in this section display only significant interactions.

Each of the sections addressed a different question. Within each question the first paragraph addresses the analysis used, whether the assumptions were fulfilled, and the procedures followed according to whether the pertinent assumptions were violated or not. Following, the hypothesis and the analysis are presented with the significant differences and interactions that were found. In the next paragraph the tendencies observed in the raw means are described. If significant interactions were found, that part of the subsection addresses the results of Bonferroni's correction, which indicate the direction of the significant effects in the interactions described in the second paragraph. Lastly, graphs are provided in order to give a more visual approach to the data.

RQ1: What differences are there between groups regarding the gender of nouns combined with concordance?

A three-way 5 (Treatment) x 2 (Concordance, agreement or assignment) x 2 (Gender, masculine or feminine) mixed ANOVA with repeated measures on the accuracy scores for Concordance and Gender was performed. Neither the main effect of Concordance, Gender, or the combination of both did not reach significance, thus showing that the assumption of sphericity was not violated since the significance value was greater than .05. According to Levene's test Normality was also not violated for either of them. The following hypothesis was tested:

*H₀: Performance across groups regarding gender does not differ
in agreement and assignment*

Significant differences were found for Treatment, thus indicating that groups behaved differently, $F(4, 104) = 3.6, p = .009, r = .4626.57$, and disproving the hypothesis that there are no differences between groups regarding gender and concordance (Table 9). There were also significant differences regarding Concordance¹⁴ (Table 10), indicating that learners showed different degrees of accuracy regarding assignment and agreement, $F(1, 104) = 58.17, p = .000, r = .2200.91$. However, no significant differences were found in its interaction with Treatment, thus showing that the effect of the different treatments on the performance was not different in Agreement compared to Assignment. There was a significant difference in the performance when it came to gender (Table 10), thus showing that accuracy rates significantly differed between masculine and feminine nouns, $F(1, 1) = 89.79, p = .000, r = .24661.15$. Just like with Concordance there were no significant

¹⁴ Variables used for statistical analyses are capitalized, while non-capitalized terms are used to refer to the general concept that is being discussed. In this case, Concordance is referring to the variable when the analysis was run. This applies throughout the chapter.

differences when taking into account the treatment. The interaction between Concordance and Gender did prove significant (Table 11), thus suggesting that accuracy differed significantly between Assignment and Agreement and Masculine and Feminine nouns, $F(1, 1) = 6.82, p = .010, r = 91.39$. When Treatment was factored into the interaction of Concordance and Gender, no significant differences were found, thus showing that the treatment group was not the source of differences in performance.

Table 9
Differences between treatment groups

Treatment	Mean difference	<i>p</i> value	Direction of effect
ANA	8.14	.031*	C>ANA
AN	10.20	.007**	C>AN

Note. C=Control; ANAB=Article, Noun, and Adjective in bold font; ANA=Article, Noun, and Adjective; ANB=Article and Noun in bold font; AN=Article and Noun. * $p < .05$, ** $p < .01$, *** $p < .001$

Table 10
Direction of effects for Concordance and Gender separately

Variables		Mean difference	<i>p</i> value	Direction of effect
Assignment	Agreement	4.53	.000***	Assignment>Agreement
Masculine	Feminine	15.15	.000***	Masculine>Feminine

Note. * $p < .05$, ** $p < .01$, *** $p < .001$

Table 11
Interaction between Concordance Gender*

Gender	Concordance	Difference	<i>p</i> value	Direction of effect
Masculine	Assignment Agreement	3.6	.000***	Assignment>Agreement
Feminine	Assignment Agreement	5.45	.000***	Assignment>Agreement

Note. * $p < .05$, ** $p < .01$, *** $p < .001$

The overall means showed a tendency to the control group being more accurate, followed by ANAB, ANB, ANA, and AN. However, the only significant interactions were found between the Control group and ANA and the control group and AN, thus suggesting

that without the enhancement learners performed significantly worse than control participants, although not significantly worse than those in an enhanced condition. The control group still outperformed the rest. Assignment showed significantly better scores than agreement, suggesting that assignment was easier to the learners. Accuracy scores were also significantly higher for masculine nouns than feminine, suggesting that the participants might have been resorting to a default masculine. These tendencies held even when combined. Figures 3 and 4 provide an overview of the means by group.

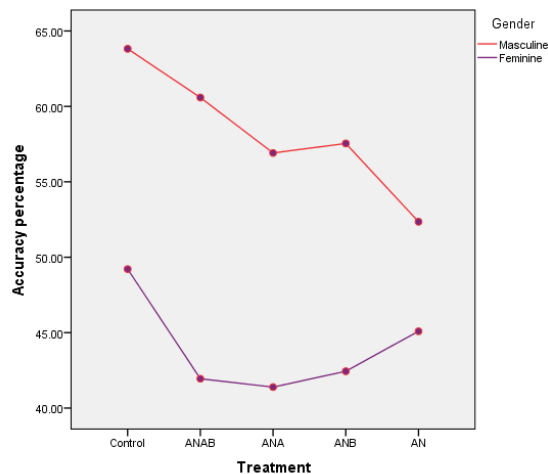


Figure 3. Mixed ANOVA results for assignment regarding gender

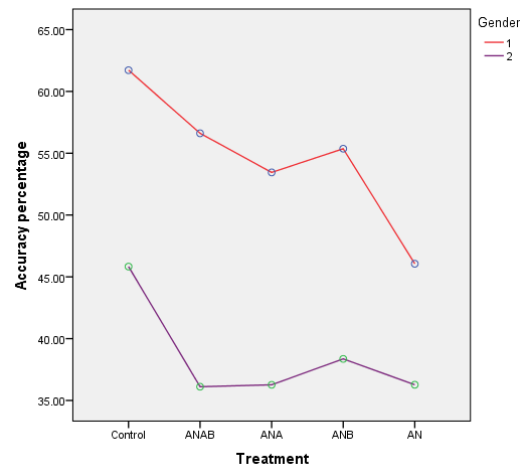


Figure 4. Mixed ANOVA results for agreement regarding gender

RQ2: What differences are there between groups regarding noun ending combined with concordance?

A three-way 5 (Treatment) x 2 (Concordance, agreement and assignment) x 3 (Noun Ending, canonical, non-canonical, and exceptions) mixed ANOVA with repeated measures on the accuracy scores for Concordance and Noun Ending was performed. The main effect of Concordance does not significantly violate the sphericity assumption; however, it is indeed violated by ending, $\chi^2(2)$, $p = .028$, since the value for the Greenhouse-Geisser correction was higher than $>.075$, therefore using the Huyn-Feldt correction ($\epsilon = .937$). The

interaction between Concordance and Noun Ending also violated the assumption of sphericity, $\chi^2(2)$, $p = .012$, $\epsilon = .924$, so the Huynh-Feldt correction was used. Canonical nouns also violated the assumption of normality both in assignment and agreement. According to Levene's test, canonical nouns violated it in assignment, $F(4, 104) = 3.12$, $p = .018$; and in agreement, $F(4, 104) = 3.03$, $p = .021$. This means we should approach the results with caution. The following hypothesis was tested:

*H0: Performance across groups regarding noun ending does not differ
in agreement and assignment*

Significant differences were found for Treatment (Table 12), thus indicating that groups behaved differently, $F(4, 104) = 3.42$, $p = .011$, $r = 16.78.07$, and disproving the hypothesis that there are no differences between groups regarding noun ending and concordance. Concordance displayed significant differences, with assignment rendering more accuracy than agreement, $F(1, 104) = 38.89$, $p = .000$, $r = 2324.2$. Its interaction with treatment did not prove significant, indicating that treatment did not explain the differences observed in concordance. Noun ending also showed significant differences (Table 13), thus reflecting that performance change depending on the noun ending, $F(1.98, 104) = 852.28$, $p = .000$, $r = 697618.4$. Its interaction with treatment proved significant as well (Table 14), meaning that treatment accounted for differences in performance regarding ending, $F(7.93, 104) = 2.34$, $p = .002$, $r = 10605.34$. Concordance also showed significant differences when noun ending was taken into account (Table 16), suggesting that behavior in nouns depending on ending could partially be accounted by concordance, $F(1.95, 104) = 15.17$, $p = .000$, $r = 670.16$. Finally, the interaction between concordance, ending and treatment was also significant (see Table 17 under interactions), $F(7.81, 104) = 2.29$, $p = .024$, $r = 403.75$.

Table 12

Differences between treatment groups

Treatment		Mean difference	<i>p</i> value	Direction of effect
C	ANA	8.50	.023*	C>ANA
	AN	9.65	.014*	C>AN

Note. C=Control; ANAB=Article, Noun, and Adjective in bold font; ANA=Article, Noun, and Adjective; ANB=Article and Noun in bold font; AN=Article and Noun. **p* < .05, ***p* < .01, ****p* < .001

Table 13

Direction of effects among Noun Endings

Endings		Mean difference	<i>p</i> value	Direction of effect
C	NC	39.13	.000***	C>NC
	E	80.57	.000***	C>E
NC	E	41.43	.000***	NC>E

Note. Noun endings are coded as C=Canonical, NC=Non-Canonical, E=Exception. **p* < .05, ***p* < .01, ****p* < .001

Table 14

*Interaction between Ending*Treatment*

Ending	Treatment	Mean difference	<i>p</i> value	Direction of effect
Canonical	AN	10.67	.025*	ANAB>AN
Non-Canonical	C ANAB	18.26	.000***	C>ANAB
	ANA	20.07	.000***	C>ANA
	ANB	16.15	.002**	C>ANB
	AN	23.70	.000***	C>AN

Note. C=Control; ANAB=Article, Noun, and Adjective in bold font; ANA=Article, Noun, and Adjective; ANB=Article and Noun in bold font; AN=Article and Noun. **p* < .05, ***p* < .01, ****p* < .001

The overall means showed a tendency to the control group being more accurate in this case as well, followed by ANAB, ANB, ANA, and AN. However, just like before, the only significant interactions were found between the Control group and ANA and the control group and AN, thus suggesting that without the enhancement learners performed significantly worse than control participants, although not significantly worse than those in an enhanced condition. Again, assignment showed significantly better scores than

agreement, suggesting that assignment was easier for the learners. Accuracy scores were also significantly higher for canonical endings, followed by non-canonical endings and exceptional endings, suggesting that participants preferred the unmarked feature. These tendencies held even when combined, except for the hierarchy of the groups changing when controlling by noun ending, ANAB>C>ANB>ANA>AN, showing that the control group was slightly outperformed by the ANAB group. Figures 5 and 6 provide an overview of the means by group.

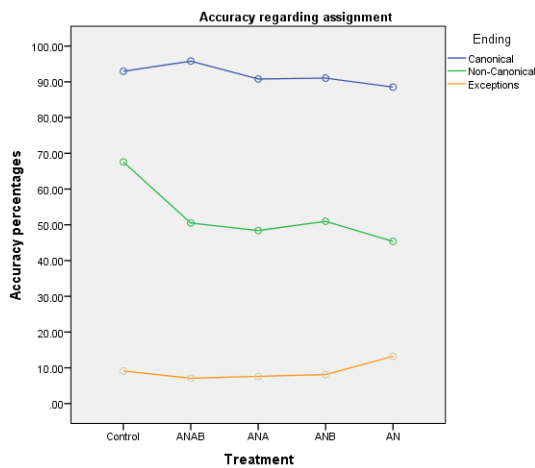


Figure 5. Mixed ANOVA results regarding assignment in noun ending

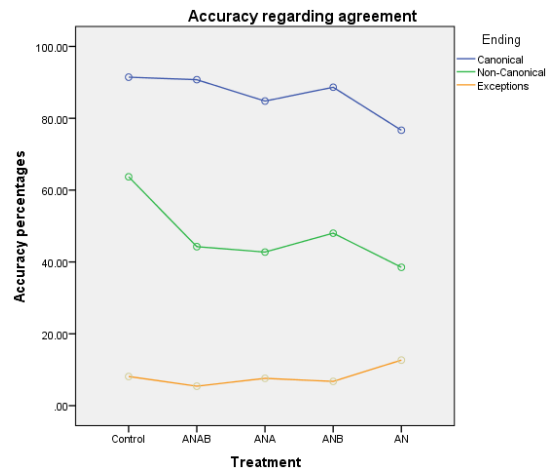


Figure 6. Mixed ANOVA results regarding agreement in noun ending

Interactions

When controlling by noun ending we do not find any significant interactions, except for the ANAB group outperforming the AN group within the nouns with a canonical ending and the control group outperforming the ANAB and ANA groups in non-canonical nouns, see Table 15 for significant interactions.

Table 15

*Interaction between Ending*Treatment*

Ending	Treatment	Mean difference	<i>p</i> value	Direction of effect
Canonical	AN	10.67	.025*	ANAB>AN
Non-Canonical	C ANAB	18.26	.000***	C>ANAB
	ANA	20.07	.000***	C>ANA
	ANB	16.15	.002**	C>ANB
	AN	23.70	.000***	C>AN

Note. C=Control; ANAB=Article, Noun, and Adjective in bold font; ANA=Article, Noun, and Adjective; ANB=Article and Noun in bold font; AN=Article and Noun. **p* < .05, ***p* < .01, ****p* < .001

Performance in assignment had significantly higher accuracy scores than agreement within each noun ending category (see Table 16).

Table 16

*Interaction between Ending*Concordance*

Ending	Concordance		Mean difference	<i>p</i> value	Direction of effect
Canonical	Asgn	Agr	5.36	.000***	Asgn>Agr
Non-Canonical	Asgn	Agr	5.12	.000***	Asgn>Agr
Exception	Asgn	Agr	.92	.008**	Asgn>Agr

Note. Asgn=Assignment, Agr=Agreement. **p* < .05, ***p* < .01, ****p* < .001

On Table 17 there is a full report of the interaction between concordance, noun ending and treatment. The differences between treatment groups, though, were altogether non-significant, the only significant interactions being found when looking into non-canonical nouns both in assignment and agreement, in that case the control group significantly outperformed all treatment groups. The only other significant interactions were found in the interaction between agreement and canonical nouns, in which the AN group significantly underperformed with regards to the control and the ANAB groups, thus suggesting that only with the consistent effect of enhancement and adjective did significant differences arise between groups.

Table 17
Interaction between Concordance Ending*Treatment*

Concordance	Ending	Treatment	Mean diff.	<i>p</i> value	Direction of effect	
Assignment	Non-Canonical	C	ANAB	17.07	.001**	C>ANAB
			ANA	19.18	.000***	C>ANA
			ANB	16.61	.002**	C>ANB
			AN	22.21	.000***	C>AN
Agreement	Canonical	C	AN	14.78	.015*	C>AN
		ANAB	AN	14.07	.016*	ANAB>AN
		ANB	AN	11.97	.044*	ANB>AN
	Non-Canonical	C	ANAB	15.45	.001**	C>ANAB
			ANA	20.95	.000***	C>ANA
			ANB	15.61	.006**	C>ANB
			AN	25.19	.000***	C>AN

Note. C=Control; ANAB=Article, Noun, and Adjective in bold font; ANA=Article, Noun, and Adjective; ANB=Article and Noun in bold font; AN=Article and Noun. **p* < .05, ***p* < .01, ****p* < .001

RQ3: What differences are there between accuracy of concordance in new versus old pairings of nouns and adjectives?

A three-way 5 (Treatment) x 2 (Concordance, agreement and assignment) x 2 (Pairing, new versus old) mixed ANOVA with repeated measures on the accuracy scores for Concordance and Pairing was performed. None of the variables violates the assumption of sphericity according to Mauchly's test. The assumption of normality is also fulfilled regarding assignment, however, regarding agreement the assumption is violated in new pairings, $F(4, 104) = 8.54, p = .000$, and old pairings, $F(4, 104) = 3.10, p = .019$. Given that some of the assumptions were violated, the conservative Greenhouse-Geisser correction is used to report on significance. The following hypothesis was tested:

H₀: Performance across groups regarding old and new pairings do not differ in agreement and assignment

Significant differences were found for Treatment (Table 18), thus indicating that groups behaved differently, $F(4, 104) = 3.39, p = .012, r = 1748.99$, and disproving the hypothesis that there are no differences between groups regarding pairing and concordance. Concordance displayed significant differences, with assignment having higher rates of accuracy than agreement, $F(1, 104) = 1929.44, p = .000, r = 160652.71$. Treatment's interaction with Concordance did not prove significant, indicating that it did not explain the differences observed in concordance. Pairing did not display any significance and neither did it in combination with treatment, concordance, nor both simultaneously, thus these interactions were not further pursued.

Table 18
Differences between treatment groups

Treatment		Mean difference	<i>p</i> value	Direction of effect
C	AN	6.50	.006**	C>AN

Note. C=Control; ANAB=Article, Noun, and Adjective in bold font; ANA=Article, Noun, and Adjective; ANB=Article and Noun in bold font; AN=Article and Noun. * $p < .05$, ** $p < .01$, *** $p < .001$

The overall means suggest that the best accuracy is found in the control group, followed by ANAB, ANB, ANA, and AN; in this case, the control group only significantly outperformed the AN group, suggesting that differences across groups were due to chance, except for the difference between the control and AN groups. There was a slightly better performance in old items over new although they did not differ significantly. Accuracy scores for concordance show that there was more accuracy in agreement than assignment, in contrast to what was found in the previous sections. Figures 7 and 8 provide an overview of the means by group.

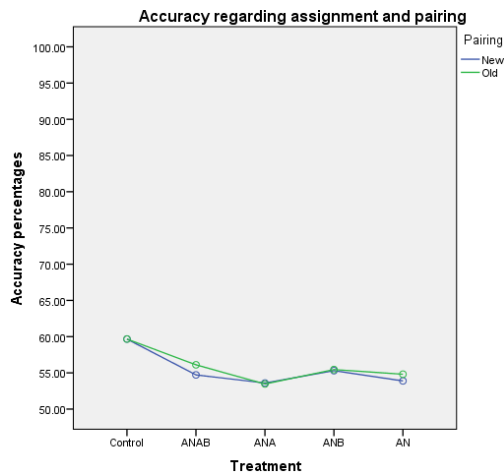


Figure 7. Mixed ANOVA results regarding assignment divided by pairing

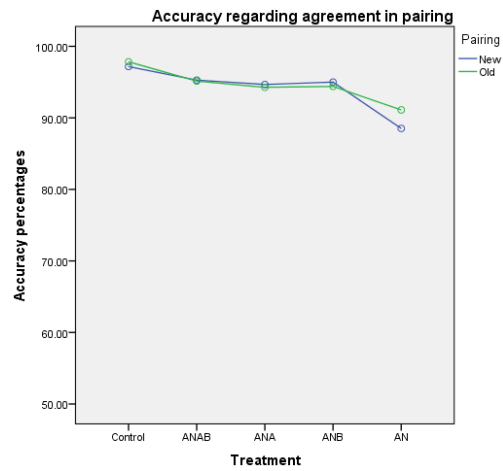


Figure 8. Mixed ANOVA results regarding agreement in pairing

RO4: What differences in performance are there between concordance, noun ending, gender and treatment?

A four-way 5 (Treatment) x 2 (Concordance, agreement or assignment) x 2 (Gender, masculine or feminine) x 3 (ending, canonical, non-canonical, and exception) mixed ANOVA with repeated measures was conducted to find group differences. According to Mauchly's test, the assumption of sphericity was not violated, except for noun ending, $\chi^2(2)$, $p = .015$, since the value for the Greenhouse-Geisser correction was higher than $>.075$, I used the Huynh-Feldt correction ($\epsilon = .922$). The interaction between Concordance and Noun Ending also violated the assumption of sphericity, $\chi^2(2)$, $p = .007$, $\epsilon = .909$, so the Huynh-Feldt correction was also used. Canonical nouns also violated the assumption of normality both in assignment and agreement. The assumption of normality was only violated regarding the assignment of masculine canonical nouns, $F(4, 104) = 3.69$, $p = .008$, and the assignment in feminine non-canonical nouns, $F(4, 104) = 3.22$, $p = .015$. The following hypothesis was tested:

H₀: There are no significant differences in performance between gender, noun ending, and concordance.

Significant differences were found for Treatment (Table 19), thus indicating that groups behaved differently, $F(4, 104) = 3.5, p = .010, r = 13735.33$, and disproving the hypothesis that there are no differences between groups in the intercept with gender, ending and concordance. Significant differences regarding Concordance were also present, indicating that learners showed different degrees of accuracy regarding assignment and agreement, $F(1, 104) = 38.90, p = .000, r = 4482.56$; learners were more accurate in agreement than assignment. However, no significant differences were found in the interaction of Concordance with Treatment, thus showing that the effect of the different treatment on the performance was not different in Agreement compared to Assignment; for this reason, this interaction was no longer pursued. A significant difference per noun ending was found (Table 20), $F(1.96, 104) = 828.05, p = .000, r = 1239$, and also in its interaction with treatment (Table 21), $F(7.84, 104) = 3.29, p = .002, r = 22058.93$. The intercept of Concordance*Gender*Ending*Treatment did not result in any significance, it was therefore not further pursued. Other interactions were not present in the software's output.

Table 19
Differences between treatment groups

Treatment	Mean difference	<i>p</i> value	Direction of effect
C ANA	8.22	.031*	C>ANA
AN	9.9	.010*	C>AN

Note. C=Control; ANAB=Article, Noun, and Adjective in bold font; ANA=Article, Noun, and Adjective; ANB=Article and Noun in bold font; AN=Article and Noun. * $p < .05$, ** $p < .01$, *** $p < .001$

Table 20

<i>Direction of effects between Endings</i>				
Endings		Mean difference	<i>p</i> value	Direction of effect
C	NC	39.2	.000***	C>NC
	E	80.41	.000***	C>E
NC	E	41.21	.000***	NC>E

Note. Noun endings are coded as C=Canonical, NC=Non-Canonical, E=Exception. **p* < .05, ***p* < .01, ****p* < .001

Table 21. *Interaction between Ending*Treatment*

Ending	Treatment		Mean difference	<i>p</i> value	Direction of effect
Canonical	C	AN	10.18	.043*	C>AN
		ANAB	11.39	.019*	ANAB>AN
Non-Canonical	C	ANAB	17.9	.001**	C>ANAB
		ANA	20.11	.000***	C>ANA
		ANB	17.97	.001**	C>ANB
		AN	24.00	.000***	C>AN

Note. C=Control; ANAB=Article, Noun, and Adjective in bold font; ANA=Article, Noun, and Adjective; ANB=Article and Noun in bold font; AN=Article and Noun. **p* < .05, ***p* < .01, ****p* < .001

The means for each of the factors indicate the same trends as found in the previous section, the groups performed from most to least accuracy in the following order: control, ANAB, ANB, ANA, and AN; noun accuracy was better among masculine nouns than feminine nouns; accuracy was also better in canonical nouns, followed by non-canonical and exceptions; and agreement outperformed assignment significantly.

RQ5: What differences are there between groups regarding the gender of nouns and the immediate and delayed post-test in assignment?

A three-way 5 (Treatment) x 2 (Test, immediate or delayed) x 2 (Gender, masculine or feminine) mixed ANOVA with repeated measures on the accuracy scores for Tests and Gender was performed. The main effect of Test, Concordance, or the combination of both

do not significantly violate the assumption of sphericity. Normality was also not violated according to Levene’s test for any of the variables. The following hypothesis was tested:

H₀: There are no significant differences in performance between gender and test regarding assignment.

Significant differences were found for Treatment (Table 22), thus indicating that groups behaved differently, $F(4, 104) = 3.29, p = .014, r = 2236.33$, and disproving the hypothesis that there are no differences between groups regarding gender and test. No variables showed significance, nor did their interactions, with the exception of Gender in which accuracy was higher in masculine nouns than feminine, $F(1, 104) = 101.69, p = .000, r = 23112$; suggesting that significant differences in the sample were caused by the differences between masculine and feminine nouns. Since they were not significant, other interactions were not addressed.

Table 22
Differences between treatment groups in the delayed post-test

Treatment	Mean difference	<i>p</i> value	Direction of effect
C ANA	6.67	.009**	C>ANA

Note. C=Control; ANAB=Article, Noun, and Adjective in bold font; ANA=Article, Noun, and Adjective; ANB=Article and Noun in bold font; AN=Article and Noun. **p* < .05, ***p* < .01, ****p* < .001

The overall means showed a tendency to the control group being more accurate, followed by ANAB, ANB, ANA, and AN. However, the only significant interactions were found between the Control group and ANA. The immediate test had a non-significant slight higher mean with respect to the delayed post-test, while accuracy for masculine nouns was significantly higher than for feminine nouns. When looking at the interaction between treatment and test the only exception to the tendencies observed above is the AN group,

who did slightly better in the delayed post-test over the immediate post-test. Figures 9 and 10 provide an overview of the means by group.

