ATTENTION TO FORM AND MEANING:
LEARNING WITHOUT AWARENESS?
AN INTERPRETABLE AND UNINTERPRETABLE APPROACH

A Dissertation
Submitted to the Faculty of the
Graduate School of Arts and Sciences
of Georgetown University
in partial fulfillment of the requirement for the
degree of
Doctor of Philosophy
in Spanish

By

Miguel Ángel Novella Gómez, M.S.

Washington, DC
August 23, 2012
ATTENTION TO FORM AND MEANING:
LEARNING WITHOUT AWARENESS?
AN INTERPRETABLE AND UNINTERPRETABLE APPROACH

Miguel Ángel Novella-Gómez

Thesis Advisors: Ronald P. Leow, Ph.D. and Héctor Campos Ph.D.

ABSTRACT

While attentional models of SLA postulate attention as crucial for learning to take place (e.g., Robinson, 1995b; Schmidt, 1990, 2001; Tomlin & Villa, 1994), the role that awareness or lack thereof plays remains debatable. Indeed, the studies that have empirically addressed the construct of unawareness reveal conflicting results (e.g., Williams, 2005 vs. Hama & Leow, 2010). Interestingly, within the recent minimalist syntactic approach to L2 learning, the notion of detectability in meaningful contrasts (Lardiere, 2009) is being postulated to play a role. In this approach, language variability responds to the specification of two types of features: interpretable and uninterpretable, and there is a debate as to whether both features are available for processing. (cf. Tsimpli & Dimitrakopoulou, 2007; Lardiere, 2009). To provide a bridge between these two areas, this dissertation sought to address (1) whether there was a distinction between the performances of 62 aware and unaware participants (advanced students of Spanish) on implicit form-meaning mappings during an incidental reading task, (2) whether type of feature [+/- animate], [+/- count], [+/- feminine]
influenced the performances, and (3) whether participants’ assignment of gender was based on [+/- feminine] noun gender or if type of noun ending ruled gender assignment. Participants completed immediate and one-week delayed posttests, including recognition, controlled written production, and comprehension tasks. The data were analyzed both quantitatively and qualitatively.

The results revealed that awareness seems to be beneficial for further processing and learning of the target items; however, level of awareness appears to be a factor that needs to be considered, especially at the delayed posttest stage. A disassociation between unawareness and further processing or learning was found, more in line with Schmidt’s noticing (1990, 2001). The qualitative analysis of the think-aloud protocols revealed that participants adopted two different strategies. While both data-driven and conceptually-driven processes (Leow, 1998; Robinson, 1995b; Williams, 1999) were found, it was revealed that another variable, namely, type of feature, seemed to have contributed to the usage of one or the other. Finally, awareness or lack thereof did not appear to play a role in participants’ reliance on ending cues for gender assignment.
DEDICATION

To my parents, Antonia Gómez Gómez (1938-2010), and Miguel Ángel Novella Gómez (1935-2011)
ACKNOWLEDGEMENTS

I want to acknowledge many people that have helped me during this time. First of all – or, as it is said in Spanish, *primero lo primero* –, I want to thank Prof. Ron Leow and Prof. Héctor Campos. There are not enough words to thank these two Professors for all that they have done for me. As I mentioned once to Prof. Campos, I have had the opportunity to meet real professors during my life as a student. Prof. Leow and Prof. Campos are a good example of dedication, knowledge and humanity. You are an inspiration.

Thank you to my readers Prof. Donna Lardiere, and Prof. Michael Ferreira. I would never have imagined that after I took a language class (Language Acquisition), I would end up being lucky to have Prof. Lardiere as a reader. Thank you very much for meeting me at Boston College and explaining to me in depth the intricacies of feature acquisition and feature reassembly. I met Prof. Michael Ferreira at the very beginning of my Ph.D. process, and I always considered him a friend, now we are colleagues.

I also need to thank Teodora Bernadeta Baba, my wife, and my new member of my family for being all supportive during this period. After so many years of knowing each other, finally we are living together. Thank you for travelling the world with me, and above all, for taking me to Romania.

I want to also thank my parents and my siblings. Unfortunately, neither of my parents could not see the fruit of this long process; however, they were, and
they have been, present all this time. Thank you for everything. My sister, Mayra Roxana, and my brother, Víctor Manuel are such a wonderful example of what a brotherhood is. Without them, all this could not be have completed.

My family in Los Angeles, and in Mexico, have been there for me too. Thank you to my uncles and aunts in Los Angeles, Jorge, Lourdes, Miguel, Arturo, Magdalena, Rosa, Fernando. I will never forget that trip to Las Vegas to attend my marriage in that city. I want to thank also my family in Mexico for all my childhood memories. Thank you to all my aunts, uncles, and cousins. Both families know what an extended family is.

I would like to thank all those instructors who allowed me to solicit learners’ participation, Charles Nagle, Pedro Cruz, Patricia Soler, José Cornelio, Francisco Cornejo, Aitzi Delgado, and Prof. Thomas Walsh.

To my friends at the Spanish Department who, through conversation and moral support, made my time in Washington and this long process an easier one. Jonathan Mercer proofread some parts of this dissertation and made valuable comments. In exchange, I only took him and his girlfriend to the Dulles airport on his way to South America once. Cecily Raynor also took the time to proofread this dissertation, as well as Silvia Marijuan. Silvia and Jonathan assist me with the statistics section. Some other friends who I need to mention Matthew Vuskovich, María José Navia, Ana Martínez, and Mercedes López.
Finally, I would like to thank some people whom I met as UNAM student and employee: Martha Jurado, Martha Elena Cortés, Harold Ormsby, Lydia García, Selma Sablow, and Morris Sablow†. Thank you to all the people who have been nice to me and I did not mention.
# TABLE OF CONTENTS

INTRODUCTION ........................................................................................................... 1
STATEMENT OF THE PROBLEM .............................................................................. 1
DEFINITION OF TERMS ............................................................................................ 7

CHAPTER 1: THEORETICAL FRAMEWORK ............................................................... 12
INTRODUCTION ........................................................................................................ 12
ATTENTION AS A VITAL CONSTITUENT FOR SECOND LANGUAGE ACQUISITION 13
ATTENTION, AWARENESS IN THE COGNITIVE SCIENCES ................................. 15
Empirical studies in regards to (un)awareness in the non-SLA field ............. 21
ATTENTION AND AWARENESS IN L2 ACQUISITION ....................................... 24
MODELS OF ATTENTION AND AWARENESS IN SLA .................................. 25
Schmidt’s noticing hypothesis ................................................................. 25
Tomlin & Villa’s Fine-grained model of attention ....................................... 27
Robinson’s Model (1995b) of attention and memory ................................. 29
Learning without awareness ................................................................. 34
MEASURING ATTENTION/AWARENESS IN SLA ......................................... 55
Verbal reports ......................................................................................... 56
Verbal reports to measure attention, awareness, and noticing in SLA .... 57
Think-alouds criticism ............................................................................. 57

CHAPTER 2: UNIVERSAL GRAMMAR AND SECOND LANGUAGE ACQUISITION ......................................................................................................................... 61
INTRODUCTION ....................................................................................................... 61
Background ........................................................................................................ 62
THEORETICAL INTRODUCTION: CURRENT MINIMALIST THEORY AND UNIVERSAL
INDEPENDENT VARIABLES ................................................................. 115
DEPENDENT VARIABLES ................................................................. 115
MATERIALS .......................................................................................... 116
  Introduction ......................................................................................... 116
  Instructional package ........................................................................... 116
  Verbalization training materials ......................................................... 117
TREATMENT / EXPERIMENTAL TASKS ............................................... 118
  Reading Materials .............................................................................. 118
ASSESSMENT TESTS ............................................................................ 123
  Introduction ......................................................................................... 123
  Critical tests ....................................................................................... 123
  Recognition test: Multiple-choice (MC) ............................................. 125
  Production test: Fill-in-the-blank (FIB) .............................................. 126
  Reading test: Multiple-choice comprehension assessment task ......... 127
  Gender test ......................................................................................... 129
MEASUREMENT OF AWARENESS ....................................................... 130
PROCEDURE .......................................................................................... 132
  Think-aloud training task ................................................................. 134
  Vocabulary Pre-training session ....................................................... 134
  Training task ...................................................................................... 138
RESEARCH DESIGN OVERVIEW: EXPERIMENTAL DESIGN .............. 139
SCORING AND CODING ....................................................................... 140
QUALITATIVE ANALYSES ................................................................... 141
THINK-ALOUD PROTOCOLS ............................................................... 141
  Animate versus Inanimate ............................................................... 142
  Mass versus Count ........................................................................... 146
STATISTICAL ANALYSIS ..................................................................... 148
CHAPTER 4: RESULTS ......................................................................... 151
<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research Question 4</td>
<td>241</td>
</tr>
<tr>
<td>Research Question 5</td>
<td>243</td>
</tr>
<tr>
<td>Research Question 6</td>
<td>244</td>
</tr>
<tr>
<td>CONCLUSIONS, LIMITATIONS AND FUTURE RESEARCH</td>
<td>251</td>
</tr>
<tr>
<td>Overview</td>
<td>251</td>
</tr>
<tr>
<td>Conclusions</td>
<td>251</td>
</tr>
<tr>
<td>Pedagogical implications</td>
<td>257</td>
</tr>
<tr>
<td>Limitations and Future Research</td>
<td>259</td>
</tr>
<tr>
<td>Limitations</td>
<td>259</td>
</tr>
<tr>
<td>Future research</td>
<td>260</td>
</tr>
<tr>
<td>APPENDIX A</td>
<td>262</td>
</tr>
<tr>
<td>APPENDIX B</td>
<td>264</td>
</tr>
<tr>
<td>APPENDIX C</td>
<td>273</td>
</tr>
<tr>
<td>APPENDIX D</td>
<td>274</td>
</tr>
<tr>
<td>APPENDIX E</td>
<td>282</td>
</tr>
<tr>
<td>APPENDIX F</td>
<td>288</td>
</tr>
<tr>
<td>APPENDIX G</td>
<td>290</td>
</tr>
<tr>
<td>BIBLIOGRAPHY</td>
<td>291</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