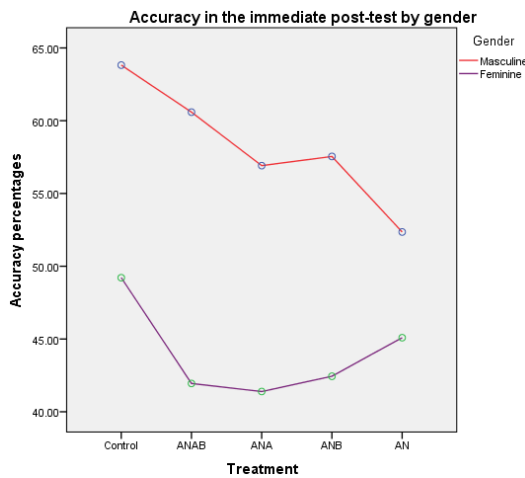


Figure 9. Mixed ANOVA results in the immediate post-test divided in assignment by gender

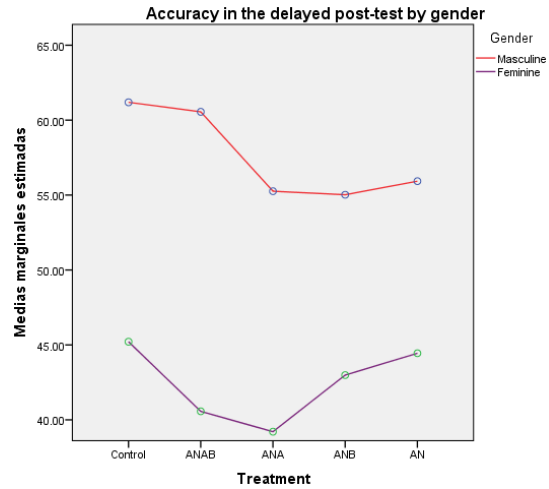


Figure 10. Mixed ANOVA results in the delayed post-test divided in assignment by gender

RQ6: What differences are there between groups regarding the gender of nouns and the immediate and delayed post-test in agreement?

A three-way 5 (Treatment) x 2 (Test, immediate or delayed) x 2 (Gender, masculine or feminine) mixed ANOVA with repeated measures on the accuracy scores for Test and Gender was performed. The main effect of Test, Gender, or the combination of both do not significantly violate the assumption of sphericity. Normality was also not violated according to Levene’s test for any of the variables. The following hypothesis was tested:

$$H_0: \text{There are no significant differences in performance between gender and test regarding agreement.}$$

Significant differences were found for Treatment (Table 23), thus indicating that groups behaved differently, $F(4, 104) = 3.74, p = .007, r = 4103.26$, and disproving the hypothesis that there are no differences between groups regarding gender and test. No

significance was found for other variables nor their interactions, except for the interaction between Test and Treatment (Table 24), $F(4, 104) = 2.53, p = .045, r = 887.29$, suggesting that the differences in our sample might be accounted for the treatment along with the test. The other significant interaction is found regarding gender where again higher accuracy rates were found for masculine nouns in contrast to feminine, $F(4, 104) = 151.14, p = .000, r = 29357.37$, suggesting that differences can be accounted by masculine and feminine nouns.

Table 23
Differences between treatment groups in the delayed post-test

Treatment		Mean difference	<i>p</i> value	Direction of effect
C	ANA	8.66	.007**	C>ANA
	AN	8.49	.021*	C>AN

Note. C=Control; ANAB=Article, Noun, and Adjective in bold font; ANA=Article, Noun, and Adjective; ANB=Article and Noun in bold font; AN=Article and Noun. * $p < .05$, ** $p < .01$, *** $p < .001$

Table 24
*Interaction between Test*Treatment in the delayed post-test*

Test	Treatment		Mean difference	<i>p</i> value	Direction of effect
Immediate Post-Test	C	ANAB	7.41	.021*	C>ANAB
		ANA	8.92	.005**	C>ANA
		ANB	6.91	.035*	C>ANB
		AN	12.61	.000***	C>AN
Delayed Post-Test	C	ANA	8.41	.001**	C>ANA
		ANAB	5.20	.034*	ANAB>ANA
		ANA	-5.63	.025*	ANA>ANB

Note. C=Control; ANAB=Article, Noun, and Adjective in bold font; ANA=Article, Noun, and Adjective; ANB=Article and Noun in bold font; AN=Article and Noun. * $p < .05$, ** $p < .01$, *** $p < .001$

The overall means show a tendency of accuracy in which the control group is best, followed by ANB, ANAB, AN, and ANA, which differs from previous analyses. Data

suggest that, when looking into agreement accuracy, unexpectedly, the presence of an adjective affected performance negatively while bold font aided accuracy more. When accounting for the gender of the nouns, we find slightly improved performance in the delayed post-test. Masculine nouns still present higher accuracy rates than feminine. We see the same tendencies in all interactions, except for the Treatment*Test, in which ANA did better in the immediate post-test than the delayed post-test. Figures 11 and 12 provide an overview of the means by group.

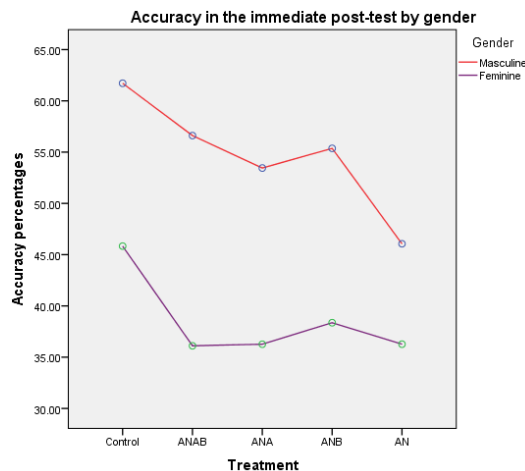


Figure 11. Mixed ANOVA results regarding the immediate post-test divided by gender in assignment

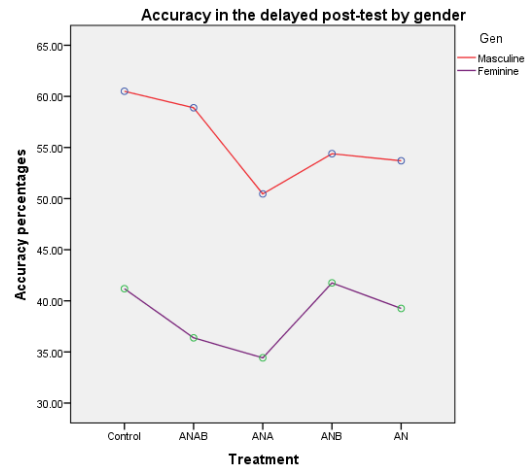


Figure 12. Mixed ANOVA results regarding the delayed post-test divided by gender in agreement

RQ7: What differences are there between groups regarding noun ending and the immediate and delayed post-test in assignment?

A three-way 5 (Treatment) x 2 (Test, immediate or delayed) x 3 (Noun Ending, canonical, non-canonical, or exceptions) mixed ANOVA with repeated measures on the accuracy scores for Tests and Noun Endings was performed. The main effect of Test and Noun Ending do not significantly violate the assumption of sphericity, however the combination thereof does, $\chi^2(2), p = .004$, since the value for the Greenhouse-Geisser correction was higher than $>.075$, I used the Huyn-Feldt correction ($\epsilon = .909$). The

assumption of normality held for all variables, except for assignment of canonical nouns, $F(4, 104) = 3.12, p = .018$. The following hypothesis was tested:

H₀: There are no significant differences in performance between noun ending and test regarding assignment.

Significant differences were found for Treatment (Table 25), thus indicating that groups behaved differently, $F(4, 104) = 3.21, p = .016, r = 3250.9$, and disproving the hypothesis that there are no differences between groups regarding gender and test. No significant values were found for Test or the Test*Treatment interaction. Nevertheless, Ending did reach significance (Table 26), $F(2, 140) = 1534.02, p = .000, r = 791745.78$; its interaction with treatment also rendered significance (see Table 27) at $F(8, 104) = 2.13, p = .034, r = 4398.6$. Test*Ending also proved significant at $F(1.92, 104) = 6.98, p = .001, r = 1107.74$ (Table 28 in interactions). The interaction Test*Ending*Treatment also proved significant, $F(7.68, 104) = 2.08, p = .042, r = 1317.15$ (Table 29 under interactions). This suggests that noun ending and its combination with other variables explain most of our data.

Table 25

Differences between treatment groups in the delayed post-test

Treatment	Mean difference	<i>p</i> value	Direction of effect
C ANA	6.55	.011*	C>ANA

Note. C=Control; ANAB=Article, Noun, and Adjective in bold font; ANA=Article, Noun, and Adjective; ANB=Article and Noun in bold font; AN=Article and Noun. * $p < .05$, ** $p < .01$, *** $p < .001$

Table 26

Direction of effects between Endings in the delayed post-test

Endings	Mean difference	<i>p</i> value	Direction of effect
C NC	41.66	.000***	C>NC

	E	85.83	.000***	C>E
NC	E	44.17	.000***	NC>E

Note. Noun endings are coded as C=Canonical, NC=Non-Canonical, E=Exception. * $p < .05$, ** $p < .01$, *** $p < .001$

Table 27

*Interaction between Ending*Treatment in the delayed post-test*

Ending	Treatment	Mean difference	p value	Direction of effect
Non-Canonical	C ANAB	11.47	.006**	C>ANAB
	ANA	15.78	.000**	C>ANA
	ANB	11.92	.005**	C>ANB
	AN	15.27	.001**	C>AN

Note. C=Control; ANAB=Article, Noun, and Adjective in bold font; ANA=Article, Noun, and Adjective; ANB=Article and Noun in bold font; AN=Article and Noun. * $p < .05$, ** $p < .01$, *** $p < .001$

The overall means show a tendency of accuracy in which the control group is best, followed by ANB, ANAB, AN, and ANA, which suggests that, just like with noun gender, unexpectedly, the main advantage would be the enhancement and within the enhancement the absence of adjectives shows better performance. As in previous analyses, accuracy is higher among canonical nouns, followed by non-canonical and exceptions. The tendencies hold within interaction, except for the AN group with regard to Treatment, since it underperformed slightly on the immediate post-test, but only by < 1 %. Figures 13 and 14 provide an overview of the means by group.

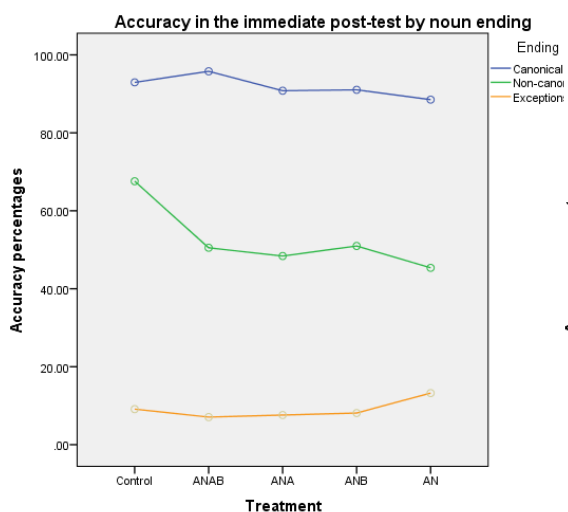


Figure 13. Mixed ANOVA results for the immediate post-test regarding noun ending in assignment

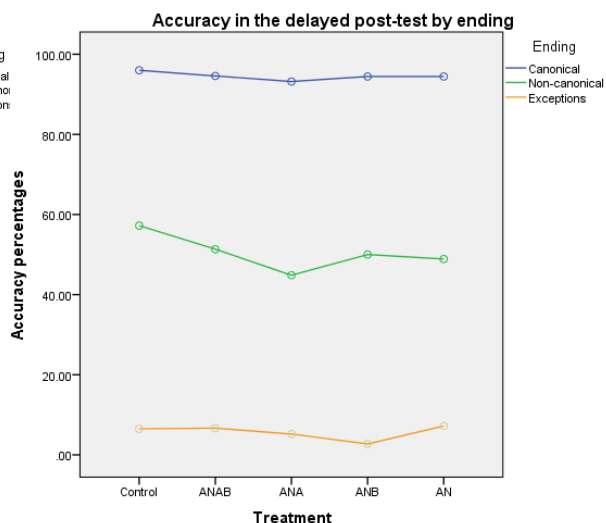


Figure 14. Mixed ANOVA results for the delayed post-test regarding noun ending in assignment

Interactions

Interestingly, in nouns with a canonical ending there is a significant improvement in accuracy in the delayed post-test in comparison to the immediate post-test, while there are no significant differences in non-canonical nouns and we find the opposite effect in exceptional endings, where better results are found in the immediate post-test (a full report of the data is provided in Table 28). The only significant interactions found in the Treatment*Test interaction showed that the control group outperformed the rest. When looking into the Test*Ending*Treatment interaction we find that regarding non-canonical nouns the control group significantly outperforms the rest of groups, but no other significant effects appear in our results (see Table 29).

Table 28

*Interaction between Ending*Test in the delayed post-test*

Ending	Test		Mean difference	<i>p</i> value	Direction of effect
Canonical	IPT	DPT	-2.71	.012*	IPT<DPT
Non-canonical	IPT	DPT	2.11	.163	-
Exception	IPT	DPT	3.37	.002**	IPT>DPT

Note. IPT=Immediate Post-Test; DPT=Delayed Post-Test. **p* < .05, ***p* < .01, ****p* < .001

Table 29

*Interaction between Treatment*Test*Ending in the delayed post-test*

Test	Ending	Treatment	Mean difference	<i>p</i> value	Direction of effect
Immediate Post-test	Non-Canonical	C ANAB	17.07	.001**	C>ANAB
		ANA	19.18	.000***	C>ANA
		ANB	16.61	.002**	C>ANB
		AN	22.21	.000***	C>AN

Note. C=Control; ANAB=Article, Noun, and Adjective in bold font; ANA=Article, Noun, and Adjective; ANB=Article and Noun in bold font; AN=Article and Noun. Only significant interactions are reported. **p* < .05, ***p* < .01, ****p* < .001

RQ8: What differences are there between groups regarding noun ending and the immediate and delayed post-test in agreement?

A three-way 5 (Treatment) x 2 (Test, immediate or delayed) x 3 (Noun Ending, canonical, non-canonical, or exceptions) mixed ANOVA with repeated measures on the accuracy scores for Tests and Noun Endings was performed. The main effect of Test does not significantly violate the assumption of sphericity, however Noun Ending does, $\chi^2(2)$, *p* = .012, since the value for the Greenhouse-Geisser correction was higher than >.075, I used the Huyn-Feldt correction (ϵ = .923); the interaction of Test*Ending also violates the assumption of sphericity, $\chi^2(2)$, *p* = .042, with the Greenhouse-Geisser value being ϵ = .944, thus the correction is also used for the interaction. The assumption of normality is only violated in agreement of canonical noun endings in the immediate post-test, *F*(4, 104)

= 3.03, $p = .021$, and delayed post-test, $F(4, 104) = 1.03, p = .001$. The following hypothesis was tested:

*H₀: There are no significant differences in performance between
noun ending and test regarding agreement.*

Significant differences were found for Treatment (Table 30), thus indicating that groups behaved differently, $F(4, 104) = 3.88, p = .006, r = 6540.12$, and disproving the hypothesis that there are no differences between groups regarding gender and test. No significant values were found for Test or the Test*Treatment interaction. Nevertheless, Ending did reach significance (Table 31), $F(2, 140) = 1193.6, p = .000, r = 720128.4$; its interaction with treatment also rendered significance (Table 32) at $F(7.8, 104) = 2.42, p = .017, r = 5846.66$. Test*Ending also proved significant (Table 33 under interactions) at $F(1.996, 104) = 7.18, p = .001, r = 1381.69$. The interaction Test*Ending*Treatment also proved significant (Table 34 under interactions), $F(7.98, 104) = 2.68, p = .008, r = 2059.34$. This suggests that noun ending and its combination with other variables explain most of our data.

Table 30
Differences between treatment groups in the delayed post-test

Treatment	Mean difference	p value	Direction of effect
C ANA	9.20	.004**	C>ANA
AN	8.33	.028*	C>AN

Note. C=Control; ANAB=Article, Noun, and Adjective in bold font; ANA=Article, Noun, and Adjective; ANB=Article and Noun in bold font; AN=Article and Noun. * $p < .05$, ** $p < .01$, *** $p < .001$

Table 31
Direction of effects between Endings in the delayed post-test

Ending	Mean difference	<i>p</i> value	Direction of effect
C NC	40.16	.000***	C>NC
E	81.86	.000***	C>E
NC E	41.70	.000***	NC>E

Note. Noun endings are coded as C=Canonical, NC=Non-Canonical, E=Exception. **p* < .05, ***p* < .01, ****p* < .001

Table 32
*Interaction between Ending*Treatment in the delayed post-test*

Ending	Treatment	Mean difference	<i>p</i> value	Direction of effect
Canonical	C AN	9.69	.049*	C>AN
Non-Canonical	C ANAB	13.25	.003**	C>ANAB
	ANA	18.42	.000***	C>ANA
	AN	17.35	.000***	C>AN

Note. C=Control; ANAB=Article, Noun, and Adjective in bold font; ANA=Article, Noun, and Adjective; ANB=Article and Noun in bold font; AN=Article and Noun. **p* < .05, ***p* < .01, ****p* < .001

The overall means show a tendency of accuracy in which the control group is best, followed by ANB, ANAB, AN, and ANA, which differs from analyses in RQs 1 through 5. Data suggest that when looking into agreement, the most helpful feature is enhancement and we obtain mixed results for the presence or absence of an adjective. The scores in the delayed post-test are higher than those for the immediate post-test and scores on canonical nouns are higher followed by those with non-canonical endings and exceptions. Tendencies hold across interactions, except for Treatment*Test, where the control and ANA groups perform slightly better in the immediate post-test. Figures 15 and 16 provide an overview of the means by group.

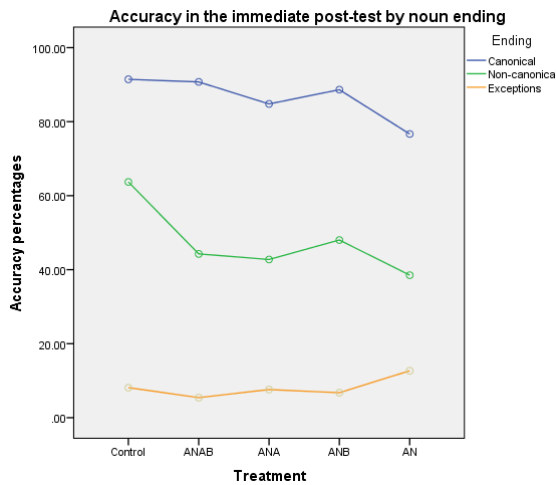


Figure 15. Mixed ANOVA results in the immediate post-test regarding noun ending in agreement

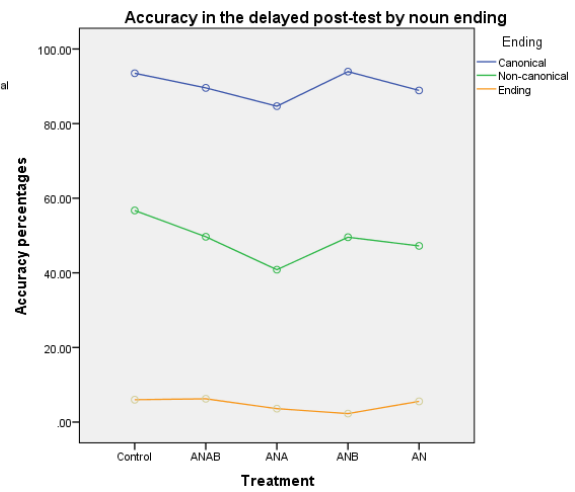


Figure 16. Mixed ANOVA results in the delayed post-test regarding noun ending in agreement

Interactions

Interestingly, in nouns with a canonical ending there is a significant improvement in accuracy in the delayed post-test in comparison to the immediate post-test, while there are no significant differences in non-canonical nouns and we find the opposite effect in exceptional endings, where better results are found in the immediate post-test (a full report of the data is provided in Table 33). The only significant interactions found in the Treatment*Test interaction showed that the control group outperformed the rest.

Table 33

Interactions between Ending*Test in the delayed post-test

Ending	Test		Mean difference	<i>p</i> value	Direction of effect
Canonical	IPT	DPT	-3.66	.023*	IPT<DPT
Non-canonical	IPT	DPT	-1.35	-	-
Exception	IPT	DPT	3.38	.001**	IPT>DPT

Note. IPT=Immediate Post-Test; DPT=Delayed Post-Test. **p* < .05, ***p* < .01, ****p* < .001

The significant interactions between Test, Ending and Treatment are reported on Table 34. We observe that in the immediate post-test all significant interactions show that for nouns with canonical ending AN underperforms in comparison to the control, ANAB, and ANB groups, which suggests that the enhancement improves accuracy significantly in comparison to exposure to only article-noun input. Regarding non-canonical nouns no differences were found among treatment groups, except for the control group outperforming the rest. In the delayed post-test canonical nouns show that the ANA group significantly underperformed with respect to the control, ANB, and ANAB, indicating again that the enhancement improves performance with regard to being exposed the noun-adjective combination without enhancement, no significant differences were however found with respect to the AN treatment.

Table 34
*Interaction between Treatment*Test*Ending in the delayed post-test*

Test	Ending	Treatment		Mean difference	<i>p</i> value	Direction of effect
Immediate Post-test	Canonical	AN	C	-14.78	.015*	AN<C
			ANAB	-14.07	.016*	AN<ANAB
		ANB	-11.97	.044*	AN<ANB	
	Non-Canonical	C	ANAB	19.45	.001**	C>ANAB
			ANA	20.95	.000***	C>ANA
			ANB	15.69	.006**	C>ANB
Delayed Post-test	Canonical	ANA	ANB	-9.27	.044*	ANA<ANB
			C	-15.88	.001**	ANA<C
		ANAB	-8.83	.046*	ANA<ANAB	

Note. C=Control; ANAB=Article, Noun, and Adjective in bold font; ANA=Article, Noun, and Adjective; ANB=Article and Noun in bold font; AN=Article and Noun. Only significant interactions are reported. **p* < .05, ***p* < .01, ****p* < .001

RQ9: What differences are there between the learning of adjectives and nouns in the ANAB, ANA and control groups?

A two-way 5 (Treatment) x 2 (Category, noun or adjective) repeated-measures ANOVA on the accuracy scores for Category was performed. The main effect of Treatment and Category did not violate the assumption of sphericity according to Mauchly's test nor did they violate the assumption of normality according to Levene's test. The following hypothesis was tested:

H₀: There are no significant differences between knowledge of adjectives and nouns between groups.

No significant differences were found per Treatment or Category, however, the interaction did render significant results, $F(2, 104) = 5.64, p = .005, r = 699.64$. The overall means show a tendency in the control group to outperform the ANAB group followed by the ANA group. Adjectives seem to have better accuracy scores by approximately 1 %, these patterns do not however hold across interactions. Despite the significant interaction mentioned above, the only significant difference was found between the control and the ANA group at $p = .021$, with the control outperforming the ANA group. This suggests that the accuracy resulting from exposure did not differ depending on input enhancement nor treatment. Figure 17 provides an overview of how the data behaved.

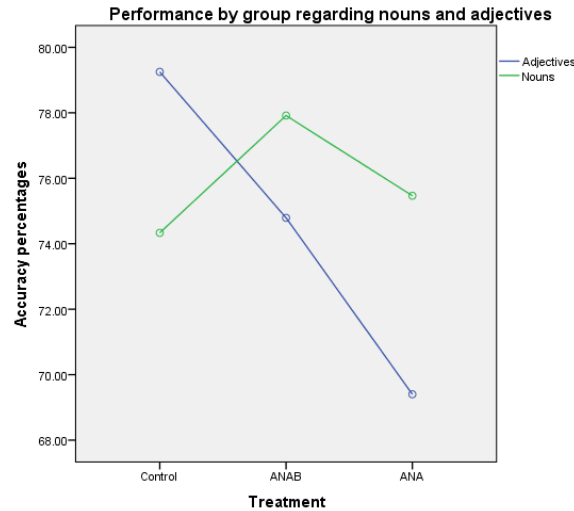


Figure 17. Accuracy scores for adjectives and nouns in the translation test.

RQ10: What differences are there between the learning of adjectives depending on their frequency in the ANAB, ANA and control groups?

A two-way 5 (Treatment) x 2 (Frequency, 2 versus 4 exposures) repeated-measures ANOVA on the accuracy scores for Frequency was used. The main effect of Treatment and Frequency did not violate the assumption of sphericity according to Mauchly's test nor did they violate the assumption of normality according to Levene's test. The following hypothesis was tested:

H₀: There are no significant differences between knowledge of adjectives based on their frequency between groups.

No significant differences were found per Treatment or Category nor their interaction. For this reason, I did not pursue the results further. The lack of statistical significance suggests that participants were not sensitive to frequency. Figure 18 provides an overview of how the data behaved.

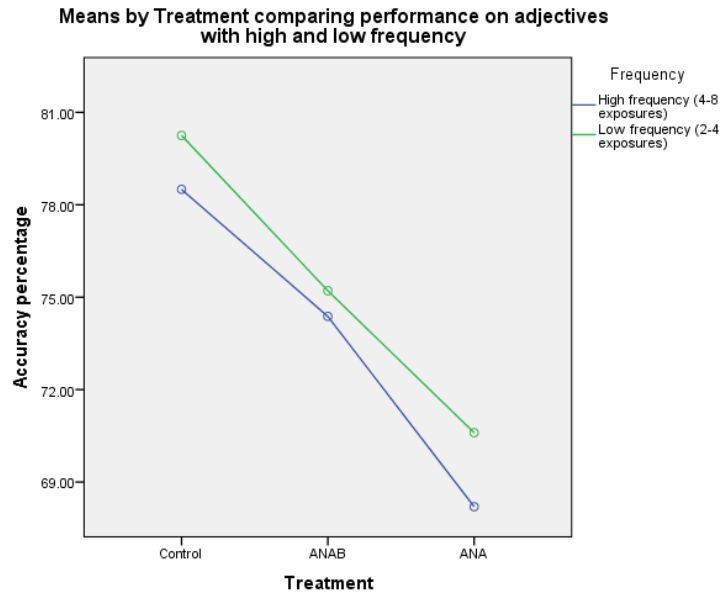


Figure 18. Accuracy scores for adjectives depending on frequency in the translation test.