Figure 1. L2 acquisition of features: Gender, Animate/inanimate, mass/Count ........................................................................................................ 79
Figure 2. Multiple-choice (recognition test) .................................................. 126
Figure 3. Fill-in-the-blank (production test) .................................................... 127
Figure 4. Research Design Overview: Experimental Design ............................. 139
Figure 5. Boxplots of the recognition task across time: Animacy ................. 154
Figure 6: Plot of the recognition task across time by groups (old items):
   Animacy ...................................................................................................... 156
Figure 7: Plot of the recognition task across time by groups (new items):
   Animacy ...................................................................................................... 157
Figure 8: Plot of the production task across time by groups (old items):
   Animacy ...................................................................................................... 160
Figure 9: Plot of the production task across time by groups (new items):
   Animacy ...................................................................................................... 161
Figure 10: Boxplot of the recognition task across time: Count ...................... 163
Figure 11: Plot of the recognition task across time by groups (old items):
   Count .......................................................................................................... 165
Figure 12: Plot of the recognition task across time by groups (new items):
   Count .......................................................................................................... 166
Figure 13: Plot of the production task across time by groups (old items):

Count................................................................................................................. 169

Figure 14: Plot of the production task across time by groups (new items):

Count................................................................................................................. 169
LIST OF TABLES

Table 1. Studies addressing awareness and unawareness ........................................ 40

Table 2. Interpretable and unintepretable features ...................................................... 65

Table 3. Percentage of correct answers in the GJT and fill-in-the-blank by aware vs. unaware participants in the first session ......................................... 105

Table 4. Percentage of correct answers in the GJT and fill-in-the-blank by aware vs. unaware participants in the second session ....................................... 105

Table 5. Level of awareness of those participants classified as aware .................. 106

Table 6. Percentage of new and old items from the fill-in-the-blank endorsed by aware and unaware learners in the first session ................................. 107

Table 7. Percentage of new and old items from the fill-in-the-blank endorsed by aware and unaware learners in the second session ................................. 107

Table 8. Target items according to the distinctions +/- animate and gender .......... 113

Table 9. Target items according to the distinctions +/- count and gender ........ 113

Table 10. Total distribution of the target items by feature: Animate vs. Inanimate and texts .............................................................................................................. 121

Table 11. Total distribution of the target items by gender and feature: Animate vs. Inanimate .............................................................................................................. 121

Table 12. Total distribution of the target items by gender and feature: Count vs. Mass and texts .............................................................................................................. 121
Table 13. Total distributions of the target items by gender and feature: Count vs. Mass .......................................................... 121

Table 14. Distribution of the target items in the paragraphs: Animate and Inanimate ................................................................. 122

Table 15. Distribution of the target items in the paragraphs: Count and Mass 122

Table 16. Distribution of the target items: Animate vs. Inanimate condition in each power point presentation .................................. 135

Table 17. Distribution of the target items Count vs. Mass condition in each power point presentation ........................................ 136

Table 18: The artificial determiner system used according to the disclosed and undisclosed features: Animate vs. inanimate .................. 137

Table 19: The artificial determiner system used according to the disclosed and undisclosed features: Count vs. mass .................. 137

Table 20: Number of participants per experimental group ..................... 152

Table 21: Descriptive statistics for recognition task across time: Animacy ...... 155

Table 22: Repeated measures ANOVA results for the recognition task: Animacy ................................................................. 155

Table 23: Descriptive statistics for production task across time: Animacy ...... 158

Table 24: Repeated measures ANOVA results for the production task: Animacy ................................................................. 159

Table 25: Descriptive statistics for recognition task across time: Count........ 164
Table 26: Repeated measures ANOVA results for the recognition task:

Count ............................................................................................................................................. 164

Table 27: Descriptive statistics for production task across time: Count............ 167

Table 28: Repeated measures ANOVA results for the recognition task:

Count ............................................................................................................................................... 168

Table 29: One-sample $t$ test summary for Unaware group: Animacy ............ 171

Table 30: One-sample $t$ test summary for the Aware group: Animacy ............ 172

Table 31: Descriptive statistics for the Recognition Task by Aware and Unaware participants: Animacy.................................................................................................................. 173

Table 32: Repeated measures ANOVA results for the Unaware participants:

Animacy ........................................................................................................................................ 173

Table 33: Repeated measures ANOVA results for the Aware participants:

Animacy ......................................................................................................................................... 174

Table 34: Descriptive statistics for the Production Task by Aware and Unaware participants: Animacy.................................................................................................................. 176

Table 35: Repeated measures ANOVA results for the Unaware participants:

Animacy ......................................................................................................................................... 176

Table 36: Repeated measures ANOVA results for the Aware participants:

Animacy ......................................................................................................................................... 177

Table 37: One-sample $t$ test summary for Unaware group: Count ............... 179

Table 38: One-sample $t$ test summary for the Aware group: Count ............... 179
Table 39: Descriptive statistics for the Recognition Task by Unaware and Aware participants: Count ................................................................. 180

Table 40: Repeated measures ANOVA results for the Unaware participants:
Count ............................................................................................ 180

Table 41: Repeated measures ANOVA results for the Aware participants:
Count .............................................................................................. 181

Table 42: Descriptive statistics for animate and inanimate responses according to the recognition and production task: Animacy................................. 185

Table 43: Repeated measures ANOVA results for the Recognition task and the Production task: Animacy ................................................................. 185

Table 44: Descriptive statistics for recognition task across time: Count .......... 187

Table 45: Repeated measures ANOVA results for the Recognition and Production task: Count ..................................................................................... 188

Table 46: Contingency table for animacy and type of feature (Aware and unaware participants): Animate ................................................................. 191

Table 47: Contingency table for animacy and type of feature (Unaware participants): Animate ................................................................. 192

Table 48: Contingency table for animacy and type of feature (Aware participants):
Animate ............................................................................................ 192

Table 49: Contingency table for animacy and gender (Aware and unaware participants): Animate ................................................................. 193
Table 50: Contingency table for animacy and gender (Unaware participants):

Animate .................................................................................................................. 193

Table 51: Contingency table for animacy and gender (Aware participants):

Animate .................................................................................................................. 193

Table 52: Recognition task: Contingency table for count and gender (Aware and
unaware participants): Count................................................................................. 194

Table 53: Production task: Contingency table for count and gender (Aware and
Unaware participants): Count................................................................................. 194

Table 54: Contingency table for animacy and gender (Aware and unaware
participants): Animacy ......................................................................................... 196

Table 55: Contingency table for animacy and gender (Unaware participants):

Animacy .................................................................................................................. 196

Table 56: Contingency table for animacy and gender (Aware participants):

Animacy .................................................................................................................. 197

Table 57: Contingency table for animacy and gender (Aware and unaware
participants): Animacy ......................................................................................... 197

Table 58: Contingency table for animacy and gender (Unaware participants):

Animacy .................................................................................................................. 198

Table 59: Contingency table for animacy and gender (Aware participants):

Animacy .................................................................................................................. 198

xx
Table 60: Recognition task: Contingency table for count and gender (Aware and unaware participants): Count........................................................................................................... 199
Table 61: Production task: Contingency table for count and gender: Count..... 200
Table 62: Canonical and Arbitrary comparisons ................................................. 201
Table 63: Descriptive statistics Recognition task Canonical and Arbitrary
  comparisons: Animacy ................................................................................. 203
Table 64: Paired-sample t test results Recognition task Canonical and Arbitrary
  comparisons: Animacy ................................................................................. 203
Table 65: Descriptive statistics Production task Canonical and Arbitrary
  comparisons: Animacy ................................................................................. 204
Table 66: Paired-sample t test results Production task Canonical and Arbitrary
  comparisons: Animacy ................................................................................. 204
Table 67: Descriptive statistics for the Recognition task: Count....................... 206
Table 68: Paired-sample t test results for the Recognition task: Count............. 206
Table 69: Descriptive statistics for the Production task: Count......................... 207
Table 70: Paired-sample t test results for the Production task: Count.............. 207
Table 71: Errors according to the noun endings Animate versus Inanimate
  experimental group .................................................................................... 250
Table 72: Errors according to the noun endings Count versus Mass experimental
  group ........................................................................................................... 250
INTRODUCTION

Statement of the problem

One of the main differences between first language (L1) and second/foreign (L2) acquisition is the fact that the resulting grammars in L2 learning are unreliable and nonconvergent. L2 grammars are unreliable because a lack of success is the norm, and L2 learners end up with grammars that are not similar to those of the target language (nonconvergence) (Bley-Vroman, 2009). Children’s use of an endowed innate language learning mechanism takes place “outside awareness”; whereas L2 language learners, according to DeKeyser (2003) use their analytical abilities to analyze the structure of the L2, and to contrast it with the L1 (p. 334).

In order to acquire an L2, some sort of attention must be paid to the input. The crucial role conceded to attention is similar to the consensus existing in cognitive psychology. In second language acquisition (SLA), the importance of attention is mainly found in those cognitive approaches that state that attention “to input is seen as essential for storage and a necessary precursor to hypothesis formation and testing” (Schmidt, 2001, p. 6).

Attention, therefore, becomes the central point in which internal (e.g., aptitude, motivation, current L2 knowledge, processing abilities etc.) and external
(e.g., the complexity and distribution of input, discoursal, instructional treatments, task characteristics etc.) factors converge (Schmidt, 2001).

The three most comprehensive and elaborated theoretical models of attention in SLA include attention as a crucial aspect for learning to take place (Schmidt, 1990; Tomlin & Villa, 1994, Robinson, 1995b); however, they disagree with regards to the role of awareness in subsequent processing of the L2 data. The number of empirical studies in SLA that focus on attention and awareness as the primary goal of research or as the interaction with other factors has grown. While these studies appear to support the important role awareness plays in input processing (e.g., Leow, 1997, 2000, Rosa & Leow, 2004a, 2004b; Rosa & O’Neill, 1999; Hama & Leow, 2010), whether awareness is crucial for learning to take place is still debatable (Faretta-Stutenberg & Morgan-Short, 2011; Hama & Leow, 2010; Leow, 2000; Leung & Williams, 2011, 2012; Williams, 2004, 2005).

Given its controversial status, the relationship between awareness and learning appears an ideal candidate to be studied under the minimalist syntactic approach. Evidence obtained from second language acquisition is contradictory. Although Interlanguages are considered incomplete and variable, some properties of the target languages appear despite the fact that they are not taught explicitly, and they present no violations of the Universal Grammar (Sorace, 2003).

Under the recent minimalist syntactic theory, L1 and L2 language acquisition can be understood as “a function from set of lexical items to meaning-
sound pairs” (Adger & Smith, 2003). This function responds to a small number of homogeneous syntactic operations in all languages. The variability in languages responds to the specification of lexical items, which are thought to be collections of morpho-phonological, syntactic, and semantic features. The model presupposes that features are of two types: interpretable and uninterpretable. Interpretable features, while relevant to syntactic computation, are also used by the semantic component. Interpretable –semantic– features are [singular], [3rd person], [+past], and [Q(uestion)]. Their counterparts, uninterpretable features –syntactic features–, are not usable by the semantic component, and they are, for example, nominative or accusative or agreement marking on a verb (Pesetsky & Torrego, 2001).

In the case of L1 acquisition, interpretable and uninterpretable features are available for further acquisition and, therefore, language acquisition is the accommodation of features from a universal inventory of available features. For L2, according to some proposals, the difference in interpretability becomes crucial. According to these proposals, interpretable features are predicted to be acquired even if the L2 differs from the native language, while the acquisition of uninterpretable features (syntactic features) might be problematic. This proposal is prompted by the view that L2 processing seems to be more semantic driven due to the L2 learner’s need of extracting meaning from the messages (Liceras, 2003). This processing bias would lead to overpass those semantic features that are
redundant from a semantic-conceptual point of view (Liceras, Zobl, Goodluck, 2009).

Other proposals consider that uninterpretable features are the subject of critical period constraints, and, as a result, not available for the L2 learner (Hawkins & Hattori, 2006; Tsimli & Dimitrakopoulou, 2007). The “Interpretability Hypothesis” states that uninterpretable –not interpretable– features vanish if they are not selected from the inventory of features provided by the Universal Grammar (UG) after a “critical period” when both features are available. These two proposals suggest that uninterpretable features might be either difficult to acquire or not be fully acquired. Interpretable features, on the other hand, are predicted not to be problematic for acquisition.