Correlations Between Dependent and Independent Variables

Given that the data did not show the expected interactions, Spearman’s rho correlation was run to see whether the relatedness of variables might help shed some light into the results. The independent variables that were observed were proficiency, answer sheet, and form recognition tests’ scores. Assignment in feminine nouns positively correlated with accuracy in adjectives ($r = .398, n = 69, p = .001$) and nouns ($r = .009, n = 69, r = .000$), agreement in feminine nouns showed similar results with adjectives ($r = .370, n = 69, r = .002$) and nouns ($r = .236, n = .236, r = .013$). The same pattern was found for feminine nouns in the delayed post-test. Performance regarding assignment of feminine nouns positively correlated with knowledge of adjectives ($r = .361, n = 69, p = .002$) and of nouns ($r = .208, n = 109, r = .030$). Agreement of feminine nouns showed similar results with respect to knowledge of adjectives ($r = .348, n = 69, p = .003$) and nouns ($r = .250, n = 109, p = .009$).

Regarding noun ending only exceptions within assignment correlated with knowledge of adjectives ($r = .003, n = 69, r = .003$) and nouns ($r = .189, n = 69, r = .049$), exceptions within agreement also correlated with adjectives ($r = .290, n = 69, n = .016$) but not nouns. The pattern still held in the delayed post-test for assignment in exceptions ($r = .280, n = 69, p = .020$) but not nouns. The correlation was replicated in relation to agreement of nouns with exceptional endings as well ($r = .294, n = 69, p = .014$).

In contrast with the immediate post-test, the delayed post-test agreement of canonical nouns correlated with knowledge of adjectives ($r = .261, n = 69, p = .030$) and of nouns ($r = .285, n = 109, p = .003$). Agreement of non-canonical nouns also correlated with knowledge of adjectives ($r = .411, n = 69, p = .000$) and nouns ($r = .259, n = 109, p = .007$).

There were also positive correlations between new pairings of nouns and adjectives in the treatment and accuracy in adjectives ($r = .405, n = 69, p = .001$) and in nouns ($p = .230, n = 109, p = .016$); the same pattern was found among the old pairings in adjectives ($p = .434, p = 69, n = .000$) and nouns ($p = .249, p = 109, r = .009$).

The correlations with proficiency, performance in the answer sheet and accuracy in performance in the form recognition test showed some general tendencies. Hours of practice or looking at the screen did not correlate systematically with anything. Given the number of variables, those that showed a relation are reported on Table 35 below and only general tendencies are described. The observed trends are that proficiency, performance in the answer sheet and performance in the form recognition display the most correlations with features that seem to pose a bigger issue to learners, feminine nouns, non-canonical nouns and exceptions, both in agreement and assignment. The most correlations are found

with proficiency. Thus indicating that proficiency and non-default features of the language tend to go together, as proficiency increases so does accuracy and vice versa.

Table 35
Correlations between dependent and independent variables

	Variables	Proficiency	Answer sheet	Form recognition
Accuracy	Nouns	.374**	.392**	.515***
	Adjectives	.138	.606***	.423***
Frequency	High	.329**	.292*	.502***
	Low	.359**	.457***	.451***
Assignment	Feminine	.436***	.262*	.290**
Agreement	Feminine	.409***	.290**	.257**
Assignment	Canonical	.111	.310**	.054
	Non-canonical	.325***	.227*	.120
	Exceptions	.158	.075	.231*
Agreement	Canonical	.169	.232*	.055
	Non-canonical	.350***	.252*	.159
	Exceptions	.186	.091	.262*
Assignment	New pairing	.315***	.309**	.240*
	Old pairing	.200*	.150	-.007
Agreement	New pairing	.224*	.105	.051
	Old pairing	.104	.128	.108
Assignment	Feminine DPT	.258*	.202	.193
Agreement	Feminine DPT	.250*	.231*	.246*
Assignment	Canonical DPT	.007	.222*	.020
Assignment	Non-canonical DPT	.281**	.165	.310
Agreement	Canonical DPT	.073	.217*	.158
	Non-Canonical DPT	.241*	.205	.359***

Note. Unless otherwise indicated, results refer to the immediate post-test.
DPT=Delayed Post-Test. * $p < .05$, ** $p < .01$, *** $p < .001$

CHAPTER VI: DISCUSSION

At first hand, the results presented might seem rather uninformative and inconclusive, however, a second look reveals that there is consistency in the variation of the factors observed. Given that most variables behaved similarly across research questions and to avoid repetition, this discussion will address them in the following order (1) differences among treatment groups, (2) differences between performance on nouns with different gender, (3) differences between concordance performance, (4) differences between noun endings, (5) differences between new and old noun-adjective pairings, and (6) correlation of variables. A discussion of learning of nouns versus adjectives will follow, together with a discussion of frequency effects. The discussion will focus on the presence and absence of differences across variables, its (dis)similarities with the literature, and the possible causes for the results obtained, which will make references to the results of the correlation analyses. Finally, a brief discussion about ultimate attainment in relation to the current results is presented.

Differences Among Treatment Groups

Results indicated significant differences among treatment groups in all research questions, however, the differences did not go as expected upon observing Bonferroni's results. At first glance the outcome seems to indicate that the treatment was ineffective as significant differences showed that only the control group behaved significantly better, which was not expected. However, results also showed that the proficiency of the control group was significantly higher than that of the treatment groups and correlations showed that proficiency did correlate with the variables that deviated from the default form,

masculine nouns with canonical endings. Previous research (e.g. Montrul et al., 2008; Martínez-Gibson, 2011) has shown that proficiency is one of the best predictors for gender assignment and agreement.

Still, why were there no differences across treatment groups? First, there were a few significant differences, although only in the specific interactions across several variables: the most frequent significant difference was between the ANAB and the AN groups, with the ANAB group outperforming AN. Nevertheless, this only occurred in the following interactions (unless otherwise indicated it refers to the immediate post-tests): Ending*Treatment, Test*Treatment (in the delayed post-test when looking into agreement), Concordance*Ending*Treatment, and Ending*Test*Treatment (in the immediate post-test for agreement). This difference that occurs in some interactions suggests that, overall, both enhancement and more exposure to agreement through adjectives, for instance, are necessary for there to be a statistically significant improvement in performance in the immediate post-test, indicating that either one or the other are not enough. In the delayed post-test for agreement, however, ANAB outperforms ANA in the Test*Treatment interaction and in the Treatment*Test*Ending interaction, suggesting that in this case the presence of an unenhanced adjective only hindered better acquisition.

Secondly, the article-noun-adjective groups might have imposed cognitive demands that were too high. As explained in the Methodology, article-noun-adjective groups (ANA and ANAB) were exposed to 70 novel items, while those in the article-noun groups (AN and ANB) only had to learn 30. These extra 40 items might have overloaded those students' working memory or made the exercise too taxing. Despite not having tested for working memory, significant differences were observed between the two types of groups (ANA and

ANAB on the one hand and AN and ANB on the other) in their performance with regard to the form recognition test, which suggests that participants in the noun-article (AN and ANB) conditions did not even register several of the items. In previous research we find that working memory correlates with sensitivity to gender agreement (Sagarra & Herschensohn, 2012), with performance only during treatment, believed to be a factor only if participants are given specific instructions to find variation within the stimuli (McDonough & Trofimovich, 2016). Working memory also affects recognition according to Malone (2018) and Varol & Erçetin (2016); although some studies advocate for the role of statistical learning, learning through statistically likely patterns, over working memory (McDonough & Trofimovich, 2016). In conclusion, working memory plays an important role with regard to performance and recollection. What also stems from the literature is that working memory needs to be sufficiently engaged and the task needs to have a certain degree of involvement for effects to be found (McDough & Trofimovich, 2016) without overloading the system (overload effects were not reported on by previous studies) in order to avoid taking up all the memory and hindering learning.

Finally, despite the lack of statistical support, some tendencies did rise in the overall performance in concordance. Among the immediate post-test analyses, the constant patterns showed better accuracy for ANAB, followed by ANB, ANA, and AN groups. This suggests that albeit they did not reach statistical significance and such outcome might be due to chance, learners did seem to benefit from the combined effects of enhancement and more exposure to agreement through adjectives. The ANB group, coming in second among the treatments, might point at the bold font enhancement having more benefits than the ANA condition which, despite providing more exposure to agreement, does not draw

attention to it. Among the delayed post-tests the pattern is different, from better to worse performance groups are ranked as ANB, ANAB, AN, and ANA (except for the immediate post-test regarding gender in agreement, which shows the same pattern as the immediate post-test). This order again suggests the superiority of the bold font enhancement but also surprisingly indicated that the presence of the adjective was hindering the process (although not enough to counter the effects of the bold font). It is surprising that it is precisely in agreement where this pattern arises for the first time, as the original expectation was that increased exposure to the adjective would aid later performance. In light of the literature on grammatical gender acquisition it can be argued that these results show tendencies in line with Lee & Huang (2008), indicating that visually enhanced input supports learning if we account for proficiency.

Given the low scores in the form recognition test and the apparently hindering effect of adjectives in the delayed post-test along with the lack of statistical difference among groups, it is my hypothesis that the difference in the total of items (a difference of 40) across conditions hindered the results and prevented differences from arising. Inasmuch a good explanation as that is, it does not explain the lack of significant differences between the AN and ANB groups, both of which were exposed to the same number of items. I would suggest that the design of the treatment with pictures linked to decontextualized sentences might have slanted the results, and that being exposed to decontextualized sentences might not be enough of a stimulus. For learners to be more engaged and register the information, it may be that a written text like in Malone (2018) or Bordag et al. (2017) is necessary, although studies such as Morgan-Short et al. (2010) and Denhovska & Serratrice (2017) did find improvement after exposure to sentences only. Nevertheless, there is a noteworthy

difference, which is that the first two studies used a target language that was already known to learners, while the latter two used a novel language; therefore, it could be hypothesized that if there is previous knowledge of the language system sentences do not suffice.

Other possible confounding variables that account for the absence of results were (1) lack of motivation of students that might have led to lack of attention, (2) abnormal behavior, e.g. a participant standing up and sitting down repeatedly during treatment or talking and complaining during it, (3) test taking strategies such as choosing first all articles and then all adjectives during the test as observed by the researcher during testing, and (4) the stimuli not being engaging enough for the students and causing lack of cognitive involvement.

Some studies have also argued the need for awareness in order for incidental learning to work (e.g. Denhovska & Serratrice, 2017), although others just argue for the necessity of noticing (e.g. Bell & Collins, 2009). Given the questions asked in the debriefing questionnaire it can only be ventured that participants both did not notice nor had awareness of the grammatical gender feature under observation. Again, there is an added difficulty to determining this, as participants did have explicit knowledge of the grammatical system according to the curriculum for the course, but awareness or noticing of the treatment items might have been necessary. The debriefing questionnaire showed that no learners discovered the target of the treatment, a possible explanation for the lack of results. Nevertheless, the studies mentioned above used novel features, while in the current study learners were already acquainted with the rule. With the current information and research it is impossible to discern whether awareness would be required for the specific new items being taught if the feature is already known. In line with Denhovska &

Serratrice (2017) these results suggest that incidental learning allows for receptive knowledge but not productive, which would be in line with learning needing to be first receptive to then become productive; so, in line with DeKeyser (2008), implicit modes of learning might require more time to reach the same levels of success.

In summary, results were inconclusive with regard to the effectiveness of the treatments for the reasons stated above and showed no statistical differences between treatment groups, even though there were tendencies that supported my hypotheses. Previous research posits enough evidence about the variables (working memory and richness of context primarily) that can affect the outcome of incidental learning of grammar and vocabulary to not disregard the treatments entirely, since most of those variables could not be measured.

Differences Among the Grammatical Gender of the Nouns

The significant differences in performance accuracy across masculine and feminine nouns in both agreement and assignment align with previous findings in the acquisition of Spanish grammatical gender (in line with Montrul et al., 2008; but against findings in Italian as reported by Martino et al., 2017). The increased accuracy in masculine nouns suggests that, when in doubt, learners resort to the masculine gender, which they might perceive as the default, as argued by Corbett (1991). The analysis showed that the treatment did not correct or level the differences in performance regarding grammatical gender. This might be due to the ineffectiveness of the treatment as explained above or due to the possibility that learners have interiorized certain tendencies and overruling them is not possible with such a short treatment. One must bear in mind that these participants were in an intermediate level I class, and that they were already familiar with the grammatical

gender system. Modifying this knowledge that might already be in the process of becoming automatized declarative knowledge might require further exposure to modify their pre-established assumptions regarding gender. As Tipurita & Jean (2014) show, explicit instruction can render positive results with learners with some proficiency of the language. Williams (2005) and Wust & Roche (2015) also provide evidence towards the effectiveness of explicit over implicit learning.

In short, performance with regard to the noun's gender conformed to previous research showing higher accuracy rates for masculine nouns and this research strengthens the claim of a default masculine for Spanish, at least among second language learners.

Differences Between Concordance

Concordance also showed significant differences, suggesting that assignment might have been easier for learners than agreement, in line with Montrul et al. (2008). Accuracy being higher in assignment seems to suggest that learners have not yet integrated the grammatical gender system of agreement and might be learning in chunks (in line with Lew-Williams & Fernald, 2010, and Tipurita & Jean, 2014; the latter only with regard to marked features). However, more clarity with regard to chunk learning is provided in the discussion about new versus old noun-adjective pairings. The pattern did however reverse in the interaction Concordance*Ending*Gender*Treatment, showing significantly higher rates of accuracy for agreement. This suggests that when taking into account all the variables involved in the study in the immediate post-test, agreement was more accurate than assignment, suggesting that the differences in performance between agreement and assignment cannot be explained by one variable alone, but rather are the result of several

factors together. Due to statistical limitations no analyses could be carried out in the delayed post-tests.

In brief, in general the results align with previous research showing that assignment has higher rates in accuracy than agreement (e.g. Montrul et al., 2008). Nevertheless, when taking into account all variables (concordance, gender, noun ending, and treatment) the reverse pattern arises, suggesting that when all these aspects are factored in, agreement outperforms assignment, which is a finding contrary to previous literature. This finding could be the byproduct of the decisions made with regard to coding and/or due to the inclusion of invariable adjectives, which might have caused an overestimation of the ability learners had to make article, nouns and adjectives agree.

Differences Among Noun Endings

As was the case regarding accuracy in masculine and feminine nouns, there is a constant pattern in accuracy of gender assignment and agreement depending on noun ending. As expected, and in line with previous research (for an example see Montrul et al., 2008), accuracy was higher in nouns with canonical endings, followed by non-canonical and exceptions. This pattern held across interactions and research questions. This suggests that the treatment did not help level accuracy across noun endings; a possibility might be that more exposures are necessary in order to facilitate the learning of non-canonical and exceptional noun endings. During the delayed post-test, nouns with a canonical ending received significantly higher accuracy scores, as reflected in the significant interaction of Test*Ending both in assignment and agreement. These findings support Zandieh & Jafarigohar's (2012) claims that the benefits from incidental learning take more time to show. Statistical results also showed that exceptional endings received significantly lower

accuracy scores. Thus, the treatment might have facilitated sensibility to the grammatical gender system of Spanish, but only towards its canonical features, which is a specification that was not provided in the literature; previous research had shown that advantages for incidental learning of grammatical features appeared after some time in the delayed post-test (e.g. Grey et al., 2014; Spada & Tomita, 2010; Zandieh & Jafarigohar, 2012). The literature nevertheless shows that such delayed effects are not available with regard to all features; this could be accounted for by DeKeyser (1995), who argues that clear-cut rules benefit from explicit instruction but fuzzy rules benefit from implicit inductive instruction. Arguably, in the current study, Spanish grammatical gender might be interpreted as a hybrid, although interestingly, the “fuzziest” of its variables are the ones that did not show the best results after the treatment.

The decrease in the accuracy of exceptions suggests that learners did not learn the specific items they were presented, but rather the pattern, which would explain why they would mistakenly assign the opposite gender to exceptions, which is counterintuitive per se. In addition, as reported in the description of the Spanish grammatical gender system in Chapter I, some of the nouns with exceptional endings might have posed an issue given that they had biological gender as well. This, as anticipated, could have been a confounding factor. Nevertheless, as Bordag et al. (2017) showed, even native speakers can be primed with wrongfully assigned gender, thus, learners should also be susceptible to it.

In summary, the tendency across the board is for nouns with canonical ending to display more accuracy, followed by non-canonical and exceptional endings. In the delayed post-test a clear significant improvement is found in performance of canonical nouns, however,

a significant decrease in accuracy appears also in exceptions, while non-canonical nouns remain the same statistically.

Differences Between New and Old Noun-Adjective Pairings

When looking into the new and old noun-adjective pairings no statistical difference was found.. The lack of significance could indicate one of two things (1) participants did not learn in chunks (Tipurita & Jean, 2014; Denhovska & Serratrice, 2017) or (2) participants just did not pay attention or parse the adjectives. The first option would show a certain degree of generalizability among known items, which would be a success for learners. There was indeed some degree of sensibility, as groups with the bold font enhancement performed slightly better and with regard to agreement all groups outperformed the AN group. The second case seems more likely, as it would explain the significant difference observed between ANA and AN, neither of which have the enhancement. However, the ANA group performed similarly to the groups with enhancement, suggesting that the presence or absence of adjectives played a role. In summary, results are inconclusive in this regard.

Correlation of Variables

Results showed that attention and recognition might be necessary for learning marked features, as there was a tendency to find correlations between performance in the marked features of grammatical gender (feminine nouns, and non-canonical and exceptional endings) and proficiency, and the answer sheet and the form recognition questionnaires. Not as many interactions were present in the delayed post-test (contra Grey et al., 2014; Spada & Tomita, 2010). Similar trends were also found for learning of nouns and adjectives. In addition, proficiency also correlated with higher accuracy rates. This

suggests that attention and recognition are necessary, as well as meaning, and sufficient for passive recall of meaning. It might be the case, though, that form-meaning connections were weaker due to the low working memory abilities of the population as shown in their inability for active recall during the pilot, which would be in line with Malone (2018). The inability for active recall might also be explained by Godfroid et al.'s (2017) findings, which suggest that frequency affects form recognition, but that in order to build form-meaning connections deliberate focus is necessary and time spent on learning the word has more explanatory power.

Learning of Adjectives and Nouns and Frequency Effects

As seen in the results, no significant differences were found between the learning of nouns and adjectives nor between the frequency of the adjectives learners were exposed to. The only differences between treatments showed that the control group outperformed the rest. This is an odd result, considering that frequency is the factor that gathers the most evidence regarding its role in language learning (see Rott, 1999; Teng, 2016; Hucking & Coady, 1999; Teng, 2016; Pellicer-Sánchez, 2015). Why did frequency not render any differences? A hypothesis formulated after consulting with native speakers of English and looking at possible translations is that there were cognates available that might have aided learning and thus cancelled out the effects of the enhancement. While the cognates were accounted for beforehand, the pre-tests showed the words were unknown to the learners and translations prioritized non-cognates. It was therefore expected that the translation they were primed with would prevail given the speed of the treatment; but this might not have been the case. In the future, research should account for the fact that if there is an available

cognate, even when not presented among the stimuli, it might affect participants' performance.

In reference to enhancement, we have repeatedly seen that it provided positive results regarding grammatical gender acquisition, meaning it did affect learners. Why then did it not affect learning? To begin with, learners were told to pay attention to the vocabulary, which means they were learning explicitly and intentionally. Research has shown that the learning that most benefits from this kind of enhancement (bold font in this case) is incidental. Thus the success of explicit learning reported on by research might not lay on its explicitness, but rather on the learners' attention being directed to the target features (e.g. Ellis et al., 2006; Sanz & Morgan-Short, 2004); and it may be that explicit intentional learning might not work if learners pay attention to the wrong feature. It could be ventured that the same hypothesis holds for frequency regarding its ineffectiveness in implicit intentional learning. Besides this theory, as mentioned, there could not have been any learning occurring and results be based on participants' guessing in the multiple-choice translation test that was used to measure their learning of nouns and adjectives. Finally, the last possible explanation could be that the students tested for knowledge of the treatment items in the pre-test, before creating the final treatment materials, were either below the bar in comparison to these groups or that the vocabulary was taught during the semester. Nevertheless, that explanation is not quite plausible, since most of the words do not belong to the thematic units of the class. Future research should aim to pre-test one level above the targeted level in order to make sure items are completely novel to participants.

In short, results showed no significant differences between accurate translation of nouns and adjectives and it also did not reflect any frequency effects, the only significant

differences showed that the control group (who had not been exposed to any stimuli) outperformed the rest. It is hypothesized that results depended upon proficiency and the learners might have benefitted from cognates to perform well or that frequency and the bold font enhancement do not render learning benefits in implicit intentional learning. It may also be argued that the tasks were not engaging enough and that that might account for part of the lack of results, which would be in line with the Involvement Load Hypothesis. However, the current data did not compare performance across tasks with different degrees of involvement and it remains unknown whether results would improve with them, so the alignment with the ILH is just a speculation.

Ultimate Attainment

Results suggest that the treatment did not bring learners closer to native-like performance, but rather that proficiency was the best metric to predict closeness to ultimate attainment. As seen in the literature review, proficiency is indeed the best predictor for performance. The differences depending on proficiency seem to indicate that improvement is possible, suggesting that eventual ultimate attainment is possible behaviorally, provided that learners do not fossilize at a certain stage (nothing can be said as far as processing since the current data did not look into that). The way the data was coded required that learners be consistent in their assignment and agreement, which means that the error rates represent either inconsistent assignment/agreement or wrongful consistent assignment/agreement. A qualitative look at the data does suggest that there were indeed systematic errors in assignment and agreement, which according to Santoro (2012) is in line with Lardiere's Feature Reassembly Hypothesis, since these errors do not reflect a

misrepresentation of the system, but rather a deficiency in the assembly of the gender feature of words.

Regarding the storing of knowledge as declarative or procedural only speculations can be put forward. If we were to assume that the knowledge documented in the delayed post-test is reflective of what has been interiorized and thus stored in the procedural memory and that such behavior is the by-product of the treatment, it seems as though only unmarked endings, canonical endings, became automatized, while the rest of the system would still be stored as factual knowledge and more susceptible to variation. In line with Ullman's (2001) Declarative/Procedural model it is then likely that the treatment might have aided in the proceduralization of unmarked noun endings.

In conclusion, results indicate that proficiency is the best predictor of attainment and that the treatment did not seem to bring learners significantly closer to native-like performance. Regarding SLA theories, the current results seem to align with the predictions made by the Feature Reassembly Hypothesis and to posit tentative evidence towards the Declarative/Procedural model. Results towards the DP model would only suggest that the unmarked noun endings were proceduralized, but not other features.

Chapter VII: Conclusions, Limitations and Future Research

This project aimed to research the viability of a new pedagogical approach to facilitate the acquisition of the Spanish grammatical gender system through incidental learning. Five groups were compared, a control, two groups with bold font enhancement of the target features and two groups without it. Of the two groups with and without enhancement one also had nouns agreeing with adjectives and one did not. The expectation was that the group with bold font enhancement and exposed to the article-noun-adjective combination (ANAB) would render better accuracy than the rest since their attention was directed to the feature and they were exposed to more of the system, its assignment and agreement, through the agreement with adjectives.

The results did not statistically support the expectation that the ANAB group would outperform the rest in all conditions; nevertheless, some tendencies that supported this expectation were present in the immediate post-tests and the delayed post-test regarding assignment by gender. The tendency found among groups in the immediate post-test was C>ANAB>ANA>ANB>AN, while in the delayed post-test we found ANB>ANAB>AN>ANA. Statistical significance was present in some of the inquiries, only with the ANAB group significantly outperforming the AN (article-noun, no enhancement) group, which is interpreted as there being a necessity of both, the enhancement and the increased exposure to agreement through the presence of an adjective, in order for there to be a significant improvement between groups. The pattern does not hold for the delayed post-test when looking into agreement by noun gender and both assignment and agreement regarding noun ending. In that case it shows that the ANB (article-noun, enhanced) group outperforms the ANA (article-noun-adjective, no

enhancement) group, suggesting that, two weeks after treatment, learners actually benefitted the most from the bold font enhancement and the presence of the adjectives worsened their performance. This was interpreted as a consequence of the effects of working memory (Sagarra & Herschensohn, 2012), which is hypothesized was overloaded given that learners in the noun-adjective groups (ANAB and ANA) were exposed to 70 novel items versus the 30 the article-noun groups (ANB and AN) were taught. This result aligns with DeKeyser's (2008) argument that incidental learning might require (1) more time for learners to acquire the feature and (2) more time for the effects of learning to show (in line with Zandieh & Jafarigohar, 2012). In general terms, the delayed post-test showed increased performance with regard to the immediate post-test. A Bonferroni correction revealed that significant interactions affected the performance on nouns with a canonical ending, improving it, while the accuracy with regard to exceptions decreased and for non-canonical nouns stayed the same. This suggests that unmarked items might be transitioning from declarative to procedural memory, which would provide further evidence for Ullman's (2001) Declarative/Procedural model.

Previous research by Montrul et al. (2008) has shown that performance is higher when it comes to assignment over agreement and for masculine nouns over feminine nouns. Accuracy for agreement was however higher than for assignment when factoring in treatment, noun ending, gender of the noun and concordance. Proficiency was the best predictor when it came to accuracy, which suggests that ultimate attainment is possible. Given how accuracy was scored—that consistent gender assignment and agreement needed to happen for their performance to be considered correct—and given the presence of consistent misassignment of gender to nouns with matching agreement, it is argued that

the current data provide support for Lardiere's (2008, 2009) Feature Reassembly Hypothesis (FRH).