However, Lardiere (2007, 2009a, 2009b) has challenged this proposal based on data obtained from a extended longitudinal case study that followed a L1 Chinese speaker learning English in an immersion context for more than twenty years. In this regard, Lardiere (2009) affirms that feature acquisition –interpretable and uninterpretable– seems to respond more to detectability peculiarities, assuming that the “basis for detectability is the observation of any formal contrast, such as the difference between student ~ students, or xuesheng ~ xueshengmen” (p. 214). The observation of these contrasts will lead the learner to associate them with some difference in meaning or grammatical function, and hence, construct a certain type of representation for it.
She eludes to express her definition of detectability; however, in the case of feature acquisition, she rejects any intrinsic distinction between acquiring interpretable and uninterpretable features: “My own view is that any feature contrast that is detectable is, in principle, ultimately acquirable (although it might not be actually acquired in any given particular case for independent reasons) (Lardiere, 2009a, p. 214).

According to Lardiere (2000, p. 103), a series of questions emerge: Are language learners able to derive syntactic representations solely through exposure to the relevant morphological forms of the target language? Or, on the contrary, the fact of having already some L1 featural knowledge constrains their hypothesis in order to match up the features with the forms? Is it possible to assemble the right combination of features into the right lexical items for each language especially in those cases in which the same features are present in both languages but they are arranged differently? What happens in those cases in which the features in the L1 do not correspond to those of the L2?

The present study will address the issue of the learning of features using an artificial determiner system that is set up following Williams’ (2004, 2005) artificial syntactic determiner system. The determiner system will test the learning and assembly of a bundle of features while reading a text for meaning. The features under examination are animacy and count (Interpretable features) and their learning or not while they interact with gender in some cases presented as
grammatical gender and others as natural gender. The former corresponds to the arbitrary assignment of gender on inanimate objects, i.e. *la mesa* ‘the table’ (feminine) or *el cuaderno* ‘the notebook’ (masculine). The latter, to the natural assignment of grammatical gender according to biological or natural gender, i.e. *el niño* ‘the boy’ (masculine), *la muchacha* ‘the girl’ (feminine).

The purpose was to test if aware and unaware participants behave differently, and if types of feature [+/– animacy] and [+/–count] play a role when they interact with gender and Spanish. Also, the study addresses whether participants’ assignation of gender is based on [+/– feminine] noun gender, or if type of noun ending determines gender assignment. Type of endings was studied according to two groups: (1) (–o and –a) as canonical markers of gender in Spanish, and (2) (–e and –n) as endings that can be either masculine and feminine. The intention was to address if there is any distinction between the availability of interpretable and uninterpretable features or if detectability plays a greater role in the case of L2 learning. This research conducted under the cognitive and attentional strand measures attention and levels of awareness using online and offline protocols to identify the possible relationship between level of awareness and the learning of the features.

The present study seeks to incorporate what previous online studies have done in regard to a major limitation that existed in the studies within the attentional strand. This limitation consists in the operationalization and
measurement of attention. It was believed that the inclusion of any post measurement was a necessary and sufficient (i.e., diaries, questionnaires, stimulated recalls) condition to establish the presence of attention, or lack thereof. Finer online measurements of attention have proved that all these measurements may suffer from memory decay or simple failure to report thought processes during exposure (Leow, 2000, Leow & Morgan-Short, 2004). Online protocols, on the other hand, have proved the efficacy of measuring attention and diminish memory consequences (Bowles, 2004; Leow, 1997a). Hybrid designs, which combine offline and online elicitation procedures, emerge as a relatively robust tool to address the issue of awareness due to apparent divergences between participants’ responses while processing the input and what they provide offline (Hama & Leow, 2010).

Definition of terms

Attention: Attention in cognition refers to the general term for selectivity in perception. Selectivity involves the focus of an organism on certain aspects of the stimuli in detriment of other aspects (Bundensen & Habekost, 2005). In SLA, according to Robinson (1995b, p. 287-8), attention can refer to “a) the process selecting the information to be processed and stored in memory; b) the capacity for processing information; c) the mental ‘effort’ involved in processing information.” Attention is defined in the present study as the mechanism that
permits the learner to select certain information in L2 input and store it in memory for further processing.

**Awareness:** Although most studies in the cognitive psychology use awareness interchangeably with consciousness, some authors have proposed a distinction between the two. Tulving (1993, p. 290) considers it important to depict a distinction between consciousnesses as “the general capacity that an individual possesses for particular kinds of subjective experience.” According to him, awareness is the particular manifestation of this general capacity at a certain time. Consciousness is never directed at something in particular, whereas the awareness is always of something. Tomlin and Villa’s (1994) definition of awareness in SLA is closely related to the definition of awareness in Tulving (1993). Awareness for them is “the particular state of mind in which an individual has undergone a specific subjective experience of some cognitive external stimulus.”

**Explicit knowledge:** This is knowledge that is conscious and controlled through processing. It is believed that explicit knowledge can be verbalized although it does not imply that these verbalizations need to be metalinguistic in nature (Ellis, N, 1994, and Ellis, R., 2009). Any rule of explanations that elucidates the usage of that linguistic structure or item qualifies as explicit knowledge.

**Explicit learning:** This type of learning is conscious and intentional. It usually comprises memorization of facts and as a result “makes heavy demands on
working memory.” (Ellis, R., 2009, p. 3). Participants are usually aware that they have learnt something.

**Features**: Under recent minimalist syntactic theory, features are at the core of language variation. Travis (2009) states that “any study of language acquisition done within this framework is now the study of the acquisition of features.”

Features are further classified between interpretable –semantic– and uninterpretable –syntactic–. Some theories have proposed that the variability in L2 language acquisition is the direct result of that distinction. Uninterpretable features may pose difficulty to acquire due to a bias in L2 language processing that favors semantic over syntactic analysis (Liceras, 2003).

Other theories that have received attention, like the Interpretability Hypothesis (Tsimpli & Mastropavlou, 2007) consider that uninterpretable, but not interpretable features, are subject to maturational constraints. According to these theories, the former are no longer available while the latter remain available for acquisition at any age.

Lardiere (2009a, 2009b) considers that variability in feature acquisition is due to learners’ potential difficulties with reconfiguring the distribution of formal features on lexical items in the L2 in a way that may be quite different from that of the L1, in addition to differences in conditioning factors that affect their realization (her Feature Reassembly Hypothesis). Therefore, the primary learning problem is the different conditioning factor between the L1 and the L2 and the
reconfiguration of the features on the lexical items of the L2 differently from the way they are presented in the L1 regardless than any distinction between interpretable and uninterpretable features *per se*.

**Implicit knowledge:** This is knowledge that is tacit and intuitive in nature. It cannot be reported orally (i.e., in rules) and it is available through automatic processing, such as fluent performance in verbal behavior (N. Ellis, 1994, and R. Ellis, 2009).

**Implicit learning:** It is the acquisition of knowledge that takes place either without deliberate intention or awareness by a process which occurs naturally, simply, and without conscious operations (N. Ellis, 1994).

**Input:** Input refers in language learning to the language data that the learner receives either orally or written.

**Intake:** Intake is that part of the input that is actually available for further processing, but it does not imply learning. Leow (1993, p. 334) defines intake as “that part of the input that has been attended to by second language learners while processing the input. Intake represents stored linguistic data that may be used for immediate recognition and does not necessarily imply language acquisition.”

**Noticing:** Noticing, according to Schmidt (1994), “is the registration of occurrence of a stimulus event in conscious awareness and subsequent storage in long-term memory). The result of noticing is intake, which can be further processed for learning or not.
Verbal reports: Verbal report refers to the data-elicitation method to collect information about the thought processes involved in completing a task. Protocol analyses—as verbal reports are also known—are classified in three types according to Ericsson & Simon (1983). Level 1 involve the verbalization of covert articulatory or oral encodings; level 2 demands description, or explication of the thought content that is not of the same nature than the language (e.g., information about odors); level 3 requires the subjects to explain their thought processes or thoughts.

Verbal reports are divided also in those that are concurrent or retrospective. The former refers to those reports elicited while participants are engaged in completing the task. The latter are those elicited after completion of the task.

In SLA verbal reports have emerged as one of the principal methods for investigating L2 learners’ thoughts and thought processes, and they have been employed and discussed in much SLA research. Think-aloud protocols—as the preferred method used in SLA—have contributed to providing a rich source of insightful information on L2 learners’ cognitive processes.
CHAPTER 1: THEORETICAL FRAMEWORK

Introduction

The explanatory goal of cognitive approaches to the study of language (either in L1 or L2) seems to be the most promising alternative to explain the complex process of what constitutes knowledge of a language. These approaches try to elucidate what it is that speakers know about a language that enables them to speak and understand it, but, above all, they seek to explain how this knowledge is represented in the mind/brain. If it is true that both L1 and L2 research acquisition ultimate goals are the same: the understanding of this mental representation in the speakers’ mind/brain, a central distinction exists between L1 and L2 approaches. In the former case, it is the norm that almost any person exposed to a particular language under the proper conditions will be able to acquire in a short period of time a language. A fairly well established mental representation of the mother tongue will be derived mostly by positive evidence taken from the input. That is to say, the effect of the input triggers the formation of a particular language without any immense variability among those particular speakers (Strozer, 1994).