From a more practical standpoint, pedagogically speaking, it seems that the only modification made in the experiment that rendered definitive results was the use of bold font, while the presence or absence of adjectives had a more variable role. Adjectives played an unstable role, suggesting that they improved the results in the immediate post-test but turned out to have hindering effects in the delayed post-test two weeks later. These effects are only arguable regarding visual enhancement, as no effects of aural enhancement were considered; the voice over had no specific features for enhancement.

In conclusion, the treatments proposed for incidental acquisition of the Spanish gender system did not seem to provide statistically relevant results. However, there were some tendencies suggesting that both enhancement via bold font and noun-adjective presence were necessary for obtaining statistically significant results. In the delayed post-test only the bold font enhancement seemed relevant while the presence of the adjective seemed to be hindering. Thus, it can be concluded that these results provide tentative support for the Declarative/Procedural Model and the Feature Reassembly Hypothesis. On visual enhancement, the presence of bold font is a clear facilitator, while the presence of an adjective is helpful in the immediate results but hindering two weeks later.

Limitations and future research

There were several shortcomings in this research derived mainly from time constraints and the scope of the project. These limitations, with possible ways of addressing them, are offered next. Despite the research design allowing for further inquiries, time did not allow for looking at the accuracy in the use of adjectives, namely whether learners forced

agreement upon invariable adjectives and vice versa, and the ability to generalize to known and unknown nouns. However, the data to explore such questions was gathered during the delayed post-test and is available. Time constraints also did not allow for measurement of working memory, which considering the number of items in the treatment that included adjectives (70), would have been a likely explanation of the differences with the non-adjectives group, which had only 30 items to become acquainted with. Future research should attempt to replicate the study with only the nouns in order to lower the cognitive load and increase comparability between groups. In addition and in retrospect, the non-adjective groups maybe should have been tested for adjectives. The constraints on time did not allow for creating a test for adjectives and nouns, which would have facilitated testing both, as two separate tests with a combined total of 70 items would have needed, versus the 40-item test that was used. Knowledge of nouns and adjectives should also have been measured in the delayed post-test so correlations between item knowledge, assignment and agreement could be measured address whether knowledge of the noun is necessary.

The debriefing questionnaire might have benefitted from asking (1) about the knowledge learners had of the Spanish grammatical gender system and (2) asking more explicitly about any patterns regarding grammatical gender assignment. This was not taken into account, as the priority was to observe whether participants noticed the patterns or not and it was assumed that, according to the curriculum at the institution, learners had received explicit instruction of the Spanish grammatical gender at previous levels. Nevertheless, it might have been relevant in order to explain our results since awareness and explicit knowledge of rules has played a role in some studies (e.g. Leow, 1998). Unfortunately, that would have required yet a separate test in order for them not to change their questions to

the debriefing questionnaire that was provided, which would have increased the time demands of the experiment on the students.

Another aspect to control for in the future is the comparability of groups, considering that the control group, which did not receive any treatment, had significantly higher proficiency and outperformed most treatment groups in most assessments. Another bias in this research is a recurrent one in SLA, common when testing undergraduate college populations. Unaccounted for factors such as motivation, might have played a significant role, since at this institution this is the last course students are required to take to fulfill the language requirement and that might have caused them to not pay as much attention as desired.

Finally, seeing the results and having spotted participants filling the items for the determiners and adjectives separately, meaning that they first chose all the determiners and then all the adjectives, it might have been a better choice to test participants in the context of a reading passage. Moreover, the context provided in the sentences might have not been meaningful enough causing the lack of results. This may suggest the necessity of meaningfulness in the contexts provided for incidental learning, as despite enhancement, mere exposure to agreement and novel items did not suffice for acquisition of the Spanish grammatical gender system nor the vocabulary and adjectives. Future research should thus (1) ensure the comparability of treatment and control groups, (2) explore gender systems with various degrees of transparency, (3) address generalizability to novel and unknown items as well as the accuracy of agreement in other linguistic items such as demonstratives, possessives, pronouns, etc., and (4) consider the use of texts with meaningful contexts for the target words both, during treatment and testing. Lastly, despite the inconclusiveness of

this research, it has provided useful information for future research into the important questions that the literature has raised about the acquisition of assignment and agreement in Spanish and how learning can be facilitated.

APPENDIX A

This appendix presents the list of nouns and adjectives that were used in the study. Some of the available cognates for the adjectives are only semantically related, but they might have been triggered upon seeing the Spanish adjectives due to the amount of overlap and aided learning due to the semantic relation.

Table A1
Nouns used in the experiment by gender and noun ending

Gender	Ending	Noun	Translation	
Masculine	Canonical	El cuerno	<i>The horn</i>	
		El teclado	<i>The keyboard</i>	
		El delito	<i>The felony</i>	
		El anillo	<i>The ring</i>	
		El vuelo	<i>The flight</i>	
	Non-canonical	El mensaje	<i>The message</i>	
		El hogar	<i>The home</i>	
		El bosque	<i>The forest</i>	
		El corte	<i>The cut</i>	
		El bigote	<i>The moustache</i>	
	Exceptions	El cura	<i>The priest</i>	
		El fantasma	<i>The ghost</i>	
		El hematoma	<i>The bruise</i>	
		El pesticida	<i>The pesticide</i>	
		El tenista	<i>The tennis player</i>	
	Feminine	Canonical	La sonrisa	<i>The smile</i>
			La beca	<i>The grant</i>
La medalla			<i>The medal</i>	
La cacerola			<i>The saucepan</i>	
La boda			<i>The wedding</i>	
Non-canonical		La cumbre	<i>The peak</i>	
		La torre	<i>The tower</i>	
		La pared	<i>The wall</i>	
		La calle	<i>The street</i>	
		La población	<i>The population</i>	
Exceptions		La soprano	<i>The soprano</i>	
		La moto	<i>The motorbike</i>	
		La biblio	<i>The library</i>	
	La seo	<i>The old cathedral</i>		
		La modelo	<i>The model</i>	

Table A2

Adjectives used in the experiment by variability and available cognates

Variability	Adjective	Frequency	Treatment Translation	Available cognate	
Variable	Abollado	4	<i>Dented</i>	-	
	Afectuoso	2	<i>Loving</i>	<i>Affectionate</i>	
	Afilado	2	<i>Sharp</i>	-	
	Amoroso	2	<i>Loving</i>	<i>Amorous</i>	
	Barato	4	<i>Cheap</i>	-	
	Caro	2	<i>Expensive</i>	-	
	Comprensivo	2	<i>Understanding</i>	-	
	Envuelto	2	<i>Wrapped</i>	-	
	Específico	4	<i>Specific</i>	<i>Specific</i>	
	Flaco	2	<i>Lean</i>	-	
	Flojo	4	<i>Weak</i>	-	
	Inflado	4	<i>Swollen</i>	<i>Inflated</i>	
	Ligero	4	<i>Light</i>	-	
	Lujoso	4	<i>Luxurious</i>	<i>Luxurious</i>	
	Peligroso	4	<i>Dangerous</i>	<i>Perilous</i>	
	Poroso	2	<i>Porous</i>	<i>Porous</i>	
	Pulcro	4	<i>Neat</i>	-	
	Rocamboloso	2	<i>Bizarre</i>	-	
	Soso	2	<i>Dull</i>	-	
	Vacío	4	<i>Empty</i>	-	
	Invariable	Agradable	2	<i>Pleasant</i>	<i>Agreeable</i>
		Agrícola	4	<i>Agricultural</i>	<i>Agricultural</i>
		Culpable	2	<i>Guilty</i>	<i>Culpable</i>
		Decepcionante	4	<i>Disappointing</i>	-
		Desagradable	4	<i>Unpleasant</i>	<i>Disagreeable</i>
		Enorme	2	<i>Enormous</i>	<i>Enormous</i>
		Formidable	4	<i>Terrific</i>	<i>Formidable</i>
		Grave	4	<i>Serious</i>	<i>Grave</i>
		Impermeable	2	<i>Waterproof</i>	<i>Impermeable</i>
		Impracticable	2	<i>Impassable</i>	<i>Impracticable</i>
		Impresionante	4	<i>Impressive</i>	<i>Impressive</i>
		Indemne	2	<i>Unharmful</i>	-
Infantil		4	<i>Childish</i>	<i>Infantile</i>	
Inolvidable		4	<i>Unforgettable</i>	-	
Insoportable		2	<i>Unbearable</i>	<i>Unsupportable</i>	
Pobre		4	<i>Poor</i>	-	
Progresista		2	<i>Progressive</i>	<i>Progressist</i>	
Sorprendente		4	<i>Surprising</i>	-	
Sostenible		2	<i>Sustainable</i>	<i>Sustainable</i>	
Vigente		2	<i>Current</i>	-	

APPENDIX B

Appendix B provides access to the different treatments that were administered to the learners. Given that each treatment is composed of 101 slides and considering that they would take up unnecessary space and that they have audio files attached, it was considered more appropriate to upload the different treatments on Filebin and provide a link to the RAR file created. For safe keeping, the file is protected with a password, which is SpGrammGendTreatments. The link where they can be found is: <https://filebin.net/8ke8n7hwcm1td4aj>

As explained in the methodology, all pictures came from Pixabay and Google images free of copyright. In occasion, pictures were modified to best match the sentences they accompanied.

APPENDIX C

Appendix C provides all of the forms and tests used during the treatment and testing sessions in the same order as they were administered. All documents were provided to all groups, except for the Answer Sheet and the Form Recognition tests, and the Debriefing Questionnaire, which were not given to the control group. In order to keep the materials as similar to those presented to the participants they are all in separate pages.

Consent Form

Whom to Contact about this study:

Principal Investigator: Mireia Toda Cosi

Department: Department of Modern Languages, Linguistics & Intercultural
Communication

Telephone number: (XXX) XXX-XX67

Vocabulary Learning Treatment Protocol

I. INTRODUCTION/PURPOSE:

I am being asked to participate in a research study. The purpose of this study is to explore enhanced ways to learn Spanish vocabulary. I am being asked to volunteer because I am taking Spanish 201 at UMBC. My involvement in this study will begin when I agree to participate and will continue until the end of the semester. About 130 persons will be invited to participate.

II. PROCEDURES:

As a participant in this study, I will be asked to sit through a PowerPoint presentation and fill in some multiple-choice questionnaires. I will be asked to come to my regular UMBC classes or after class if I am participating outside my classroom environment. My participation in this study will last for two visits. The first session will last for 40-50 minutes and the second one will last for 20-30 minutes. No personal identifying information will appear in the study.

III. RISKS AND BENEFITS:

My participation in this study does not involve any significant risks and I have been informed that my participation in this research will not benefit me personally, but I might learn some vocabulary and help develop better vocabulary teaching materials. I have been informed that participation in this study may involve a moderated risk of tiredness. I have also been informed that my participation in this research will not benefit me personally, but will help improve current Spanish teaching materials.

IV. CONFIDENTIALITY:

Any information learned and collected from this study in which I might be identified will remain confidential and will be disclosed ONLY if I give permission. The investigator (s) will attempt to keep my personal information confidential. To help protect my confidentiality, all participants there will be represented by a numerical ID. Upon completion I will preserve the data stored securely in their hard-copy format.

Only the investigator and members of the research team will have access to these records. If information learned from this study is published, I will not be identified by name. By signing this form, however, I allow the research study

investigator to make my records available to the University of Maryland Baltimore County (UMBC) Institutional Review Board (IRB) and regulatory agencies as required to do so by law.

Consenting to participate in this research also indicates my agreement that all information collected from me individually may be used by current and future researchers in such a fashion that my personal identity will be protected. Such use will include presentations at scientific or professional meetings, publishing in scientific journals, sharing anonymous information with other researchers for checking the accuracy of study findings and for future approved research that has the potential for improving human knowledge.

V. COMPENSATION/COSTS:

My participation in this study will involve no cost to me. I will be provided extra credit in my Spanish class as well as light refreshments.

VI. CONTACTS AND QUESTIONS:

The principal investigator(s), Mireia Toda Cosi has offered to and has answered any and all questions regarding my participation in this research study. If I have any further questions, I can contact Mireia Toda Cosi at mire1@umbc.edu or (XXX) XXX-XX67.

If I have any questions about my rights as a participant in this research study, contact the Office of Research Protections and Compliance at (410) 455-2737 or compliance@umbc.edu.

VII. VOLUNTARY PARTICIPATION

I have been informed that my participation in this research study is voluntary and that I am free to withdraw or discontinue participation at any time I have been informed that data collected for this study will be retained by the investigator and analyzed even if I choose to withdraw from the research. If I do choose to withdraw, the investigator and I have discussed my withdrawal and the investigator may use my information up to the time I decide to withdraw. My withdrawal will not impact my grade in the class under any circumstances

I will be given a copy of this consent form to keep.

VII. SIGNATURE FOR CONSENT

The above-named investigator has answered my questions and I agree to be a research participant in this study.

Participant's Name: _____ Date: _____

Participant's Signature: _____ Date: _____

Investigator's Signature: _____ Date: _____

Lextale and Linguistic Background Questionnaire

Vocabulary Test Spanish

Hi, this is a test of Spanish vocabulary. On the next page you will find 90 sequences of letters that look “Spanish”. Only some of them are real words. Please, indicate the words you know (or of which you are convinced they are Spanish words, even though you would not be able to give their precise meaning). Be careful, however: Errors are penalised. So, there is no point in trying to increase your score by adding tallies to “words” you’ve never seen before!

All you have to do is to tick the box next to the words you know. If, for instance, in the example below you recognise “sí”, “sacapuntas”, “bien”, and “casa”, you indicate this as follows:

Estímulo	Palabra?	Estímulo	Palabra?
depiste		priba	
sí	✓	pelasula	
coné		bien	✓
calpar		casa	✓
joten		lejo	
sacapuntas	✓	pretantas	

- Name: _____
 - Age: _____
 - Gender: male / female
 - Do you have a learning disability, condition or any kind of hearing or sight impairment that is not medicated or corrected? _____
 - Native language (language/s you learnt from your parents) _____
 - Do you speak other languages besides English? _____
 - If yes, what languages? _____
 - Do you practice or are exposed to Span outside of the classroom or your assignments? _____
 - Please, explain, how many hours a week? What kind of exposure? (i.e. I watch two hours of telenovelas a week, I listen to 3 hours of Spanish music on the car, etc.) _____
-
- How many years of Spanish instruction did you get before coming to UMBC? _____

Turn the page to start the test. Many thanks in advance!

Please return the test to the experimenter.

Estímulo	Palabra?
terzo	
pellizcar	
pulmones	
batillón	
zapato	
tergiversar	
pésimo	
cadeña	
hacha	
antar	
cenefa	
asesinato	
helar	
yunque	
regar	
abracar	
floroso	
arsa	
brecedad	
ávido	
capillo	
lacayo	
lampera	
látigo	
bisagra	
secuestro	
acutación	
merodear	
decar	
alardio	

Estímulo	Palabra?
pandilla	
fatacidad	
pauca	
aviso	
rompido	
loro	
granuja	
estornudar	
torpe	
alfombra	
rebuscar	
cadallo	
canela	
cuchara	
jilguero	
martillo	
cartinar	
ladrón	
ganar	
flamida	
candado	
camisa	
vegada	
fomentar	
nevar	
musgo	
tacaño	
plaudir	
besar	
matar	

Estímulo	Palabra?
seda	
flaco	
esposante	
orgullosa	
bizcocho	
hacido	
cabello	
alegre	
engatusar	
temblo	
polvoriento	
pemición	
hervidor	
cintro	
yacer	
atar	
tiburón	
frondoso	
tropaje	
hormiga	
pozo	
empirador	
guante	
escudo	
laúd	
barato	
grodo	
acantilado	
prisa	
clavel	

Name: _____

Answer Sheet¹⁵

Example:

El tomate rojo está encima del queso

A. B. C. D.

El tomate verde se come con ensaladas

A. B. C. D.

- | | |
|---|--|
| 1. El delito grave que cometió era un robo | 14. La medalla ligera es de lata |
| A. B. C. D. | A. B. C. D. |
| 2. El delito peligroso acabó en un incendio | 15. La sonrisa agradable de la joven con el ipad me gustó |
| A. B. C. D. | A. B. C. D. |
| 3. El teclado impermeable se puede mojar | 16. La sonrisa sosa de la chica no me gustaba |
| A. B. C. D. | A. B. C. D. |
| 4. El teclado envuelto es mi regalo | 17. La boda fue desagradable porque los novios no se querían |
| A. B. C. D. | A. B. C. D. |
| 5. El vuelo emocionante de ayer me sorprendió | 18. La boda barata se hizo en la ciudad |
| A. B. C. D. | A. B. C. D. |
| 6. Había muchas opciones, pero cogí el vuelo barato | 19. La cacerola vacía no tenía comida |
| A. B. C. D. | A. B. C. D. |
| 7. Puedes beber con el cuerno poroso | 20. La cacerola impresionante tiene mucha comida |
| A. B. C. D. | A. B. C. D. |
| 8. El animal seguía con el cuerno indemne | 21. El hogar rocambolés pertenece a la bruja |
| A. B. C. D. | A. B. C. D. |
| 9. No se podía poner el anillo abollado | 22. El hogar sostenible protege el medioambiente |
| A. B. C. D. | A. B. C. D. |
| 10. El anillo impresionante de la novia era de su madre | 23. El bosque decepcionante está cerca de la ciudad |
| A. B. C. D. | A. B. C. D. |

¹⁵ This Answer Sheet was given to the ANA group, all sentences received the enhancement or showed the appropriate agreement depending on the group.

11. Le dieron una beca específica para estudiar arte
A. B. C. D.
12. Solo dan la beca agrícola a chicos
A. B. C. D.
13. Una granja recibe la medalla agrícola
A. B. C. D.
27. El corte inflado se infectó
A. B. C. D.
28. Tuvo que ir al hospital por el corte grave
A. B. C. D.
29. Todos han visto el bigote sorprendente de Dalí
A. B. C. D.
30. Un bigote pulcro debe recortarse cada día
A. B. C. D.
31. Subo a la cumbre afilada
A. B. C. D.
32. Es imposible subir a la cumbre impracticable
A. B. C. D.
33. Como no tenía comida la población estaba floja
A. B. C. D.
34. Necesitamos más escuelas para la población infantil
A. B. C. D.
35. Había muchos dibujos en la pared sorprendente
A. B. C. D.
36. Esa pared específica tiene un mosaico
A. B. C. D.
37. La torre enorme está en la derecha
A. B. C. D.
24. El bosque vacío será reforestado
A. B. C. D.
25. Mi madre me dio el mensaje amoroso
A. B. C. D.
26. "El mensaje vigente" es el título del libro
A. B. C. D.
44. El tenista inolvidable ganó
A. B. C. D.
45. El hematoma desagradable se va con hielo
A. B. C. D.
46. El hematoma está inflado por el golpe
A. B. C. D.
47. El cura culpable fue a la cárcel
A. B. C. D.
48. El cura comprensivo le perdonó
A. B. C. D.
49. El pesticida peligroso no se puede usar
A. B. C. D.
50. El pesticida formidable funciona muy bien
A. B. C. D.
51. La soprano progresista no apoya a Trump
A. B. C. D.
52. La soprano flaca canta muy bien
A. B. C. D.
53. Tras el accidente la moto abollada terminó en la basura
A. B. C. D.
54. La moto es decepcionante en comparación con el avión para viajes largos
A. B. C. D.

38. La torre cara es más alta que la casa
A. B. C. D.
39. No hay basura en la calle pulcra
A. B. C. D.
40. En la calle pobre la gente no tiene casas
A. B. C. D.
41. El fantasma insoportable está en la habitación
A. B. C. D.
42. El fantasma afectuoso está en la cocina
A. B. C. D.
43. El tenista flojo no terminó
A. B. C. D.
55. La biblio pobre tiene pocos libros
A. B. C. D.
56. La biblio lujosa tiene los mejores libros
A. B. C. D.
57. La seo formidable está en la derecha
A. B. C. D.
58. La seo lujosa está en la izquierda
A. B. C. D.
59. La modelo ligera pesa 90 libras
A. B. C. D.
60. La modelo infantil tiene 5 años
A. B. C. D.

Form Recognition Test (ANAB and ANA Groups)

Circle or underline **ONLY** the words that you **HAVE** seen during the PowerPoint.

sonrisa	culpable	guajalote	específico
apogeo	querella	ataúd	veleidad
bulto	afectuoso	biblio	estribación
obispo	insoportable	anillo	agrícola
cumbre	grieta	terratendiente	hélice
truhan	huelga	cacerola	ligero
mensaje	marea	navegador	vacío
güisqui	poroso	bogavante	lombriz
colmillo	indemne	corte	decepcionante
criado	ciudadanía	bigote	anatema
cura	envuelto	calle	hacha
cervatillo	impermeable	enjambre	sofocando
fantasma	rocambolesco	pista	sorprendente
cuerno	tumba	seo	inflado
mendigo	sostenible	balsa	domiciliario
teclado	cara	pesticida	rocoso
casco	sepultura	eyección	desagradable
respaldo	morcilla	sucursal	desamparado
hogar	enorme	vuelo	abollado
charco	sastre	alcalde	cotidiano
torre	flaca	modelo	súbito
soprano	progresista	azufre	repentino
sosa	delito	rebrote	lujoso
agradable	declive	auge	pobre
aldeano	tifón	boda	halagüeño
afilada	beca	población	costoso
impracticable	calamar	caridad	impresionante
respingón	medalla	ceguedad	exótico
lino	bosque	tenista	cerrojo
amoroso	pared	servidumbre	pulcro
sotana	buque	peligroso	formidable
vigente	hematoma	grave	bélico
comprensivo	moto	peste	barato

Form Recognition Test (ANB and AN Groups)

Circle or underline **ONLY** the words that you **HAVE** seen during the PowerPoint.

sonrisa	ataúd
apogeo	biblio
cumbre	anillo
mensaje	terrateniente
güisqui	cacerola
criado	navegador
cura	bogavante
cervatillo	corte
fantasma	bigote
cuerno	calle
mendigo	enjambre
teclado	pista
casco	seo
respaldo	balsa
hogar	pesticida
charco	eyección
torre	sucursal
soprano	vuelo
aldeano	alcalde
sastre	modelo
delito	azufre
declive	auge
tifón	boda
beca	población
calamar	caridad
medalla	ceguedad
hélice	tenista
bosque	servidumbre
pared	
buque	
hematoma	
moto	
guajalote	

Multiple Choice Translation Test (ANAB, ANA, and Control Groups)

Please, underline the right English translation to the given Spanish translation just like in the example. Be attentive, as a last resort, if you really don't know the words, select "I don't know", but that must be the very last resort.

Example: *El tomate rojo está encima de la comida*

- a) *The red apple is on the food*
- b) **The red tomato is on the food**
- c) *The green apple is on the food*
- d) *The green tomato is on the food*
- e) *I don't know*

- | | |
|---|--|
| 1) Mi madre me dio el mensaje soso | e) I don't know |
| a) My mom gave me the loving message | |
| b) My mom gave me the dull message | 5) Solo dan la beca agrícola a chicos |
| c) My mom gave me the loving letter | a) They only give that specific grant to boys |
| d) My mom gave me the dull letter | b) They only give that specific loan to boys |
| e) I don't know | c) They only give the agricultural grant to boys |
| 2) La sonrisa afectuosa de la chica no me gustaba | d) They only give the agricultural loan to boys |
| a) I didn't like the affectionate lips of the girl | e) I don't know |
| b) I didn't like the pleasant lips of the girl | |
| c) I didn't like the affectionate smile of the girl | 6) La medalla estaba vacía de significado |
| d) I didn't like the pleasant smile of the girl | a) The medal was empty of meaning |
| e) I don't know | b) The medal was full of meaning |
| 3) El delito específico acabó en un incendio | c) The trophy was empty of meaning |
| a) The specific felony ended up in a fire | d) The trophy was full of meaning |
| b) The specific robbery ended up in a fire | e) I don't know |
| c) The dangerous felony ended up in a fire | 7) El bosque pobre será reforestado |
| d) The dangerous robbery ended up in a fire | a) The poor forest will be reforested |
| e) I don't know | b) The rich forest will be reforested |
| 4) La torre impracticable es de Rapunzel | c) The poor jungle will be reforested |
| a) Rapunzel's tower is impassable | d) The rich jungle will be reforested |
| b) Rapunzel's tower is expensive | e) I don't know |
| c) Rapunzel's castle is impassable | 8) Esa pared barata tiene un mosaico |
| d) Rapunzel's castle is expensive | a) That specific wall has a mosaic |
| | b) That cheap wall has a mosaic |

- c) That specific ceiling has a mosaic
d) That cheap ceiling has a mosaic
e) I don't know
- 9) La soprano progresista no apoya a Trump
a) The talkative contralto doesn't support Trump
b) The progressive contralto doesn't support Trump
c) The talkative soprano doesn't support Trump
d) The progressive soprano doesn't support Trump
e) I don't know
- 10) El fantasma comprensivo está en la cocina
a) The affectionate ghost is in the kitchen
b) The affectionate spirit is in the kitchen
c) The understanding ghost is in the kitchen
d) The understanding spirit is in the kitchen
e) I don't know
- 11) El teclado sostenible se carga con energía solar
a) The sustainable board charges with solar energy
b) The unusable board charges with solar energy
c) The sustainable keyboard charges with solar energy
d) The unusable board charges with solar energy
e) I don't know
- 12) El hematoma está inflado por el golpe
a) The swollen bruise was because he got hit
b) The swollen wound was because he got hit
c) The serious bruise was because he got hit
d) The serious wound was because he got hit
e) I don't know
- 13) La moto es ligera en comparación con el avión
a) The motorcycle is disappointing in comparison to the plane
b) The bicycle is light in comparison to the plane
c) The bicycle is disappointing in comparison to the plane
d) The motorcycle is light in comparison to the plane
e) I don't know
- 14) La biblio lujosa tiene los mejores libros
a) The poor library has the best books
b) The poor bookstore has the best books
c) The luxurious bookstore has the best books
d) The luxurious library has the best books
e) I don't know
- 15) El anillo sorprendente de la novia era de su madre
a) The bride's surprising ring was her mother's
b) The bride's surprising bracelet was her mother's
c) The bride's modest ring was her mother's
d) The bride's modest bracelet was her mother's
e) I don't know
- 16) La cacerola impresionante tiene mucha comida
a) The impressive saucepan has lots of food
b) The impressive pot has lots of food
c) The disposable saucepan has lots of food
d) The disposable pot has lots of food
e) I don't know

- 17) Tuvo que ir al hospital por el corte desagradable
- He had to go to the hospital because of the serious wound
 - He had to go to the hospital because of the unpleasant wound
 - He had to go to the hospital because of the unpleasant cut
 - He had to go to the hospital because of the serious cut
 - I don't know
- 18) Un bigote pulcro debe recortarse cada día
- A neat beard should be trimmed every day
 - A messy beard should be trimmed every day
 - A neat moustache should be trimmed every day
 - A messy moustache should be trimmed every day
 - I don't know
- 19) No hay basura en la calle decepcionante
- There is no trash in the disappointing avenue
 - There is no trash in the disappointing street
 - There is no trash in the dangerous avenue
 - There is no trash in the dangerous street
 - I don't know
- 20) La seo formidable está en la derecha
- The terrific old church is on the right
 - The awful old church is on the right
 - The terrific old cathedral is on the right
 - The awful old cathedral is on the right
 - I don't know
- 21) Ese pesticida abollado no se puede usar
- That specific herbicide can't be used
 - That specific pesticide can't be used
 - That dented herbicide can't be used
 - That dented pesticide can't be used
 - I don't know
- 22) Ayer saltamos del avión, el vuelo fue inolvidable
- We jumped off the plane yesterday, the flight was unforgettable
 - We jumped off the plane yesterday, the trip was unforgettable
 - We jumped off the plane yesterday, the flight was cheap
 - We jumped off the plane yesterday, the trip was cheap
 - I don't know
- 23) La modelo infantil tiene 5 años
- The weak actress is 5 years old
 - The child actress is 5 years old
 - The weak model is 5 years old
 - The child model is 5 years old
 - I don't know
- 24) La boda grave se hizo en la ciudad
- The cheap wedding took place in the city
 - The cheap funeral took place in the city
 - The serious wedding took place in the city
 - The serious funeral took place in the city
 - I don't know
- 25) Como no tenía comida la población era peligrosa
- Because they had no food the population was weak
 - Because they had no food the population was dangerous
 - Because they had no food the people were weak
 - Because they had no food the people were dangerous
 - I don't know
- 26) El tenista flojo no terminó
- The strong tennis player didn't finish
 - The strong soccer player didn't finish
 - The weak soccer player didn't finish
 - The weak tennis player didn't finish
 - I don't know

- 27) La sonrisa agradable de la joven con el ipad me gustó
- I liked the pleasant smile of the girl with the ipad
 - I liked the pleasant lips of the girl with the ipad
 - I liked the dull smile of the girl with the ipad
 - I liked the dull lips of the girl with the ipad
 - I don't know
- 28) Es imposible subir a la cumbre enorme
- It is impossible to climb the impressive peak
 - It is impossible to climb the impressive mountain
 - It is impossible to climb the enormous mountain
 - It is impossible to climb the enormous peak
 - I don't know
- 29) "El mensaje vigente" es el título del libro
- "The loving message" is the title of the book
 - "The loving letter" is the title of the book
 - "The current message" is the title of a book
 - "The current letter" is the title of a book
 - I don't know
- 30) El cura culpable fue a la cárcel
- The guilty priest went to jail
 - The guilty nun went to jail
 - The understanding priest went to jail
 - The understanding nun went to jail
 - I don't know
- 31) El fantasma insoportable está en la cocina
- The unbearable ghost is in the kitchen
 - The unbearable spirit is in the kitchen
 - The dull spirit is in the kitchen
 - The dull ghost is in the kitchen
- 32) El cuerno impermeable no se mojaba
- The waterproof horn couldn't get wet
 - The weak horn couldn't get wet
 - The waterproof bone couldn't get wet
 - The weak horn couldn't get wet
 - I don't know
- 33) El teclado envuelto es mi regalo
- The unusable keyboard is my gift
 - The wrapped keyboard is my gift
 - The unusable mouse is my gift
 - The wrapped mouse is my gift
 - I don't know
- 34) El hogar resultó indemne después de la inundación
- The home was unharmed after the flood
 - The house was unharmed after the flood
 - The house was sustainable after the flood
 - The home was sustainable after the flood
 - I don't know
- 35) La torre cara es más alta que la casa
- The impressive tower is taller than the house
 - The impressive castle is taller than the house
 - The expensive castle is taller than the house
 - The expensive tower is taller than the house
 - I don't know
- 36) La soprano amorosa canta muy bien
- The lean soprano sings really well
 - The loving contralto sings really well
 - The lean contralto sings really well
 - The loving soprano sings really well
 - I don't know

- 37) El cura flaco le perdonó
- a) The understanding nun forgave him
 - b) The understanding priest forgave him
 - c) The lean nun forgave him
 - d) The lean priest forgave him
 - e) I don't know
- 38) Subí a la cumbre afilada
- a) I climbed the impassable peak
 - b) I climbed the impassable mountain
 - c) I climbed the sharp peak
 - d) I climbed the sharp mountain
 - e) I don't know
- 39) Puedes beber del cuerno poroso
- a) You can drink from the weak horn
 - b) You can drink from the weak bone
 - c) You can drink from the porous horn
 - d) You can drink from the porous bone
 - e) I don't know
- 40) El hogar rocambolesco pertenece a la bruja
- a) The solvent home belongs to the witch
 - b) The solvent house belongs to a witch
 - c) The bizarre house belongs to a witch
 - d) The bizarre home belongs to a witch
 - e) I don't know

Multiple Choice Translation Test (ANB and AN)

Please, underline the right English translation to the given Spanish translation just like in the example. Be attentive, as a last resort, if you really don't know the words, select "I don't know", but that must be the very last resort.

Example: *El tomate está encima de la comida*

f) *The apple is on the food*

g) **The tomato is on the food**

h) *The pear is on the food*

i) *The potato is on the food*

j) *I don't know*

1. Mi madre me dio el mensaje
 - a. My mom gave me the package
 - b. My mom gave me the message
 - c. My mom gave me the letter
 - d. My mom gave me the envelope
 - e. I don't know
2. La sonrisa de la chica no me gustaba
 - a. I didn't like the lips of the girl
 - b. I didn't like the eyes of the girl
 - c. I didn't like the smile of the girl
 - d. I didn't like the smirk of the girl
 - e. I don't know
3. El delito acabó en un incendio
 - a. The felony ended up in a fire
 - b. The robbery ended up in a fire
 - c. The assault ended up in a fire
 - d. The break-in ended up in a fire
 - e. I don't know
4. La torre es de Rapunzel
 - a. The tower is Rapunzel's
5. Solo dan la beca a chicos
 - a. They only give that grant to boys
 - b. They only give that loan to boys
 - c. They only give the credit to boys
 - d. They only give the award to boys
 - e. I don't know
6. Una granja recibe la medalla
 - a. A farm receives the trophy
 - b. A farm receives the award
 - c. A farm receives the grant
 - d. A farm receives the medal
 - e. I don't know
7. El bosque será reforestado
 - a. The jungle will be reforested
 - b. The reservoir will be reforested
 - c. The garden will be reforested
 - d. The forest will be reforested
 - e. I don't know
8. Esa pared tiene un mosaico
 - b. The mansion is Rapunzel's
 - c. The castle is Rapunzel's
 - d. The fortress is Rapunzel's
 - e. I don't know

- a. That ceiling has a mosaic
 - b. That roof has a mosaic
 - c. That wall has a mosaic
 - d. That door has a mosaic
 - e. I don't know
9. La soprano no apoya a Trump
- a. The contralto doesn't support Trump
 - b. The tenor doesn't support Trump
 - c. The singer doesn't support Trump
 - d. The soprano doesn't support Trump
 - e. I don't know
10. El fantasma está en la cocina
- a. The ghost is in the kitchen
 - b. The spirit is in the kitchen
 - c. The presence is in the kitchen
 - d. The ghoul is in the kitchen
 - e. I don't know
11. El teclado se carga con energía solar
- a. The board charges with solar energy
 - b. The mouse charges with solar energy
 - c. The keyboard charges with solar energy
 - d. The screen charges with solar energy
 - e. I don't know
12. El hematoma se va con hielo
- a. The bruise disappears with ice
 - b. The wound disappears with ice
 - c. The injury disappears with ice
 - d. The pain disappears with ice
 - e. I don't know
13. Tras el accidente la moto terminó en la basura
- a. After the accident, the bicycle ended up in the trash
 - b. After the accident, the car ended up in the trash
 - c. After the accident, the motorcycle ended up in the trash
 - d. After the accident, the plane ended up in the trash
 - e. I don't know
14. La biblio tiene los mejores libros
- a. The kiosk has the best books
 - b. The stand has the best books
 - c. The bookstore has the best books
 - d. The library has the best books
 - e. I don't know
15. El anillo de la novia era de su madre
- a. The bride's ring was her mother's
 - b. The bride's bracelet was her mother's
 - c. The bride's necklace was her mother's
 - d. The bride's diamond was her mother's
 - e. I don't know
16. La cacerola tiene mucha comida
- a. The tray has lots of food
 - b. The grill has lots of food
 - c. The saucepan has lots of food
 - d. The pot has lots of food
 - e. I don't know

17. Tuvo que ir al hospital por el corte
- He had to go to the hospital because of the wound
 - He had to go to the hospital because of the injury
 - He had to go to the hospital because of the accident
 - He had to go to the hospital because of the cut
 - I don't know
18. Un bigote debe recortarse cada día
- A beard should be trimmed every day
 - An eyebrow should be trimmed every day
 - A moustache should be trimmed every day
 - A sideburn should be trimmed every day
 - I don't know
19. No hay basura en la calle
- There is no trash in the avenue
 - There is no trash in the street
 - There is no trash in the road
 - There is no trash in the alley
 - I don't know
20. La seo está en la derecha
- The old convent is on the right
 - The old church is on the right
 - The old chapel is on the right
 - The old cathedral is on the right
 - I don't know
21. Ese pesticida no se puede usar
- That herbicide can't be used
 - That pesticide can't be used
 - That insecticide can't be used
 - That fungicide can't be used
22. El vuelo de ayer me sorprendió
- Yesterday's flight surprised me
 - Yesterday's trip surprised me
 - Yesterday's ride surprised me
 - Yesterday's adventure surprised me
 - I don't know
23. La modelo tiene 5 años
- The actress is 5 years old
 - The performer is 5 years old
 - The model is 5 years old
 - The singer is 5 years old
 - I don't know
24. La boda se hizo en la ciudad
- The wedding took place in the city
 - The funeral took place in the city
 - The birth took place in the city
 - The baptism took place in the city
 - I don't know
25. Necesitamos más escuelas para la población
- We need more schools for the people
 - We need more schools for the population
 - We need more schools for the children
 - We need more schools for the town
 - I don't know

26. El tenista no terminó
- The tennis player didn't finish
 - The soccer player didn't finish
 - The basketball player didn't finish
 - The football player didn't finish
 - I don't know
27. La sonrisa de la joven con el ipad me gustó
- I liked the smile of the girl with the ipad
 - I liked the lips of the girl with the ipad
 - I liked the eyes of the girl with the ipad
 - I liked the face of the girl with the ipad
 - I don't know
28. Es imposible subir a la cumbre
- It is impossible to climb the cliff
 - It is impossible to climb the abyss
 - It is impossible to climb the mountain
 - It is impossible to climb the peak
 - I don't know
29. "El mensaje" es el título del libro
- "The message" is the title of the book
 - "The letter" is the title of the book
 - "The email" is the title of a book
 - "The package" is the title of a book
 - I don't know
30. El cura fue a la cárcel
- The priest went to jail
 - The nun went to jail
 - The canon went to jail
 - The bishop went to jail
 - I don't know
31. El fantasma está en la cocina
- The ghost is in the kitchen
 - The spirit is in the kitchen
 - The presence is in the kitchen
 - The ghoul is in the kitchen
 - I don't know
32. El cuerno no se mojaba
- The bone couldn't get wet
 - The horn couldn't get wet
 - The antler couldn't get wet
 - The paw couldn't get wet
 - I don't know
33. El teclado es mi regalo
- The keyboard is my gift
 - The board is my gift
 - The mouse is my gift
 - The screen is my gift
 - I don't know
34. El hogar protege el medioambiente
- The condo protects the environment
 - The flat protects the environment
 - The house protects the environment
 - The home protects the environment
 - I don't know

35. La torre es más alta que la casa
- The palace is taller than the house
 - The castle is taller than the house
 - The building is taller than the house
 - The tower is taller than the house
 - I don't know
36. La soprano canta muy bien
- The soprano sings really well
 - The contralto sings really well
 - The tenor sings really well
 - The mezzosoprano sings really well
 - I don't know
37. El cura le perdonó
- The nun forgave him
 - The canon forgave him
 - The bishop forgave him
 - The priest forgave him
 - I don't know
38. Subí a la cumbre
- I climbed the peak
 - I climbed the mountain
 - I climbed the cliff
 - I climbed the abyss
 - I don't know
39. Puedes beber del cuerno
- You can drink from the wing
 - You can drink from the antler
 - You can drink from the horn
 - You can drink from the bone
 - I don't know
40. El hogar pertenece a la bruja
- The home belongs to the witch
 - The condo belongs to a witch
 - The house belongs to a witch
 - The flat belongs to a witch
 - I don't know

Multiple Choice Concordance Test

Name: _____

Choose the appropriate article and adjective for the noun (in bold font). You have to choose 1 article out of the 2 options you are given and 1 adjective out of the 3 you are given, just like in the example. You can circle or underline your choice, just make it clear:

Example: El / la **tomate** verdo / verda / verde se come con la ensalada →

El / la **tomate** verdo / verda / verde se come con la ensalada

El / la **comida** está muy bueno / buena / buen →

El / la **comida** está muy bueno / buena / buen

1. El / la **sonrisa** afectuoso / afectuosa / afectuose de la chica no me gustaba
2. El / la **tenista** flojo / floja / floje no terminó
3. Necesitamos más escuelas para el / la **población** infantil / infantil / infantil
4. Es imposible subir a el / la **cumbre** enorme / enorma / enorme
5. Mi madre me dio el / la **mensaje** soso / sosa / sose
6. El / la **boda** barato / barata / barat se hizo en la ciudad
7. El / la **cura** flaco / flaca / flac le perdonó
8. El / la **modelo** pobro / pobra / pobre tiene 5 años
9. El / la **fantasma** comprensivo / comprensiva / comprensive está en la cocina
10. El / la **cuerno** impermeablo / impermeabla / impermeable no se mojaba
11. El / la **teclado** envuelto / envuelta / envuelto es mi regalo
12. Ayer saltamos del avión, el / la **vuelo** fue inolvidable / inolvidable / inolvidable
13. El / la **pesticida** formidablo / formidabla / formidable funciona muy bien

14. El / la **hogar** resultó indemno / indemna / indemne después de la inundación
15. El / la **torre** caro / cara / care es más alta que la casa
16. El / la **seo** formidable / formidabla / formidable está en la derecha
17. El / la **soprano** amoroso / amorosa / amorose canta muy bien
18. El / la **delito** grave / grava / grave que cometió era un robo
19. Le dieron un / una **beca** flojo / floja / floje para estudiar arte
20. Una granja recibe el / la **medalla** agrícola / agrícولا / agrícولة
21. El / la **bosque** inolvidable / inolvidabla / inolvidable está cerca de la ciudad
22. Ese / esa **pared** barato / barata / barate tiene un mosaico
23. El / la **hematoma** está inflado / inflada / inflade por el golpe
24. El / la **moto** no es impresionante / impresionanta / impresionante en comparación con el avión para viajes largos
25. El / la **biblio** infantil / infantil / infantil tiene pocos libros
26. El / la **anillo** sorprendente / sorprendenta / sorprendente de la novia era de su madre
27. El / la **cacerola** ligero / ligera / liger no tenía comida
28. El / la **corte** inflado / inflada / inflade se infectó
29. Todos han visto el / la **bigote** decepcionante / decepcionanta / decepcionante de Dalí
30. No hay basura en el / la **calle** peligroso / peligrosa / peligros
31. En el / la **calle** pobre / pobra / pobre la gente no tiene casas
32. El / la **seo** pulcro / pulcra / pulcre está en la izquierda
33. Ese / esa **pesticida** específico / específica / específico no se puede usar

34. Había muchas opciones, pero cogí el / la **vuelo** lujoso / lujosa / lujose
35. El / la **modelo** ligero / ligera / ligere pesa 90 libras
36. El / la **hematoma** grave / grava / grave se va con hielo
37. Tuvo que ir al hospital por el / la **corte** desagradable / desagradable / desagradable
38. El / la **boda** fue decepcionante / decepcionante / decepcionante porque los novios no se querían
39. Como no tenía comida el / la **población** era peligroso / peligrosa / peligrosa
40. Tras el accidente el / la **moto** abollado / abollada / abollada terminó en la basura
41. El / la **biblio** lujoso / lujosa / lujosa tiene los mejores libros
42. El / la **tenista** desagradable / desagradable / desagradable ganó
43. El / la **sonrisa** agradable / agradable / agradable de la joven con el ipad me gustó
44. Subí a el / la **cumbre** afilado / afilada / afilada
45. El / la **cura** culpable / culpable / culpable fue a la cárcel
46. El / la **hogar** rocambolesco / rocambolesca / rocambolesco pertenece a la bruja
47. El / la **delito** específico / específica / específico acabó en un incendio
48. El / la **bosque** vacío / vacía / vacía será reforestado
49. "El / la **mensaje** vigenta / vigenta / vigenta" es el título del libro
50. No se podía poner el / la **anillo** abollado / abollada / abollada
51. El / la **teclado** sostenible / sostenible / sostenible se carga con energía solar
52. El / la **torre** impracticable / impracticable / impracticable es de Rapunzel

53. Un / una **bigote** pulcro / pulcra / pulcre debe recortarse cada día
54. El / la **cacerola** impresionante / impresionanta / impresionante
tiene mucha comida
55. Había muchos dibujos en el / la **pared** sorprendente /
sorprendenta / sorprendente
56. Solo dan el / la **beca** agrícola / agrícola/ agrícola a chicos
57. El / la **medalla** estaba vacío / vacía / vacíe de significado
58. El / la **fantasma** insoportable / insoportable / insoportable está en la
habitación
59. Puedes beber con el / la **cuerno** poroso / porosa / porose
60. El / la **soprano** progresista / progresista / progresista no apoya a
Trump

Debriefing Questionnaire

Name: _____

1. When answering on your answer sheet, did you read the sentences on your answer sheet or only on the screen?

2. Did you notice any patterns during the PowerPoint or the Testing? *Yes / No*
If yes, which ones? _____

3. Did you know the words the PowerPoint taught you beforehand?

4. Did you feel like you knew the words in the translation activity and the last activity? *Yes / No*

5. What do you think was the goal of the experiment?

Delayed Post-Test for Known Nouns

Name: _____

Choose the appropriate article and adjective for the noun (in bold font). You have to choose 1 article out of the 2 options you are given and 1 adjective out of the 3 you are given, just like in the example. You can circle or underline your choice, just make it clear:

Example: El / la **tomate** verdo / verda / verde se come con la ensalada →

El / la **tomate** verdo / verda / verde se come con la ensalada

El / la **comida** está muy bueno / buena / buen →

El / la **comida** está muy bueno / buena / buen

1. El / la **perro** negro / negra / negre es de Miguel y Beatriz
2. El / la **radio** viejo / vieja / vieje no funciona
3. El / la **discusión** es fácil / fácil / fácil de resolver
4. El / la **hermano** mayoro / mayora / mayor cuida al resto
5. El / la **foto** interesanto / interesanta / interesante es del polo norte
6. El / la **camiseta** verdo / verda / verde se hizo en Estados Unidos
7. El / la **satélite** inteligente / inteligente / inteligente controla todo el internet
8. El / la **carnaval** tranquilo / tranquila / tranquil nos aburrió mucho
9. El / la **hermana** activo / activa / active hace muchísimo deporte
10. El / la **casa** es normalo / normala / normal, no tiene nada especial
11. Todo el mundo mira el / la **programa** popularo / populara / popular
12. El / la **bar** al que fuimos ayer era grando / granda / grande
13. El / la **planeta** rápido / rápida / rápid da la vuelta al sol a mucha velocidad
14. El / la **tenedor** roso / rosa / rose era de juguete

15. No se puede hablar con el / la **dentista** difícil / difícil / difícil,
nunca escucha
16. Resolvió el / la **actividad** fácil / fácil / fácil sin problemas
17. El / la **madre** simpático / simpática / simpático invita a todos los niños
a su casa
18. El / la **carne** gris / gris / gris no se puede comer
19. El / la **televisión** está rojo / roja / rojo porque le cayó pintura encima
20. El / la **madre** paciente / paciente / paciente perdona a todos sus hijos
21. El / la **día** pasó muy rápido / rápida / rápido, ni me di cuenta
22. El / la **música** que escucha Lina es interesante / interesante /
interesante
23. El / la **bebida** es claro / clara / clara, puedes ver a través
24. El / la **artista** era difícil / difícil / difícil, jamás le gustaba nada
25. El / la **mano** marrón / marrón / marrón porque tiene barro encima
26. El / la **discusión** complicado / complicada / complicada no tenía
solución
27. Mi madre me dio el / la **zapato** bonito / bonita / bonito y me gusta mucho
28. El / la **mapa** interesante / interesante / interesante te lleva a un tesoro
29. El / la **atleta** alto / alta / alto ganó la competición
30. El / la **camiseta** bonito / bonita / bonito tiene muchas flores
estampadas
31. El / la **libro** naranja / naranja / naranja habla sobre política
32. El / la **gato** blanco / blanca / blanco tiene muchos años
33. El / la **foto** viejo / vieja / viejo es de mis abuelos
34. El / la **radio** azul / azul / azul es de mi abuela
35. El / la **hermano** pequeño / pequeña / pequeño se porta muy bien
36. El / la **zapato** verde / verde / verde no gusta a nadie

37. Un / una **tomate** normalo / normala / normal es de color rojo
38. El / la **dentista** está enfermo / enferma / enferm, tiene un resfriado
39. El / la **casa** es bajo / baja /baje, solo tiene un piso
40. El / la **hermana** de Ana es pobro / pobra / pobre
41. El / la **bar** tranquilo / tranquila / tranquil tiene pocos clientes
42. Nadie puede resolver el / la **actividad** complicado / complicada / camplicad
43. El / la **artista** está enfermo / enferma / enferme con cáncer
44. El / la **carnaval** en Brasil es muy grando / granda / grande
45. El / la **tenedor** está frío / fría / fríe porque es de metal
46. El / la **carne** rojo / roja / roje es buena para la salud
47. El / la **televisión** gris / grisa / gris no tenía color, solo blanco y negro
48. El / la **programa** era largo / larga / largue porque tenía mucha información
49. Cuidan muchísimo a el / la **perro** por eso está felizo / feliza / feliz
50. El / la **gato** naranjo / naranja / naranje vive en la calle
51. El / la **mano** feo / fea / fe es de un gorila
52. Pasamos el / la **día** alegre / alegre / alegre en el parque
53. El / la **atleta** joven / jovena / joven no puede participar en la competición
54. El / la **satélite** malo / mala / mal espiaba a los ciudadanos
55. El / la **música** claro / clara / clare se escucha muy bien
56. El / la **mapa** nuevo / nueva / nueve incluye muchos países
57. El / la **planeta** caliente / caliente / caliente está en llamas
58. El / la **tomate** bajo / baja / baje sabe mejor que el de arriba
59. El / la **libro** blanco / blanca / blanc fue un regalo de María

60. El / la **bebida** interesante / interesanta / interesante tenía frutas
tropicales

Delayed Post-Test for Treatment Nouns

Name: _____

Choose the appropriate article and adjective for the noun (in bold font). You have to choose 1 article out of the 2 options you are given and 1 adjective out of the 3 you are given, just like in the example. You can circle or underline your choice, just make it clear:

Example: El / la **tomate** verdo / verda / verde se come con la ensalada →

El / la **tomate** verdo / verda / verde se come con la ensalada

El / la **comida** está muy bueno / buena / buen →

El / la **comida** está muy bueno / buena / buen

1. El / la **sonrisa** afectuoso / afectuosa / afectuose de la chica no me gustaba
2. El / la **tenista** flojo / floja / floje no terminó
3. Necesitamos más escuelas para el / la **población** infantil / infantil / infantil
4. Es imposible subir a el / la **cumbre** enorme / enorme / enorme
5. Mi madre me dio el / la **mensaje** soso / sosa / soso
6. El / la **boda** barato / barata / barat se hizo en la ciudad
7. El / la **cura** flaco / flaca / flac le perdonó
8. El / la **modelo** pobro / pobra / pobre tiene 5 años
9. El / la **fantasma** comprensivo / comprensiva / comprensive está en la cocina
10. El / la **cuerno** impermeablo / impermeabla / impermeable no se mojaba
11. El / la **teclado** envuelto / envuelta / envuelto es mi regalo
12. Ayer saltamos del avión, el / la **vuelo** fue inolvidable / inolvidable / inolvidable

13. El / la **pesticida** formidable / formidabla / formidable funciona muy bien
14. El / la **hogar** resultó indemno / indemnada / indemne después de la inundación
15. El / la **torre** caro / cara / care es más alta que la casa
16. El / la **seo** formidable / formidabla / formidable está en la derecha
17. El / la **soprano** amoroso / amorosa / amorosa canta muy bien
18. El / la **delito** grave / grava / grave que cometió era un robo
19. Le dieron un / una **beca** flojo / floja / floje para estudiar arte
20. Una granja recibe el / la **medalla** agrícola / agrícolica / agrícolica
21. El / la **bosque** inolvidable / inolvidable / inolvidable está cerca de la ciudad
22. Ese / esa **pared** barato / barata / barata tiene un mosaico
23. El / la **hematoma** está inflado / inflada / inflado por el golpe
24. El / la **moto** no es impresionante / impresionante / impresionante en comparación con el avión para viajes largos
25. El / la **biblio** infantil / infantil / infantil tiene pocos libros
26. El / la **anillo** sorprendente / sorprendente / sorprendente de la novia era de su madre
27. El / la **cacerola** ligero / ligera / ligero no tenía comida
28. El / la **corte** inflado / inflada / inflado se infectó
29. Todos han visto el / la **bigote** decepcionante / decepcionante / decepcionante de Dalí
30. No hay basura en el / la **calle** peligroso / peligrosa / peligrosos
31. En el / la **calle** pobre / pobre / pobre la gente no tiene casas
32. El / la **seo** pulcro / pulcra / pulcra está en la izquierda

33. Ese / esa **pesticida** específico / específica / específic no se puede usar
34. Había muchas opciones, pero cogí el / la **vuelo** lujoso / lujosa / lujose
35. El / la **modelo** ligero / ligera / ligere pesa 90 libras
36. El / la **hematoma** grave / grava / grave se va con hielo
37. Tuvo que ir al hospital por el / la **corte** desagradable / desagradable
38. El / la **boda** fue decepcionante / decepcionante / decepcionante porque los novios no se querían
39. Como no tenía comida el / la **población** era peligroso / peligrosa / peligrosa
40. Tras el accidente el / la **moto** abollado / abollada / abollade terminó en la basura
41. El / la **biblio** lujoso / lujosa / lujose tiene los mejores libros
42. El / la **tenista** desagradable / desagradable / desagradable ganó
43. El / la **sonrisa** agradable / agradable / agradable de la joven con el ipad me gustó
44. Subí a el / la **cumbre** afilado / afilada / afilade
45. El / la **cura** culpable / culpable / culpable fue a la cárcel
46. El / la **hogar** rocambolesco / rocambolesca / rocambolesque pertenece a la bruja
47. El / la **delito** específico / específica / específic acabó en un incendio
48. El / la **bosque** vacío / vacía / vacíe será reforestado
49. "El / la **mensaje** vigente / vigente / vigente" es el título del libro
50. No se podía poner el / la **anillo** abollado / abollada / abollade
51. El / la **teclado** sostenible / sostenible / sostenible se carga con energía solar

52. El / la **torre** impracticable / impracticable / impracticable es de Rapunzel
53. Un / una **bigote** pulcro / pulcra / pulcre debe recortarse cada día
54. El / la **cacerola** impresionante / impresionante / impresionante tiene mucha comida
55. Había muchos dibujos en el / la **pared** sorprendente / sorprendente
56. Solo dan el / la **beca** agrícola / agrícola/ agrícola a chicos
57. El / la **medalla** estaba vacío / vacía / vacía de significado
58. El / la **fantasma** insostenible / insostenible / insostenible está en la habitación
59. Puedes beber con el / la **cuerno** poroso / porosa / porosa
60. El / la **soprano** progresista / progresista / progresista no apoya a Trump

Delayed Post-Test of Unknown Nouns

Name: _____

Choose the appropriate article and adjective for the noun (in bold font). You have to choose 1 article out of the 2 options you are given and 1 adjective out of the 3 you are given, just like in the example. You can circle or underline your choice, just make it clear:

Example: El / la **tomate** verdo / verda / verde se come con la ensalada →

El / la **tomate** verdo / verda / verde se come con la ensalada

El / la **comida** está muy bueno / buena / buen →

El / la **comida** está muy bueno / buena / buen

1. El / la **yugo** ligero / ligera / ligere era fácil de usar
2. El / la **casco** chocanto / chocanta / chocante sorprendió a todo el mundo
3. La camisa tenía un / una **mancha** inquietanto / inquietanta / inquietante
4. El / la **borrego** árabo / áraba / árabe es de Arabia
5. El / la **hacha** del asesinato es muy inquietanto / inquietanta / inquietante
6. El / la **crueldad** contundento / contundenta / contundente del tirano asustaba a todo el mundo
7. El / la **huelga** repentino / repentina / repentin tomó a los turistas por sorpresa
8. El / la **obispo** tenazo / tenaza / tenaz no se rinde nunca
9. El / la **hazaña** bélico / bélica / belic acabó con muchos muertos
10. El / la **aurora** en Alaska es inolvidablo / inolvidabla / inolvidable
11. Las abejas de el / la **enjambre** sanguinario / sanguinaria / sanguinare picaron a los niños
12. El / la **perdiz** cuentisto / cuentista / cuentiste siempre cuenta historias
13. El / la **servidumbre** halagüeño / halagüeña / halagüeñe dice cosas buenas al jefe

14. El / la **comerciante** piadoso / piadosa / piados perdonó el dinero que se le debía
15. El / la **sastre** mordazo / mordaza / mordaz hablaba mal de sus clientes
16. El / la **peste** es infeccioso / infecciosa / infecciose y mata gente
17. El / la **pulpo** árabo / áraba / árabe es un plato riquísimo
18. El / la **eyección** aleatorio / aleatoria / aleatoricid del astronauta hizo que muriera
19. El / la **terrateniente** fielo / fiela / fiel era muy buena persona
20. Los soldados preferían el / la **casco** ligero / ligera / ligere
21. El tiburón se comió el / la **calamar** fielo / fiela / fiel
22. Nadie quiere sufrir el / la **anatema** costoso / costosa / costose
23. Ese / esa **hazaña** bélico / bélica / bélic ganó la guerra
24. Aquella es el / la **aula** susodicho / susodicha / susodiche.
25. El / la **yugo** chocanto / chocanta / chocante no era de madera
26. El / la **anatema** exigento / exigenta / exigente es un castigo muy duro
27. El / la **hacha** susodicho / susodicha / susodiche se usó para cortar el árbol
28. La ley protege el / la **borrego** autóctono / autóctona / autócton
29. El / la **calamar** desamparado / desamparada / desamparad no tiene familia
30. Es normal, cada día se hace el / la **mancha** cotidiano / cotidiana / cotidian
31. El / la **huelga** era comprensible / comprensible / comprensible, querían seguro médico
32. El / la **marea** repentino / repentina / repentin inundó la ciudad
33. El / la **sastre** piadoso / piadosa / piados ayudaba a todos
34. La gente se aburrió con el / la **aurora** aguado / aguada / aguade

35. El / la **enjambre** inverosímilo / inverosímila / inverosímil ocupa toda la casa
36. El / la **crueidad** del rey era infeccioso / infecciosa / infecciose
37. La gente prefiere comer el / la **pulpo** autóctono / autóctona / autóctone
38. El / la **comerciante** mordazo / mordaza / mordaz mentía a sus clientes
39. Se trata de el / la **perdiz** idóneo / idónea / idóne para la comida
40. El / la **peste** contundento / contundenta / contundente mató mucha gente en Europa
41. El / la **servidumbre** humildo / humilda / humilde nunca pide más dinero
42. El / la **eyección** unánimo / unánima / unánime mandó el astronauta lejos
43. El / la **obispo** sofocado / sofocada / sofocade corría para llegar a la reunión
44. El / la **terratiente** desamparado / desamparada / desamparade había perdido todas sus cosas
45. El / la **aula** fugazo / fugaza / fugaz desapareció
46. El / la **marea** estivalo / estivala / estival pasa cada verano

APPENDIX D

Appendix D presents all statistical results for the study in extension, including non-significant interactions for consultation. After the results there is a guide to how the data was coded.

RQ1: What differences are there between groups regarding the gender of nouns combined with concordance?

Table C1

Differences between treatment groups

Treatment		Mean difference	<i>p</i> value	Direction of effect
C	ANAB	6.33	.216	-
	ANA	8.14	.031*	C>ANA
	ANB	6.72	.170	-
	AN	10.20	.007**	C>AN
ANAB	ANA	1.81	1.000	-
	ANB	.39	1.000	-
	AN	3.87	1.000	-
ANA	ANB	-1.42	1.000	-
	AN	2.06	1.000	-
ANB	AN	3.48	1.000	-

Note. C=Control; ANAB=Article, Noun, and Adjective in bold font; ANA=Article, Noun, and Adjective; ANB=Article and Noun in bold font; AN=Article and Noun. **p* < .05, ***p* < .01, ****p* < .001

Table C2

Direction of effects for Concordance and Gender separately

Variables		Mean difference	<i>p</i> value	Direction of effect
Assignment	Agreement	4.53	.000***	Assignment>Agreement
Masculine	Feminine	15.15	.000***	Masculine>Feminine

Note. **p* < .05, ***p* < .01, ****p* < .001

Table C3

Interaction between Concordance Gender*

Gender	Concordance		Difference	<i>p</i> value	Direction of effect
Masculine	Assignment	Agreement	3.6	.000***	Assignment>Agreement
Feminine	Assignment	Agreement	5.45	.000***	Assignment>Agreement

Note. **p* < .05, ***p* < .01, ****p* < .001

When looking into group interactions controlling by Concordance we found that the only significant interactions in both agreement and assignment were found between the control group and the rest of groups (Table C4), which outperformed the rest, suggesting that differences among treatment groups were due to chance or other factors.

Table C4
Interaction between Concordance Treatment*

Concordance	Treatment		Mean difference	<i>p</i> value	Direction of effect	
Assignment	C	ANAB	5.25	.041*	C>ANAB	
		ANA	7.37	.004**	C>ANA	
		ANB	6.53	.013*	C>ANB	
		AN	7.79	.005**	C>AN	
	ANAB	ANA	2.11	.380	-	
		ANB	1.27	.608	-	
		AN	2.54	.333	-	
	ANA	ANB	-.84	.732	-	
		AN	.43	.869	-	
	ANB	AN	1.27	.635	-	
	Agreement	C	ANAB	7.41	.021*	C>ANAB
			ANA	8.92	.005**	C>ANA
ANB			6.91	.035*	C>ANB	
AN			12.61	.000***	C>AN	
ANAB		ANA	1.51	.616	-	
		ANB	-.50	.871	-	
		AN	5.2	.114	-	
ANA		ANB	-2.01	.513	-	
		AN	3.69	.256	-	
ANB		AN	5.70	.089	-	

Note. C=Control; ANAB=Article, Noun, and Adjective in bold font; ANA=Article, Noun, and Adjective; ANB=Article and Noun in bold font; AN=Article and Noun. **p* < .05, ***p* < .01, ****p* < .001

When controlling by noun gender we find the same effects there were on the treatment analysis above, the control group only significantly outperforms the groups without enhancement. In addition, the ANAB group also outperforms the AN group, suggesting that only the combination of enhancement and presence of adjectives make cause a noticeable difference. Regarding feminine gender we find that the control group only

significantly outperforms the groups with an adjective (ANAB and ANA), suggesting that the groups that were exposed to agreement with the adjective as well significantly underperformed in comparison to the control group while the groups without adjective (ANB and AN) did not. Full report is on Table C5 below.

Table C5
*Interaction between Gender*Treatment*

Gender	Treatment	Mean difference	<i>p</i> value	Direction of effect	
Masculine	C	ANAB	4.17	.247	-
		ANA	7.59	.035*	C>ANA
		ANB	6.31	.087	-
		AN	13.56	.001**	C>AN
	ANAB	ANA	3.42	.314	-
		ANB	2.15	.540	-
		AN	9.39	.012*	ANAB>AN
	ANA	ANB	-1.27	.713	-
		AN	5.98	.105	-
	ANB	AN	7.25	.056	-
		AN	7.25	.056	-
	Feminine	C	ANAB	8.5	.028*
ANA			8.7	.023*	C>ANA
ANB			7.12	.070	-
AN			6.84	.097	-
ANAB		ANA	.199	.956	-
		ANB	-1.37	.712	-
		AN	-1.65	.674	-
ANA		ANB	-1.57	.670	-
		AN	-1.85	.635	-
ANB		AN	-2.80	.944	-
		AN	-2.80	.944	-

Note. C=Control; ANAB=Article, Noun, and Adjective in bold font; ANA=Article, Noun, and Adjective; ANB=Article and Noun in bold font; AN=Article and Noun. **p* < .05, ***p* < .01, ****p* < .001

On Table C6 the interactions between groups controlled by concordance and gender are reported. In assignment controlled by masculine nouns we find that the results found in the masculine gender nouns alone hold, but any significant interactions in the feminine nouns disappear. Controlling by agreement the effects in the masculine nouns still hold and add a new significant interaction that suggests the ANB group outperforms the AN. Thus,

the AN group is systematically outperformed by those with enhancement, while the presence of adjectives does not seem to suffice to result in significant differences. In the feminine group we find more significant interactions, with the control group outperforming every group (ANAB, ANA, and AN) except for the ANB group, suggesting that those groups significantly underperformed in comparison to the control group, but again not finding significant differences between treatment groups.

Table C6
Interaction between Concordance Gender*Treatment*

Concordance	Gender	Treatment	Mean difference	p value	Direction of effect	
Assignment	Masculine	C	ANAB	3.23	.338	-
			ANA	6.91	.040*	C>ANA
			ANB	6.28	.070	-
			AN	11.47	.002**	C>AN
		ANAB	ANA	3.67	.249	-
			ANB	3.04	.354	-
			AN	8.233	.019*	ANAB>AN
		ANA	ANB	-.63	.846	-
			AN	4.56	.186	-
		Feminine	C	ANAB	7.27	.075
	ANA			7.82	.053	-
	ANB			6.77	.103	-
	AN			4.12	.344	-
	ANAB		ANA	.55	.885	-
			ANB	-.50	.900	-
			AN	-3.15	.451	-
	ANA		ANB	-1.05	.788	-
			AN	-3.70	.371	-
	ANB		AN	-2.65	.533	-
		AN	-2.65	.533	-	
Agreement	Masculine	C	ANAB	5.11	.218	-
			ANA	8.27	.045*	C>ANA
			ANB	6.35	.134	-
			AN	15.66	.001**	C>AN
		ANAB	ANA	3.17	.417	-
			ANB	1.25	.757	-
			AN	10.56	.014*	ANAB>AN
		ANA	ANB	-1.92	.630	-
			AN	7.39	.082	-
		ANB	AN	9.31	.034*	ANB>AN
			AN	9.31	.034*	ANB>AN

Feminine	C	ANAB	9.72	.013*	C>ANAB
		ANA	9.57	.014*	C>ANA
		ANB	7.47	.060	-
		AN	9.56	.023*	C>AN
	ANAB	ANA	-.16	.966	-
		ANB	-2.25	.550	-
		AN	-.16	.968	-
	ANA	ANB	-2.09	.574	-
		AN	-.00	.999	-
	ANB	AN	2.09	.606	-

Note. C=Control; ANAB=Article, Noun, and Adjective in bold font; ANA=Article, Noun, and Adjective; ANB=Article and Noun in bold font; AN=Article and Noun. * $p < .05$, ** $p < .01$, *** $p < .001$

RQ2: What differences are there between groups regarding noun ending combined with concordance?

Table C7
Differences between treatment groups

Treatment		Mean difference	p value	Direction of effect
C	ANAB	6.52	.191	-
	ANA	8.50	.023*	C>ANA
	ANB	6.51	.208	-
	AN	9.65	.014*	C>AN
ANAB	ANA	1.97	1.000	-
	ANB	.04	1.000	-
	AN	3.13	1.000	-
ANA	ANB	-1.93	1.000	-
	AN	1.16	1.000	-
ANB	AN	3.09	1.000	-

Note. C=Control; ANAB=Article, Noun, and Adjective in bold font; ANA=Article, Noun, and Adjective; ANB=Article and Noun in bold font; AN=Article and Noun. * $p < .05$, ** $p < .01$, *** $p < .001$

Table C8
Direction of effects among Noun Ending

Ending		Mean difference	p value	Direction of effect
C	NC	39.13	.000***	C>NC
	E	80.57	.000***	C>E
NC	E	41.43	.000***	NC>E

Note. Noun Ending are coded as C=Canonical, NC=Non-Canonical, E=Exception. * $p < .05$, ** $p < .01$, *** $p < .001$

When looking into group interactions controlling by Concordance I found that the only significant interactions in both agreement and assignment were between the control group and the rest of groups (Table C9), which outperformed the rest, suggesting that differences among treatment groups were due to chance or other factors.

Table C9
Interaction between Concordance Treatment*

Concordance	Treatment		Mean difference	<i>p</i> value	Direction of effect
Assignment	C	ANAB	5.42	.036*	C>ANAB
		ANA	7.61	.003**	C>ANA
		ANB	6.50	.014*	C>ANB
		AN	7.51	.007**	C>AN
	ANAB	ANA	2.19	.364	-
		ANB	1.08	.664	-
		AN	2.09	.429	-
	ANA	ANB	-1.11	.653	-
		AN	-.108	.967	-
	ANB	AN	1.00	.709	-
		AN	1.00	.709	-
	Agreement	C	ANAB	7.62	.020*
ANA			9.37	.004**	C>ANA
ANB			6.62	.046*	C>ANB
AN			11.80	.001**	C>AN
ANAB		ANA	1.75	.564	-
		ANB	-.997	.751	-
		AN	4.18	.209	-
ANA		ANB	-2.75	.377	-
		AN	2.43	.460	-
ANB		AN	5.18	.128	-
		AN	5.18	.128	-

Note. C=Control; ANAB=Article, Noun, and Adjective in bold font; ANA=Article, Noun, and Adjective; ANB=Article and Noun in bold font; AN=Article and Noun. **p* < .05, ***p* < .01, ****p* < .001

Table C10
*Interaction between Ending*Treatment*

Ending	Treatment		Mean difference	<i>p</i> value	Direction of effect
Canonical	C	ANAB	-1.07	.815	-
		ANA	4.39	.333	-
		ANB	2.36	.613	-
		AN	9.60	.052	-
	ANAB	ANA	5.46	.207	-
		ANA	5.46	.207	-

Non-Canonical	ANA	ANB	3.43	.442	-	
		AN	10.67	.025*	ANAB>AN	
	ANB	ANB	-2.04	.644	-	
		AN	5.21	.266	-	
	C	AN	7.25	.133	-	
		ANAB	18.26	.000***	C>ANAB	
	Exception	ANAB	ANA	20.07	.000***	C>ANA
			ANB	16.15	.002**	C>ANB
		ANA	AN	23.70	.000***	C>AN
			ANAB	1.81	.706	-
ANB		ANB	-2.11	.670	-	
		AN	5.44	.299	-	
C		ANAB	-3.92	.425	-	
		AN	3.63	.484	-	
ANAB		ANB	7.55	.158	-	
		ANAB	2.36	.524	-	
	ANA	1.01	.783	-		
	ANB	1.18	.755	-		
	AN	-4.35	.275	-		
	ANAB	-1.35	.699	-		
	ANB	-1.19	.742	-		
	AN	-6.71	.080	-		
ANA	ANB	.165	.963	-		
ANB	AN	-5.36	.158	-		
	AN	-5.52	.157	-		

Note. C=Control; ANAB=Article, Noun, and Adjective in bold font; ANA=Article, Noun, and Adjective; ANB=Article and Noun in bold font; AN=Article and Noun. * $p < .05$, ** $p < .01$, *** $p < .001$

Table C11
*Interaction between Ending*Concordance*

Ending	Treatment	Mean difference	p value	Direction of effect
Canonical	Asgn Agr	5.36	.000***	Asgn>Agr
Non-Canonical	Asgn Agr	5.12	.000***	Asgn>Agr
Exception	Asgn Agr	.92	.008**	Asgn>Agr

Note. Asgn=Assignment, Agr=Agreement. * $p < .05$, ** $p < .01$, *** $p < .001$

Table C12
Interaction between Concordance Ending*Treatment*

Concordance	Ending	Treatment	Mean difference	p value	Direction of effect	
Assignment	Canonical	C	ANAB	-2.84	.469	-
			ANA	2.14	.581	-
			ANB	1.91	.634	-

			AN	4.43	.294	-
		ANAB	ANA	4.99	.179	-
			ANB	4.75	.215	-
			AN	7.27	.074	-
		ANA	ANB	-.239	.950	-
			AN	2.28	.569	-
		ANB	AN	2.52	.540	-
	Non-Canonical	C	ANAB	17.07	.001**	C>ANAB
			ANA	19.18	.000***	C>ANA
			ANB	16.61	.002**	C>ANB
			AN	22.21	.000***	C>AN
		ANAB	ANA	2.11	.658	-
			ANB	-.46	.925	-
			AN	5.14	.324	-
		ANA	ANB	-2.57	.598	-
			AN	3.03	.557	-
	Exception	ANB	AN	5.60	.291	-
		C	ANAB	2.03	.599	-
			ANA	1.51	.693	-
			ANB	.994	.801	-
			AN	-4.13	.320	-
		ANAB	ANA	-.52	.887	-
			ANB	-1.03	.783	-
			AN	-6.15	.123	-
		ANA	ANB	-.52	.890	-
			AN	-5.64	.154	-
		ANB	AN	-5.12	.208	-
Agreement	Canonical	C	ANAB	.70	.900	-
			ANA	6.64	.233	-
			ANB	2.81	.624	-
			AN	14.78	.015*	C>AN
		ANAB	ANA	5.94	.263	-
			ANB	2.10	.700	-
			AN	14.07	.016*	ANAB>AN
		ANA	ANB	-3.84	.479	-
			AN	8.13	.157	-
		ANB	AN	11.97	.044*	ANB>AN
	Non-Canonical	C	ANAB	15.45	.001**	C>ANAB
			ANA	20.95	.000***	C>ANA
			ANB	15.61	.006**	C>ANB
			AN	25.19	.000***	C>AN
		ANAB	ANA	1.50	.771	-
			ANB	-3.76	.483	-
			AN	5.74	.311	-
		ANA	ANB	-5.26	.321	-
			AN	4.24	.450	-

Exception	ANB	AN	9.50	.101	-	
	C	ANAB	2.69	.456	-	
		ANA	.51	.886	-	
		ANB	1.36	.713	-	
		AN	-4.57	.240	-	
		ANAB	ANA	-2.18	.522	-
			ANB	-1.34	.704	-
			AN	-7.26	.053	-
		ANA	ANB	.85	.808	-
			AN	-5.08	.170	-
		ANB	AN	-5.93	.120	-

Note. C=Control; ANAB=Article, Noun, and Adjective in bold font; ANA=Article, Noun, and Adjective; ANB=Article and Noun in bold font; AN=Article and Noun. * $p < .05$, ** $p < .01$, *** $p < .001$

RQ3: What differences are there between accuracy of concordance in new versus old pairings of nouns and adjectives?

Table C13
Differences between treatment groups

Treatment		Mean difference	p value	Direction of effect
C	ANAB	3.27	.597	-
	ANA	4.58	.083	-
	ANB	3.55	.457	-
	AN	6.50	.006**	C>AN
ANAB	ANA	1.13	1.000	-
	ANB	.28	1.000	-
	AN	3.23	.709	-
ANA	ANB	-1.04	1.000	-
	AN	1.92	1.000	-
ANB	AN	2.96	1.000	-

Note. C=Control; ANAB=Article, Noun, and Adjective in bold font; ANA=Article, Noun, and Adjective; ANB=Article and Noun in bold font; AN=Article and Noun. * $p < .05$, ** $p < .01$, *** $p < .001$

When looking into group interactions controlling by Concordance I found that the only significant interactions in assignment were between the control group and the rest of groups (Table C14) and between the AN group and the control and ANAB groups in agreement, indicating that AN underperformed significantly in comparison to the control and ANAB

groups, suggesting again that only enhanced input and the exposure to the combination of a noun and adjectives cause a significantly different performance.

Table C14
Interaction between Concordance Treatment*

Concordance	Treatment		Mean difference	<i>p</i> value	Direction of effect
Assignment	C	ANAB	4.25	.041*	C>ANAB
		ANA	6.13	.003**	C>ANA
		ANB	4.29	.044*	C>ANB
		AN	5.32	.018*	C>AN
	ANAB	ANA	1.88	.334	-
		ANB	.04	.985	-
		AN	1.07	.616	-
	ANA	ANB	-1.85	.355	-
		AN	-.82	.698	-
	ANB	AN	1.03	.635	-
		ANAB	2.29	.349	-
	Agreement	C	ANAB	2.29	.349
ANA			3.03	.212	-
ANB			2.80	.262	-
AN			7.69	.004**	C>AN
ANAB		ANA	.74	.748	-
		ANB	.51	.830	-
		AN	5.39	.034*	ANAB>AN
ANA		ANB	-.23	.922	-
		AN	4.65	.064	-
ANB		AN	4.88	.059	-

Note. C=Control; ANAB=Article, Noun, and Adjective in bold font; ANA=Article, Noun, and Adjective; ANB=Article and Noun in bold font; AN=Article and Noun. **p* < .05, ***p* < .01, ****p* < .001

Controlling by new pairings the control group significantly outperforms the ANA and AN group, suggesting that without boldening treatment groups did worse than the control group, there are however no differences between treatment groups. The only exception being the ANB group, which outperformed the AN group, indicating that bolding might have had an effect on the adjective-noun combination groups. In the old pairing subdivision, the control group outperformed all groups (ANA, ANB, and AN) with the exception of the ANAB; again suggesting that the treatments did not differ significantly

from one another, but that ANAB still performed better than the rest and close to the control group (see Table C15 below).

Table C15
*Interaction between Pairing*Treatment*

Pairing	Treatment		Mean difference	<i>p</i> value	Direction of effect
New	C	ANAB	3.42	.074	-
		ANA	4.28	.025*	C>ANA
		ANB	3.27	.094	-
		AN	7.21	.001**	C>AN
	ANAB	ANA	.87	.629	-
		ANB	-.15	.935	-
		AN	3.80	.054	-
	ANA	ANB	-1.02	.579	-
		AN	2.93	.132	-
	ANB	AN	3.95	.050*	ANB>AN
		AN	3.13	.091	-
	Old	C	ANAB	3.13	.091
ANA			4.88	.008**	C>ANA
ANB			3.83	.043	C>ANB
AN			5.79	.004**	C>AN
ANAB		ANA	1.76	.312	-
		ANB	.70	.696	-
		AN	2.67	.161	-
ANA		ANB	-1.06	.551	-
		AN	.90	.630	-
ANB		AN	1.96	.310	-

Note. C=Control; ANAB=Article, Noun, and Adjective in bold font; ANA=Article, Noun, and Adjective; ANB=Article and Noun in bold font; AN=Article and Noun. **p* < .05, ***p* < .01, ****p* < .001

Controlling by concordance and pairing, other differences arose between treatments (for a full account refer to Table C16). Within assignment and new pairing, only the control group significantly outperformed other groups, the ANAB, ANA, and AN, with no significant differences across groups. Among old pairings, the only significant differences showed the control group outperforming the ANA and AN groups, indicating again that without bolding performance decreased. Regarding the interaction between agreement and new pairs the AN group significantly underperformed in contrast to all other groups,

suggesting that, again, the being exposed to only an article noun combination without any enhancement cause participants to underperform. Lastly, in the old pairings AN also significantly underperformed with regard to the ANB and the control groups, suggesting again a lack of sensibility to the treatments.

Table C16
Interaction between Concordance Pairing*Treatment*

Concordance	Pairing	Treatment	Mean difference	<i>p</i> value	Direction of effect		
Assignment	New	C	ANAB	4.94	.037*	C>ANAB	
			ANA	6.07	.010*	C>ANA	
			ANB	4.36	.071	-	
			AN	5.78	.023*	C>AN	
		ANAB	ANA	1.12	.613	-	
			ANB	-.58	.800	-	
			AN	.83	.730	-	
			ANA	ANB	-1.70	.453	-
		ANA	ANB	AN	-0.29	.904	-
			ANB	AN	1.41	.566	-
			C	ANAB	3.56	.110	-
				ANA	6.20	.006**	C>ANA
	ANB	4.21		.064	-		
	AN	4.85		.043*	C>AN		
	ANAB	ANA	2.64	.207	-		
		ANB	.657	.761	-		
		AN	1.3	.570	-		
		ANA	ANB	-1.99	.353	-	
	ANA	AN	-1.35	.551	-		
		ANB	AN	.640	.783	-	
C		ANAB	1.89	.498	-		
		ANA	2.50	.365	-		
	ANB	2.17	.446	-			
	AN	8.65	.004**	C>AN			
ANAB	ANA	.61	.816	-			
	ANB	.28	.918	-			
	AN	6.76	.020*	ANAB>AN			
	ANA	ANB	-.33	.901	-		
ANA	AN	6.15	.032*	ANA>AN			
	ANB	AN	6.48	.028*	ANB>AN		
	Old	C	ANAB	2.69	.286	-	
			ANA	3.57	.155	-	
ANB			3.44	.183	-		
AN			6.72	.014*	C>AN		

	ANAB	ANA	.87	.714	-
		ANB	.75	.762	-
		AN	4.03	.123	-
	ANA	ANB	-.13	.958	-
		AN	3.16	.221	-
	ANB	AN	3.28	.216	-

Note. C=Control; ANAB=Article, Noun, and Adjective in bold font; ANA=Article, Noun, and Adjective; ANB=Article and Noun in bold font; AN=Article and Noun. * $p < .05$, ** $p < .01$, *** $p < .001$

RQ4: What differences in performance are there between concordance, noun ending, gender and treatment?

Table C17

Differences between treatment groups

Treatment		Mean difference	p value	Direction of effect
C	ANAB	6.36	.222	-
	ANA	8.22	.031*	C>ANA
	ANB	7.39	.094	-
	AN	9.9	.010*	C>AN
ANAB	ANA	1.88	1.000	-
	ANB	1.04	1.000	-
	AN	3.55	1.000	-
ANA	ANB	-.83	1.000	-
	AN	1.68	1.000	-
ANB	AN	2.51	1.000	-

Note. C=Control; ANAB=Article, Noun, and Adjective in bold font; ANA=Article, Noun, and Adjective; ANB=Article and Noun in bold font; AN=Article and Noun. * $p < .05$, ** $p < .01$, *** $p < .001$

Table C18

Direction of effects between Ending

Ending		Mean difference	p value	Direction of effect
C	NC	39.2	.000***	C>NC
	E	80.41	.000***	C>E
NC	E	41.21	.000***	NC>E

Note. Noun Ending are coded as C=Canonical, NC=Non-Canonical, E=Exception. * $p < .05$, ** $p < .01$, *** $p < .001$

Table C19
*Interaction between Ending*Treatment*

Ending	Treatment		Mean difference	<i>p</i> value	Direction of effect	
Canonical	C	ANAB	-1.21	.794	-	
		ANA	3.53	.443	-	
		ANB	3.03	.522	-	
		AN	10.18	.043*	C>AN	
	ANAB	ANA	4.73	.281	-	
		ANB	4.24	.349	-	
		AN	11.39	.019*	ANAB>AN	
	ANA	ANB	-.49	.913	-	
		AN	6.66	.162	-	
	ANB	AN	7.15	.144	-	
	Non-Canonical	C	ANAB	17.9	.001**	C>ANAB
			ANA	20.11	.000***	C>ANA
ANB			17.97	.001**	C>ANB	
AN			24.00	.000***	C>AN	
ANAB		ANA	2.22	.645	-	
		ANB	.08	.988	-	
		AN	6.11	.245	-	
ANA		ANB	-2.14	.663	-	
		AN	3.89	.454	-	
ANB		AN	6.04	.260	-	
Exception		C	ANAB	2.38	.523	-
			ANA	1.03	.781	-
	ANB		1.16	.759	-	
	AN		-4.48	.263	-	
	ANAB	ANA	-1.35	.700	-	
		ANB	-1.21	.738	-	
		AN	-6.85	.076	-	
	ANA	ANB	-.14	.969	-	
		AN	-5.64	.149	-	
	ANB	AN	-5.64	.150	-	

Note. C=Control; ANAB=Article, Noun, and Adjective in bold font; ANA=Article, Noun, and Adjective; ANB=Article and Noun in bold font; AN=Article and Noun. **p* < .05, ***p* < .01, ****p* < .001

Significant interactions across factors can be found on Table C20 below. Considering the complexity and multiple factors involved in the comparison, only significant interactions are reported. The only regular and recurrent interactions found reveal that that within assignment of non-canonical masculine nouns, the AN group is significantly

outperformed by the other groups, except for the control group, thus suggesting again that the lack of enhancement and adjective had a negative effect on participants' performance. We find this pattern in other interactions, however, they do not hold across all treatments or within its subcategory completely. The other pattern we find is in assignment of non-canonical feminine nouns, but in this case it is the control group that again outperforms the treatment groups, showing that there are no differences in performance across groups.

Table C20
Interaction between Concordance Ending* Gender* Treatment*

Concordance	Ending	Gender	Treatment		Mean difference	<i>p</i> value	Direction of effect	
Assignment	Canonical	Masculine	AN	C	-16.36	.022*	AN<C	
			AN	ANAB	-16.11	.019*	AN< ANAB	
		Non-Canonical	Masculine	AN	C	-13.89	.049*	AN<C
				AN	C	-33.28	.000***	AN<C
			AN	ANAB	-19.86	.019*	AN< ANAB	
			AN	ANA	-18.98	.024*	AN< ANA	
			AN	ANB	-18.46	.032*	AN< ANB	
	Feminine	C	ANAB	24.75	.000***	C > ANAB		
		C	ANA	26.45	.000***	C > ANA		
		C	ANB	21.34	.003**	C > ANB		
	Agreement	Exception	Feminine	ANAB	AN	-11.94	.024*	ANAB <AN
				C	ANA	16.80	.023*	C> ANA
		Non-Canonical	Masculine	C	ANB	14.96	.049*	C> ANB
				C	AN	30.89	.000***	C>AN
ANB				AN	15.93	.042*	ANB >AN	
C				ANAB	21.00	.006**	C> ANAB	
Feminine			C	ANA	22.90	.003**	C> ANA	
			C	ANB	20.77	.008**	C> ANB	

Note. C=Control; **ANAB**=Article, Noun, and Adjective in bold font; **ANA**=Article, Noun, and Adjective; **ANB**=Article and Noun in bold font; AN=Article and Noun. **p* < .05, ***p* < .01, ****p* < .001

RQ5: What differences are there between groups regarding the gender of nouns and the immediate and delayed post-test in assignment?

Table C21

Differences between treatment groups in the delayed post-test

Treatment		Mean difference	<i>p</i> value	Direction of effect
C	ANAB	3.95	.484	-
	ANA	6.67	.009**	C>ANA
	ANB	5.36	.091	-
	AN	5.41	.122	-
ANAB	ANA	2.72	1.000	-
	ANB	1.42	1.000	-
	AN	1.46	1.000	-
ANA	ANB	-1.31	1.000	-
	AN	-1.26	1.000	-
ANB	AN	.04	1.000	-

Note. C=Control; ANAB=Article, Noun, and Adjective in bold font; ANA=Article, Noun, and Adjective; ANB=Article and Noun in bold font; AN=Article and Noun. **p* < .05, ***p* < .01, ****p* < .001

Table C22 below shows all relevant interactions, which are only present in the immediate post-test, none in the delayed post-test. Within the immediate post-test significant interactions are only found among masculine nouns. The ANA group is outperformed by the control group and AN underperformed significantly with regard to the control and the ANAB groups. This shows an advantage for enhanced input and more exposure to agreement.

Table C22

*Interaction between Test*Gender*Treatment in the delayed post-test*

Test	Gender	Treatment		Mean difference	<i>p</i> value	Direction of effect
Immediate Post-test	Masculine	C	ANAB	3.23	.338	-
			ANA	6.91	.040*	C>ANA
			ANB	6.28	.070	-
			AN	11.47	.002**	C>AN
		ANAB	ANA	3.67	.249	-
			ANB	3.04	.364	-
			AN	8.23	.019*	ANAB>AN

		ANA	ANB	-.63	.846	-
			AN	4.56	.186	-
	Feminine	ANB	AN	5.19	.144	-
		C	ANAB	7.27	.075	-
			ANA	7.82	.053	-
			ANB	6.77	.103	-
			AN	4.12	.344	-
		ANAB	ANA	.554	.885	-
			ANB	-.50	.900	-
			AN	-3.15	.451	-
		ANA	ANB	-1.05	.788	-
			AN	-3.70	.371	-
	Masculine	ANB	AN	-2.65	.533	-
Delayed		C	ANAB	.64	.837	-
Post-test			ANA	5.93	.054	-
			ANB	6.17	.052	-
			AN	5.27	.114	-
		ANAB	ANA	5.30	.071	-
			ANB	5.53	.068	-
			AN	4.63	.147	-
		ANA	ANB	.23	.937	-
			AN	-.67	.832	-
	Feminine	ANB	AN	-.90	.780	-
		C	ANAB	4.65	.149	-
			ANA	6.00	.061	-
			ANB	2.23	.496	-
			AN	.77	.823	-
		ANAB	ANA	1.36	.653	-
			ANB	-2.42	.439	-
			AN	-3.88	.241	-
		ANA	ANB	-3.78	.223	-
			AN	-5.24	.112	-
		ANB	AN	-1.46	.665	-

Note. C=Control; ANAB=Article, Noun, and Adjective in bold font; ANA=Article, Noun, and Adjective; ANB=Article and Noun in bold font; AN=Article and Noun. * $p < .05$, ** $p < .01$, *** $p < .001$

RQ6: What differences are there between groups regarding the gender of nouns and the immediate and delayed post-test in agreement?

Table C23

Differences between treatment groups in the delayed post-test

Treatment		Mean difference	<i>p</i> value	Direction of effect
C	ANAB	3.90	.502	-
	ANA	6.55	.011*	C>ANA
	ANB	5.34	.090	-
	AN	5.28	.139	-
ANAB	ANA	2.66	1.000	-
	ANB	1.45	1.000	-
	AN	1.38	1.000	-
ANA	ANB	-1.21	1.000	-
	AN	-1.27	1.000	-
ANB	AN	-.06	1.000	-

Note. C=Control; ANAB=Article, Noun, and Adjective in bold font; ANA=Article, Noun, and Adjective; ANB=Article and Noun in bold font; AN=Article and Noun. **p* < .05, ***p* < .01, ****p* < .001

Table C24

*Interaction between Test*Treatment in the delayed post-test*

Test	Treatment	Mean difference	<i>p</i> value	Direction of effect	
Immediate Post-Test	C	ANAB	7.41	.021*	C>ANAB
		ANA	8.92	.005**	C>ANA
		ANB	6.91	.035*	C>ANB
		AN	12.61	.000***	C>AN
	ANAB	ANA	1.51	.616	-
		ANB	-.50	.871	-
		AN	5.20	.114	-
	ANA	ANB	-2.01	.513	-
		AN	3.69	.256	-
	ANB	AN	5.70	.089	-
		C	ANAB	3.21	.212
	Delayed Post-Test	C	ANA	8.41	.001**
ANB			2.77	.291	-
AN			4.36	.115	-
ANAB		ANA	5.20	.034*	ANAB>ANA
		ANB	-.44	.861	-
		AN	1.15	.662	-
ANA		ANB	-5.63	.025*	ANA>ANB
		AN	-4.04	.125	-
ANB		AN	1.59	.555	-

Note. C=Control; ANAB=Article, Noun, and Adjective in bold font; ANA=Article, Noun, and Adjective; ANB=Article and Noun in bold font; AN=Article and Noun. * $p < .05$, ** $p < .01$, *** $p < .001$

As stated above, no interactions were significant. Table C25 below shows all the significant interactions. In the immediate post-test it is only the control group that significantly outperforms the groups ANA and AN in masculine nouns and the AN group, who underperforms significantly with regard to the ANAB and ANB groups, the combination of these results suggests that the presence of the enhancement improved performance. Among feminine nouns, the control group significantly outperformed the ANAB, ANA, and AN groups. In the delayed post-test when accounting for masculine nouns there were less significant interactions, with just the ANA group underperforming in comparison to the control and the ANAB groups. Among feminine nouns ANA underperformed significantly with regard to the control and ANB group, which could indicate that the unenhanced version with adjectives overwhelmed and confused participants. However, given that the overall interactions were not significant, results need to be considered within the context.

Table C25
*Interaction between Test*Gender*Treatment in the delayed post-test*

Test	Gender	Treatment	Mean difference	p value	Direction of effect	
Immediate Post-test	Masculine	C	ANAB	5.11	.218	-
			ANA	8.27	.045*	C>ANA
			ANB	6.35	.134	-
			AN	15.66	.001**	C>AN
		ANAB	ANA	3.17	.417	-
			ANB	1.25	.757	-
	AN		10.56	.014*	ANAB>AN	
	Feminine	ANA	ANB	-1.20	.630	-
			AN	7.39	.082	-
			ANB	9.31	.034*	ANB>AN
		C	ANAB	9.72	.013*	C>ANAB

Delayed Post-test	Masculine	ANAB	ANA	9.57	.014*	C>ANA
			ANB	7.47	.060	-
			AN	9.56	.023*	C>AN
		ANA	ANB	-1.16	.966	-
			ANB	-2.25	.550	-
			AN	-1.16	.968	-
		ANB	ANB	-2.09	.574	-
			AN	-.00	.999	-
			AN	2.09	.606	-
		C	ANAB	1.61	.672	-
	ANA		10.04	.009**	C>ANA	
	ANB		6.11	.117	-	
	Feminine	ANAB	ANA	8.43	.020*	ANAB>ANA
			ANB	4.50	.226	-
			AN	5.19	.187	-
		ANA	ANB	-3.94	.284	-
			AN	-3.25	.403	-
			AN	.69	.863	-
		C	ANAB	4.81	.147	-
			ANA	6.77	.041*	C>ANA
ANB			-.56	.867	-	
ANAB		ANA	1.93	.586	-	
	ANA	1.96	.530	-		
	ANB	-5.37	.097	-		
ANA	AN	-2.88	.398	-		
	ANB	-7.33	.023*	ANA<ANB		
	AN	-4.84	.154	-		
ANB	AN	2.49	.473	-		

Note. C=Control; ANAB=Article, Noun, and Adjective in bold font; ANA=Article, Noun, and Adjective; ANB=Article and Noun in bold font; AN=Article and Noun. * $p < .05$, ** $p < .01$, *** $p < .001$

RQ7: What differences are there between groups regarding noun ending and the immediate and delayed post-test in assignment?

Table C26

Differences between treatment groups in the delayed post-test

Treatment		Mean difference	p value	Direction of effect
C	ANAB	5.31	.364	-
	ANA	8.66	.007**	C>ANA
	ANB	4.84	.611	-
	AN	8.49	.021*	C>AN
ANAB	ANA	3.35	1.000	-
	ANB	-.47	1.000	-

	AN	3.18	1.000	-
ANA	ANB	-3.82	1.000	-
	AN	-.17	1.000	-
ANB	AN	3.65	1.000	-

Note. C=Control; ANAB=Article, Noun, and Adjective in bold font; ANA=Article, Noun, and Adjective; ANB=Article and Noun in bold font; AN=Article and Noun. * $p < .05$, ** $p < .01$, *** $p < .001$

Table C27

Direction of effects between Ending in the delayed post-test

Ending		Mean difference	p value	Direction of effect
C	NC	41.66	.000***	C>NC
	E	85.83	.000***	C>E
NC	E	44.17	.000***	NC>E

Note. Noun Ending are coded as C=Canonical, NC=Non-Canonical, E=Exception. * $p < .05$, ** $p < .01$, *** $p < .001$

Table C28

*Interaction between Ending*Treatment in the delayed post-test*

Ending	Treatment		Mean difference	p value	Direction of effect		
Canonical	C	ANAB	-.71	.822	-		
		ANA	2.47	.432	-		
		ANB	1.73	.593	-		
		AN	2.99	.380	-		
	ANAB	ANA	3.19	.288	-		
		ANB	2.44	.430	-		
		AN	3.70	.258	-		
		ANA	-.74	.808	-		
	ANB	AN	.52	.873	-		
		AN	1.26	.705	-		
		Non-Canonical	C	ANAB	11.47	.006**	C>ANAB
				ANA	15.78	.000**	C>ANA
ANB	11.92			.005**	C>ANB		
AN	15.27			.001**	C>AN		
ANAB	ANA	ANA	4.30	.263	-		
		ANB	.44	.991	-		
		AN	3.80	.365	-		
		ANB	-3.86	.325	-		
	ANA	AN	-.51	.903	-		
		ANB	3.36	.431	-		
		Exception	C	ANAB	.93	.756	-
				ANA	1.41	.635	-
ANB	2.38			.435	-		
AN	-2.42			.450	-		

ANAB	ANA	.48	.866	-
	ANB	1.45	.618	-
	AN	-3.35	.277	-
ANA	ANB	.98	.735	-
	AN	3.83	.211	-
ANB	AN	-4.81	.128	-

Note. C=Control; ANAB=Article, Noun, and Adjective in bold font; ANA=Article, Noun, and Adjective; ANB=Article and Noun in bold font; AN=Article and Noun. * $p < .05$, ** $p < .01$, *** $p < .001$

Table C29

*Interaction between Ending*Test in the delayed post-test*

Ending	Test		Mean difference	p value	Direction of effect
Canonical	IPT	DPT	-2.71	.012*	IPT<DPT
Non-canonical	IPT	DPT	2.11	.163	-
Exception	IPT	DPT	3.37	.002**	IPT>DPT

Note. IPT=Immediate Post-Test; DPT=Delayed Post-Test. * $p < .05$, ** $p < .01$, *** $p < .001$

Table C30

*Interaction between Treatment*Test*Ending in the delayed post-test*

Test	Ending	Treatment	Mean difference	p value	Direction of effect
Immediate Post-test	Non-Canonical	C ANAB	17.07	.001**	C>ANAB
		ANA	19.18	.000***	C>ANA
		ANB	16.61	.002**	C>ANB
		AN	22.21	.000***	C>AN

Note. C=Control; ANAB=Article, Noun, and Adjective in bold font; ANA=Article, Noun, and Adjective; ANB=Article and Noun in bold font; AN=Article and Noun. Only significant interactions are reported. * $p < .05$, ** $p < .01$, *** $p < .001$

RQ8: What differences are there between groups regarding noun ending and the immediate and delayed post-test in agreement?

Table C31

Differences between treatment groups in the delayed post-test

Treatment		Mean difference	<i>p</i> value	Direction of effect
C	ANAB	5.59	.297	-
	ANA	9.20	.004**	C>ANA
	ANB	5.05	.536	-
	AN	8.33	.028*	C>AN
ANAB	ANA	3.61	1.000	-
	ANB	-.54	1.000	-
	AN	2.73	1.000	-
ANA	ANB	-4.15	.933	-
	AN	-.88	1.000	-
ANB	AN	3.27	1.000	-

Note. C=Control; ANAB=Article, Noun, and Adjective in bold font; ANA=Article, Noun, and Adjective; ANB=Article and Noun in bold font; AN=Article and Noun. **p* < .05, ***p* < .01, ****p* < .001

Table C32

Direction of effects between Ending in the delayed post-test

Ending		Mean difference	<i>p</i> value	Direction of effect
C	NC	40.16	.000***	C>NC
	E	81.86	.000***	C>E
NC	E	41.70	.000***	NC>E

Note. Noun Ending are coded as C=Canonical, NC=Non-Canonical, E=Exception. **p* < .05, ***p* < .01, ****p* < .001

Table C33

*Interaction between Ending*Treatment in the delayed post-test*

Ending	Treatment		Mean difference	<i>p</i> value	Direction of effect
Canonical	C	ANAB	2.31	.611	-
		ANA	7.74	.088	-
		ANB	1.18	.798	-
		AN	9.69	.049*	C>AN
	ANAB	ANA	5.43	.207	-
		ANB	-1.13	.799	-
		AN	7.38	.116	-
	ANA	ANB	-6.56	.137	-
		AN	1.96	.673	-
	ANB	AN	8.51	.076	-

Non-Canonical	C	ANAB	13.25	.003**	C>ANAB	
		ANA	18.42	.000***	C>ANA	
		ANB	11.43	.012	-	
		AN	17.35	.000***	C>AN	
	ANAB	ANA	5.17	.213	-	
		ANB	-1.81	.671	-	
		AN	4.10	.365	-	
	ANA	ANB	-6.98	.101	-	
		AN	-1.07	.811	-	
	Exception	ANB	AN	5.91	.200	-
			C	ANAB	1.22	.653
		C	ANA	1.46	.589	-
ANB			2.54	.360	-	
AN			-2.06	.480	-	
ANAB		ANA	.23	.928	-	
		ANB	1.32	.619	-	
		AN	-3.29	.242	-	
ANA		ANB	1.09	.679	-	
		AN	-3.52	.207	-	
ANB	AN	-4.61	.109	-		

Note. C=Control; ANAB=Article, Noun, and Adjective in bold font; ANA=Article, Noun, and Adjective; ANB=Article and Noun in bold font; AN=Article and Noun. * $p < .05$, ** $p < .01$, *** $p < .001$

Table C34

*Interactions between Ending*Test in the delayed post-test*

Ending	Test		Mean difference	p value	Direction of effect
Canonical	IPT	DPT	-3.66	.023*	IPT<DPT
Non-canonical	IPT	DPT	-1.35	-	-
Exception	IPT	DPT	3.38	.001**	IPT>DPT

Note. IPT=Immediate Post-Test; DPT=Delayed Post-Test. * $p < .05$, ** $p < .01$, *** $p < .001$

Table C35

*Interaction between Treatment*Test*Ending in the delayed post-test*

Test	Ending	Treatment		Mean difference	p value	Direction of effect	
Immediate Post-test	Canonical	AN	C	-14.78	.015*	AN<C	
			ANAB	-14.07	.016*	AN<ANAB	
			ANB	-11.97	.044*	AN<ANB	
	Non-Canonical	C	C	ANAB	19.45	.001**	C>ANAB
				ANA	20.95	.000***	C>ANA
				ANB	15.69	.006**	C>ANB
				AN	25.19	.000***	C>AN

Delayed	Canonical	ANA	ANB	-9.27	.044*	ANA<ANB
Post-test		ANA	C	-15.88	.001**	ANA<C
			ANAB	-8.83	.046*	ANA<ANAB

Note. C=Control; ANAB=Article, Noun, and Adjective in bold font; ANA=Article, Noun, and Adjective; ANB=Article and Noun in bold font; AN=Article and Noun. Only significant interactions are reported. * $p < .05$, ** $p < .01$, *** $p < .001$

Correlations

Table C36

Correlations between dependent and independent variables

	Variables	Proficiency	Answer sheet	Form recognition
Accuracy	Nouns	.374**	.392**	.515***
	Adjectives	.138	.606***	.423***
Frequency	High	.329**	.292*	.502***
	Low	.359**	.457***	.451***
Assignment	Feminine	.436***	.262*	.290**
Agreement	Feminine	.409***	.290**	.257**
Assignment	Canonical	.111	.310**	.054
	Non-canonical	.325***	.227*	.120
	Exceptions	.158	.075	.231*
Agreement	Canonical	.169	.232*	.055
	Non-canonical	.350***	.252*	.159
	Exceptions	.186	.091	.262*
Assignment	New pairing	.315***	.309**	.240*
	Old pairing	.200*	.150	-.007
Agreement	New pairing	.224*	.105	.051
	Old pairing	.104	.128	.108
Assignment	Feminine DPT	.258*	.202	.193
Agreement	Feminine DPT	.250*	.231*	.246*
Assignment	Canonical DPT	.007	.222*	.020
Assignment	Non-canonical DPT	.281**	.165	.310
Agreement	Canonical DPT	.073	.217*	.158
	Non-Canonical DPT	.241*	.205	.359***

Note. Unless otherwise indicated, results refer to the immediate post-test. DPT=Delayed Post-Test. * $p < .05$, ** $p < .01$, *** $p < .001$

Data coding

Example word: *casa* ‘house’ (feminine)

Example adjectives: *blanco* ‘white’ (variable, shows grammatical gender) and *verde* ‘green’ (invariable, does NOT show grammatical gender)

Assignment: The article matches the gender of the target noun, i.e. *la casa*

Agreement: The article and the adjective match for grammatical gender, even if the gender assigned is wrong. For instance, *la casa blanca* would be correct, but if a learner wrote *el casa blanco*, they would still know the ins and outs of making things agree.

See note after the table*

Table C
Data coding

Oversimplification	Code	Description	Example	Correct assignment	Agreement
Right	1	Assignment and agreement correct	<i>var.</i> <i>La casa blanca</i>	YES	YES
			<i>invar.</i> <i>La casa verde</i>	YES	YES
	2	Correct assignment and variable adjective made invariable	<i>La casa blanc</i>	YES	YES
Wrong	3	Correct assignment and invariable adjective forced to be variable to agree	<i>La casa verda</i>	YES	YES
	4	Correct assignment but wrong agreement	<i>La casa blanco</i>	YES	NO
	5	Correct assignment but invariable adjective made	<i>La casa verdo</i>	YES	NO

6	variable with wrong gender Wrong assignment but correct agreement	<i>var.</i>	<i>El casa blanca</i>	NO	NO
7	Wrong assignment with invariable adjective made variable with correct agreement		<i>El casa verda</i>	NO	NO
8	Wrong assignment with invariable adjective made variable with incorrect agreement		<i>El casa verdo</i>	NO	YES
9	Wrong assignment with a variable adjective made invariable		<i>El casa blanc</i>	NO	YES
10	Wrong assignment but consistent agreement		<i>El casa blanco</i>	NO	YES
11	Wrong assignment but correct agreement	<i>inv.</i>	<i>El casa verde</i>	NO	YES

* There should probably be another column to the right called “consistency”. Each noun appears twice, so consistency would have to take into consideration whether assignment

on the one hand has been correctly assigned in both instances or not. Same goes for agreement.

Examples:

consistent assignment and agreement: successful acquisition of gender and agreement

La casa blanca AND *La casa verde* → consistent assignment and agreement

La casa blanc AND *La casa verde* → consistent assignment and agreement

La casa blanc AND *La casa verda* → consistent assignment and agreement

OR

wrong but consistent assignment and agreement: learnt the wrong gender but acquired the mechanics of gender agreement

El casa blanco AND *El casa verde* → wrong but consistent assignment and agreement

El casa blanc AND *El casa verde* → wrong but consistent assignment and agreement

El casa blanco AND *El casa verdo* → wrong but consistent assignment and agreement

El casa blanc AND *El casa verdo* → wrong but consistent assignment and agreement

OR

consistent assignment but inconsistent agreement: only acquired gender but did not internalized the gender system

La casa blanc AND *La casa verdo* → consistent assignment but inconsistent agreement

La casa blanco AND *La casa verde* → consistent assignment but inconsistent agreement

La casa blanco AND *La casa verde* → consistent assignment but inconsistent agreement

La casa blanca AND *La casa verdo* → consistent assignment but inconsistent agreement

REFERENCES

- Ahmad, D. (2012). Intentional vs. incidental vocabulary learning. *Arts and Humanities, 1*(1).
- Ahmadi, S. R. M. (2017). The Impact of Incidental and Intentional L2 Vocabulary Learning on Depths and Breadth Dimensions of Vocabulary Knowledge. *The Buckingham Journal of Language and Linguistics, 10*, 1-17.
- Alarcón, I. V. (2011). Spanish gender agreement under complete and incomplete acquisition: Early and late bilinguals' linguistic behavior within the noun phrase. *Bilingualism: Language and Cognition, 14*(3), 332-350.
- Ansarin, A. A., & Bayazidi, A. (2016). Task type and incidental L2 vocabulary learning: Repetition versus task involvement load. *Southern African Linguistics and Applied Language Studies, 34*(2), 135-146.
- Arzt, J., & Kost, C. (2016). Effect of different teaching techniques on the acquisition of grammatical gender by beginning German second language learners. *Die Unterrichtspraxis/Teaching German, 49*(1), 1-23.
- Bebout, J., & Belke, E. (2017). Language play facilitates language learning: Optimizing the input for gender-like category induction. *Cognitive research: principles and implications, 2*(1), 11.
- Bell, P. K., & Collins, L. (2009). 'It's vocabulary'/'it's gender': learner awareness and incidental learning. *Language Awareness, 18*(3-4), 277-293.

- Blom, E., Polišenská, D., & Weerman, F. (2008). Articles, adjectives and age of onset: The acquisition of Dutch grammatical gender. *Second language research*, 24(3), 297-331.
- Blom, E., Polišenská, D., & Unsworth, S. (2008). The acquisition of grammatical gender in Dutch. *Second Language Research*, 24(3), 259-265.
- Bordag, D., Kirschenbaum, A., Opitz, A., Rogahn, M., & Tschirner, E. (2016). Incidental acquisition of grammatical features during reading in L1 and L2. *Studies in Second Language Acquisition*, 38(3), 445-483.
- Bordag, D., Kirschenbaum, A., Rogahn, M., & Tschirner, E. (2017). The role of orthotactic probability in incidental and intentional vocabulary acquisition L1 and L2. *Second Language Research*, 33(2), 147-178.
- Bruton, A., Lopez, M. G., & Mesa, R. E. (2011). Incidental L2 vocabulary learning: An impracticable term?. *TESOL Quarterly*, 45(4), 759-768.
- Bull, W. E. (1965). *Spanish for teachers: Applied linguistics*. Krieger Pub Co.
- Caffarra, S., Siyanova-Chanturia, A., Pesciarelli, F., Vespignani, F., & Cacciari, C. (2015). Is the noun ending a cue to grammatical gender processing? An ERP study on sentences in Italian. *Psychophysiology*, 52(8), 1019-1030.
- Chini, M. (1995). *Genere grammaticale e acquisizione: aspetti della morfologia nominale in italiano L2* (Vol. 14). Franco Angeli.

- Choo, L. B., Lin, D. T. A., & Pandian, A. (2012). Language learning approaches: A review of research on explicit and implicit learning in vocabulary acquisition. *Procedia-Social and Behavioral Sciences*, *55*, 852-860.
- Clegg, J. H. (2010). Native Spanish speaker intuition in noun gender assignment. *Language design: journal of theoretical and experimental linguistics*, *12*, 0005-18.
- Cobb, T. (1997). Is there any measurable learning from hands-on concordancing?. *System*, *25*(3), 301-315.
- Corbett, G. G. (1991). *Gender*. Cambridge University Press.
- Cornips, L., & Hulk, A. (2008). Factors of success and failure in the acquisition of grammatical gender in Dutch. *Second Language Research*, *24*(3), 267-295.
- Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of verbal learning and verbal behavior*, *11*(6), 671-684.
- DeKeyser, R. M. (1995). Learning second language grammar rules: An experiment with a miniature linguistic system. *Studies in second language acquisition*, *17*(3), 379-410.
- DeKeyser, R. (2008). 11 Implicit and Explicit Learning. *The handbook of second language acquisition*, *27*, 313.
- De Martino, M., Bracco, G., Postiglione, F., & Laudanna, A. (2017). The influence of grammatical gender and suffix transparency in processing Italian written nouns. *The Mental Lexicon*, *12*(1), 107-128.

- Denhovska, N., & Serratrice, L. (2017). Incidental learning of gender agreement in L2. *Journal of psycholinguistic research, 46*(5), 1187-1211.
- de Oliveira Santos, V. (2015). Can colors, voices, and images help learners acquire the grammatical gender of German nouns?. *Language Teaching Research, 19*(4), 473-498.
- Desrochers, A., Gélinas, C., & Wieland, L. D. (1989). An application of the mnemonic keyword method to the acquisition of German nouns and their grammatical gender. *Journal of Educational Psychology, 81*(1), 25.
- Dussias, P. E., Kroff, J. R. V., Tamargo, R. E. G., & Gerfen, C. (2013). When gender and looking go hand in hand: Grammatical gender processing in L2 Spanish. *Studies in Second Language Acquisition, 35*(2), 353-387.
- Ellis, N. (1993). Rules and instances in foreign language learning: Interactions of explicit and implicit knowledge. *European Journal of Cognitive Psychology, 5*(3), 289-318.
- Ellis, R. (1995). Modified oral input and the acquisition of word meanings. *Applied Linguistics, 16*(4), 409-441.
- Ellis, N. C., & Schmidt, R. (1997). Morphology and longer distance dependencies: Laboratory research illuminating the A in SLA. *Studies in Second Language Acquisition, 19*(2), 145-171.

- Ellis, R., & He, X. (1999). The roles of modified input and output in the incidental acquisition of word meanings. *Studies in second language acquisition*, 21(2), 285-301.
- Ellis, R., Loewen, S., & Erlam, R. (2006). Implicit and explicit corrective feedback and the acquisition of L2 grammar. *Studies in second language acquisition*, 28(2), 339-368.
- Fischer, U. (1994). Learning words from context and dictionaries: An experimental comparison. *Applied Psycholinguistics*, 15(4), 551-574.
- Foote, R. (2011). Integrated knowledge of agreement in early and late English–Spanish bilinguals. *Applied Psycholinguistics*, 32(1), 187-220.
- Foucart, A., & Frenck-Mestre, C. (2011). Grammatical gender processing in L2: Electrophysiological evidence of the effect of L1–L2 syntactic similarity. *Bilingualism: Language and Cognition*, 14(3), 379-399.
- Gabriele, A., Fiorentino, R., & Bañón, J. A. (2013). Examining second language development using event-related potentials: a cross-sectional study on the processing of gender and number agreement. *Linguistic Approaches to Bilingualism*, 3(2), 213-232.
- Gass, S. (1999). Discussion: Incidental Vocabulary Learning. *Studies in Second Language Acquisition*, 21(2), 319.

- Godfroid, A., Ahn, J., Choi, I., Ballard, L., Cui, Y., Johnston, S., ... & Yoon, H. J. (2018). Incidental vocabulary learning in a natural reading context: an eye-tracking study. *Bilingualism: Language and Cognition*, 21(3), 563-584.
- Grey, S., Williams, J. N., & Rebuschat, P. (2014). Incidental exposure and L3 learning of morphosyntax. *Studies in Second Language Acquisition*, 36(4), 611-645.
- Grüter, T., Lew-Williams, C., & Fernald, A. (2012). Grammatical gender in L2: A production or a real-time processing problem?. *Second Language Research*, 28(2), 191-215.
- Hawkins, R. (2005). Revisiting wh-movement: The availability of an uninterpretable [wh] feature in interlanguage grammars. In *Proceedings of the 7th Generative Approaches to Second Language Acquisition Conference (GASLA 2004)* (Vol. 124, p. 137). Somerville, MA: Cascadilla.
- Hawkins, R., & Casillas, G. (2008). Explaining frequency of verb morphology in early L2 speech. *Lingua*, 118(4), 595-612.
- Hawkins, R., & Chan, C. Y. H. (1997). The partial availability of Universal Grammar in second language acquisition: The „failed functional features hypothesis“. *Second Language Research*, 13(3), 187-226.
- Holley, F. M. (1973). A Study of Vocabulary Learning in Context: The Effect of New-Word Density in German Reading Materials. *Foreign Language Annals*, 6(3), 339-347.

- Hopp, H. (2012). Grammatical gender in adult L2 acquisition: Relations between lexical and syntactic variability. *Second Language Research*, 29(1), 33-56.
- Hopp, H. (2016). Learning (not) to predict: Grammatical gender processing in second language acquisition. *Second Language Research*, 32(2), 277-307.
- Hopp, H., & Lemmerth, N. (2018). Lexical and syntactic congruency in L2 predictive gender processing. *Studies in Second Language Acquisition*, 40(1), 171-199.
- Hu, H. C. M., & Nassaji, H. (2016). Effective vocabulary learning tasks: Involvement load hypothesis versus technique feature analysis. *System*, 56, 28-39.
- Huckin, T., & Coady, J. (1999). Incidental vocabulary acquisition in a second language: A review. *Studies in second language acquisition*, 21(2), 181-193.
- Hulstijn, J. H. (1995). Not all grammar rules are equal: Giving grammar instruction its proper place in foreign language teaching. *Attention and awareness in foreign language learning*, 359-386.
- Hulstijn, J. (2008). 12 Incidental and Intentional Learning. *The handbook of second language acquisition*, 27, 349.
- Hulstijn, J., Hollander, M., & Greidanus, T. (1996). Incidental vocabulary learning by advanced foreign students: The influence of marginal glosses, dictionary use, and reoccurrence of unknown words. *The Modern Language Journal*, 80, 327–339.
- Hulstijn, J. H., & Laufer, B. (2001). Some empirical evidence for the involvement load hypothesis in vocabulary acquisition. *Language learning*, 51(3), 539-558.

- Izura, C., Cuertos, F., & Brysbaert, M. (2014). Lextale-Esp: A test to rapidly and efficiently assess the Spanish vocabulary size. *Psicologica*, 35, 49-66.
- Jedynak, M., & Pytlarz, J. (2012). The issue of gender in multiple language acquisition. *Brno studies in English*, 38(1).
- Krashen, S. (1985). *The Input Hypothesis*. London: Longman.
- Lardiere, D. (2008). Feature assembly in second language acquisition. In Liceras J, Zobl H and Goodluck H (Eds.). *The role of formal features in second language acquisition*, 106-140.
- Lardiere, D. (2009). Some thoughts on the contrastive analysis of features in second language acquisition. *Second language research*, 25(2), 173-227.
- Lee, J. F. (2002). The incidental acquisition of Spanish: Future tense morphology through reading in a second language. *Studies in Second Language Acquisition*, 24(1), 55-80.
- Lee, S. K., & Huang, H. T. (2008). Visual input enhancement and grammar learning: A meta-analytic review. *Studies in Second Language Acquisition*, 30(3), 307-331.
- Lemhöfer, K., Schriefers, H., & Hanique, I. (2010). Native language effects in learning second-language grammatical gender: A training study. *Acta Psychologica*, 135(2), 150-158.
- Leow, R. P. (1998). Toward operationalizing the process of attention in SLA: Evidence for Tomlin and Villa's (1994) finegrained analysis of attention. *Applied Psycholinguistics*, 19(1), 133-159.

- Lew-Williams, C. (2009). *Real-Time Processing of Gender-Marked Articles by Native and Non-Native Spanish-Speaking Children and Adults*. ProQuest LLC. 789 East Eisenhower Parkway, PO Box 1346, Ann Arbor, MI 48106.
- Lew-Williams, C., & Fernald, A. (2007). Young children learning Spanish make rapid use of grammatical gender in spoken word recognition. *Psychological Science*, *18*(3), 193-198.
- Lew-Williams, C., & Fernald, A. (2009). Fluency in using morphosyntactic cues to establish reference: How do native and non-native speakers differ. In *Proceedings of the 33rd annual boston university conference on language development* (Vol. 1, pp. 290-301).
- Lew-Williams, C., & Fernald, A. (2010). Real-time processing of gender-marked articles by native and non-native Spanish speakers. *Journal of memory and language*, *63*(4), 447-464.
- Loewen, S. (2005). Incidental focus on form and second language learning. *Studies in second language acquisition*, *27*(3), 361-386.
- Long, M. H. (1981). Input, interaction, and second-language acquisition. *Annals of the New York academy of sciences*, *379*(1), 259-278.
- Long, M. H. (1991). "Focus on form: a design feature in language teaching methodology". In de Bot, K.; Ginsberg, R.; Kramersch, C. (eds.). *Foreign language research in cross-cultural perspective*. Amsterdam: John Benjamins. pp. 39–52.

- Lyster, R. (2010). Enseignement centré sur la forme et acquisition du genre grammatical en français L2. *Canadian Journal of Applied Linguistics/Revue canadienne de linguistique appliquée*, 13(2), 73-94.
- Mackey, A. (1999). Input, interaction, and second language development: An empirical study of question formation in ESL. *Studies in second language acquisition*, 21(4), 557-587.
- Malone, J. (2018). Incidental vocabulary learning in SLA: effects of frequency, aural enhancement, and working memory. *Studies in Second Language Acquisition*, 40(3), 651-675.
- Martínez-Gibson, E. A. (2011). A comparative study on gender agreement errors in the spoken Spanish of heritage speakers and second language learners. *Porta Linguarum: revista internacional de didáctica de las lenguas extranjeras*, (15), 177-193.
- Matteini, S. (2014). L2 multiple nominal feature agreement in German: the acquisition of grammatical gender in a formal environment. *International Review of Applied Linguistics in Language Teaching*, 52(4), 313-355.
- McDonough, K. I. M., & Trofimovich, P. (2016). The role of statistical learning and working memory in L2 speakers' pattern learning. *The Modern Language Journal*, 100(2), 428-445.
- Mirzaii, M. (2012). Implicit Vs. Explicit Vocabulary Learning: Which Approach Serves Long-term Recall Better?. *3L: Language, Linguistics, Literature*, 18(2).

- Montrul, S., Foote, R., & Perpiñán, 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.
- Morgan-Short, K., & Bowden, H. W. (2006). Processing instruction and meaningful output-based instruction: effects on second language development. *Studies in Second Language Acquisition*, 28(1), 31-65.
- Morgan-Short, K., Sanz, C., Steinhauer, K., & Ullman, M. T. (2010). Second language acquisition of gender agreement in explicit and implicit training conditions: An event-related potential study. *Language learning*, 60(1), 154-193.
- Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language learning*, 50(3), 417-528.
- Pellicer-Sánchez, A. (2016). Incidental L2 vocabulary acquisition from and while reading: An eye-tracking study. *Studies in Second Language Acquisition*, 38(1), 97-130.
- Presson, N., Macwhinney, B., & Tokowicz, N. (2014). Learning grammatical gender: The use of rules by novice learners. *Applied Psycholinguistics*, 35(4), 709-737.
- Ramos, F. D. R. (2015). Incidental vocabulary learning in second language acquisition: A literature review. *Profile Issues in Teachers Professional Development*, 17(1), 157-166.
- Robinson, P. (1996). Learning simple and complex second language rules under implicit, incidental, rule-search, and instructed conditions. *Studies in second language acquisition*, 18(1), 27-67.

- Robinson, P. (2005). Cognitive abilities, chunk-strength, and frequency effects in implicit artificial grammar and incidental L2 learning: Replications of Reber, Walkenfeld, and Hernstadt (1991) and Knowlton and Squire (1996) and their relevance for SLA. *Studies in Second Language Acquisition*, 27(2), 235-268.
- Rodina, Y., & Westergaard, M. (2017). Grammatical gender in bilingual Norwegian–Russian acquisition: The role of input and transparency. *Bilingualism: Language and cognition*, 20(1), 197-214.
- Rosa, E., & O'Neill, M. D. (1999). Explicitness, intake, and the issue of awareness: Another piece to the puzzle. *Studies in second language acquisition*, 21(4), 511-556.
- Rott, S. (1999). The effect of exposure frequency on intermediate language learners' incidental vocabulary acquisition and retention through reading. *Studies in second language acquisition*, 21(4), 589-619.
- Sagarra, N., & Herschensohn, J. (2013). Processing of gender and number agreement in late Spanish bilinguals. *International Journal of Bilingualism*, 17(5), 607-627.
- Santoro, M. (2012). Morphological Variability in interlanguage grammars: New evidence from the acquisition of gender and number in Italian determiner phrases and direct object pronouns. *Canadian Journal of Applied Linguistics/Revue canadienne de linguistique appliquée*, 15(1), 167-189.
- Sanz, C., & Morgan-Short, K. (2004). Positive evidence versus explicit rule presentation and explicit negative feedback: A computer-assisted study. *Language Learning*, 54(1), 35-78.

- Schmidt, R. W. (1990). The role of consciousness in second language learning¹. *Applied linguistics*, 11(2), 129-158.
- Schmidt, R. (1995). Consciousness and foreign language learning: A tutorial on the role of attention and awareness in learning. *Attention and awareness in foreign language learning*, 9, 1-63.
- Schwartz, B. D., & Sprouse, R. A. (1996). L2 cognitive states and the Full Transfer/Full Access model. *Second language research*, 12(1), 40-72.
- Shahpari, N., Shamshiri, H. R., & Rashidi, N. (2014). Intentional vs. Incidental vocabulary learning & Iranian EFL Learners' Retention. *Enjoy Teaching Journal*, 2(3).
- Siegelman, N., & Arnon, I. (2015). The advantage of starting big: Learning from unsegmented input facilitates mastery of grammatical gender in an artificial language. *Journal of Memory and Language*, 85, 60-75.
- Smith, N. V., & Tsimpli, I. M. (1995). *The mind of a savant: Language learning and modularity*. Blackwell Publishing.
- Spada, N., & Tomita, Y. (2010). Interactions between type of instruction and type of language feature: A meta-analysis. *Language learning*, 60(2), 263-308.
- Spinner, P. (2013). The second language acquisition of number and gender in Swahili: A Feature Reassembly approach. *Second Language Research*, 29(4), 455-479.
- Suzuki, Y., & DeKeyser, R. (2017). The interface of explicit and implicit knowledge in a second language: Insights from individual differences in cognitive aptitudes. *Language Learning*, 67(4), 747-790.

- Tammenga-Helmantel, M., Arends, E., & Canrinus, E. T. (2014). The effectiveness of deductive, inductive, implicit and incidental grammatical instruction in second language classrooms. *System*, 45, 198-210.
- Teng, F. (2015). The effectiveness of extensive reading on EFL learners' vocabulary learning: Incidental versus intentional learning. *BELT-Brazilian English Language Teaching Journal*, 6(1), 66-80.
- Teng, F. (2016). The effects of word exposure frequency on incidental learning of the depth of vocabulary knowledge. *GEMA Online® Journal of Language Studies*, 16(3).
- Teschner, R. V., & Russell, W. M. (1984). The gender patterns of Spanish nouns: An inverse dictionary-based analysis. *Hispanic Linguistics*, 1(1), 115-132.
- Tipurita, M. E., & Jean, G. (2014). Enseignement explicite du genre des noms en français: expérimentation au primaire en classe d'immersion. *Canadian Modern Language Review*, 70(3), 279-302.
- Tokowicz, N., & MacWhinney, B. (2005). Implicit and explicit measures of sensitivity to violations in second language grammar: An event-related potential investigation. *Studies in second language acquisition*, 27(2), 173-204.
- Ullman, M. T. (2001). A neurocognitive perspective on language: The declarative/procedural model. *Nature reviews neuroscience*, 2(10), 717.
- VanPatten, B., & Oikkenon, S. (1996). Explanation versus structured input in processing instruction. *Studies in Second Language Acquisition*, 18(4), 495-510.

- Varol, B., & Erçetin, G. (2016). Effects of working memory and gloss type on L2 text comprehension and incidental vocabulary learning in computer-based reading. *Procedia-Social and Behavioral Sciences*, 232, 759-768.
- Vigliocco, G., Vinson, D. P., Paganelli, F., & Dworzynski, K. (2005). Grammatical gender effects on cognition: implications for language learning and language use. *Journal of Experimental Psychology: General*, 134(4), 501.
- Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics*, 28, 46–65.
- Williams, J. N. (2005). Learning without awareness. *Studies in Second Language Acquisition*, 27(2), 269-304.
- Wust, V., & Roche, C. (2015). The Impact of Form-Focused Instruction and Consciousness-Raising Activities on L2 Learners' Acquisition of Grammatical Gender in French. *French Review*, 88(3), 37-52.
- Xie, H., Zou, D., Lau, R. Y., Wang, F. L., & Wong, T. L. (2015). Generating incidental word-learning tasks via topic-based and load-based profiles. *IEEE multimedia*, 23(1), 60-70.
- Zandieh, Z., & Jafarigohar, M. (2012). The Effects of Hypertext Gloss on Comprehension and Vocabulary Retention under Incidental and Intentional Learning Conditions. *English Language Teaching*, 5(6), 60-71.