On the contrary, a crucial difference between L1 and L2, especially in those cases of L2 adult acquisition, is the variability in the ultimate level of
attainment. Bley-Vroman (2009), in a recent revision of the Fundamental Difference Hypothesis, states that L2 learning differentiates from L1 in two major concepts: it is unreliable and nonconvergent. L2 grammars are unreliable because a lack of success is the norm, and L2 learners end up with L2 grammars that are not similar to those of the target language (nonconvergence). If being exposed to the input even for an extended period of time does not guarantee a fully developed grammar as some immersion studies seem to conclude (Swain, 1985, 1993; Swain and Lapkin, 1995), some means to direct learners’ attention the input might be necessary.

Attention as a vital constituent for second language acquisition

A variety of models mirror the increasing and constant evolution of the role of cognitive and psycholinguistic factors in the process of language acquisition. Different threads of research have investigated diverse aspects of cognition in SLA. Since the function of the input is essential, it is believed by some that the way it is presented affects the processes that undergo in the learner’s mind (Doughty, 2003). Some SLA theories accentuate the importance of attention and awareness more than others. Approaches, such as connectionists, sociocultural as well as strong versions of the innateness position – active even in the case of L2 adult acquisition – diminish the role of the input and the attention to the input as well (Schmidt, 2004).
Cognitive versions of L2 development, on the contrary, see attention to the input essential for storage and a crucial factor for hypothesis formation and testing (Robinson, 1995b; Schmidt, 1990, 1993, 1995, 2001; Tomlin & Villa, 1994; VanPatten, 1994). A mutual premise of these models is that learning does not take place without attention. The role of awareness has gained momentum in SLA development and acquisition as cognitive sciences have evolved in the last years and the methods and instruments to measure awareness are more precise and sophisticated (Cf. Ericson & Simon, 1993 and Leow, 2000 for the usefulness of concurrent verbal reports; Merikle, Smilek, & Eastwood, 2001 for a review of subjective and objective measures as methods to measure awareness; Seth, Dienes, Cleerenans, Overgaard, & Pessoa, 2008 for brain measures).

Attention and awareness are believed to have consequences that impact almost all aspects of SLA, such as the development of the interlanguages, fluency, the function of individual differences (e.g., motivation, aptitude, learning strategies, etc.), and the means in which the role of interaction, negotiation for meaning and the forms and methods of instruction assist language learning (Schmidt, 2004). The debate among SLA researchers over the role of attention, and the presence or lack thereof of awareness reflects the cognitive debate between those that advocate that language is an unconscious process and those that assume a compulsory role for a conscious process in SLA.
Attention, awareness in the cognitive sciences

The disagreement over the function of attention and awareness in learning is not particular of the SLA field. It has been the object of extensive inquiry in psychology and, more recently, in the modern cognitive sciences. James studied attention and awareness in early dates as 1890 (Marchetti, 2001). Years later, behaviorist psychology, which overshadowed mentalist approaches for a period of time, considered that consciousness was an empty concept and could not be investigated because its observation was impossible (Lyons, 1986 cited in Schmidt, 1990).

After more than a hundred years, opinions are as diverse and contrasting as they were at the beginning of the debate. They range from those who have a profound conviction that an important aspect of mental life is embedded in the unconscious (Litman & Reber, 2005) to perspectives that emphasize that the single idea of a complex and abstract mental life is illusory (Shanks & St. John, 1994). This discrepancy can be perceived in state-of-the-art articles about implicit learning. Litman and Reber (2005, p. 447) conclude that “Over the past several decades, it has become increasingly clear that implicit processes, those that operate largely outside of the spotlight of consciousness, play a significant role in most of the interesting things that humans beings do.”

Shanks (2005), on the contrary, has the following concluding remarks:
Research on implicit learning has provided a rich seam of new information concerning the processes that control learning and the relationships between learning, attention and awareness. In my view it has yet to be proved beyond reasonable doubt that there exists a form of learning that precedes both unintentionally and unconsciously. (p. 216)

The challenging issues about the role of attention and awareness can be summarized in three different areas: a) the different results of empirical studies that support or contradict the possibility of implicit learning are in great deal the consequence of vagueness in definition, b) the difficulty of proving the presence or lack thereof of awareness, and c) the difficulty of measuring different levels of awareness.

In regards to the first problem, the terms awareness and consciousness have been used indistinctively, or without a clear statement of the relationship involving attention and awareness. Most researchers do not make a clear cut distinction between awareness and consciousness. In cognitive psychology, Merikle, Smilek, and Eastwood (2001) bind consciousness and awareness: “any evidence that perception is not necessarily accompanied by an awareness of perceiving attracts attention because it challenges the idea that perception implies consciousness.” Tulving (1993), on the other hand, favors a distinction between both terms, in which consciousness is the “general capacity,” whereas awareness
is “a particular manifestation” of that general capacity. In this regard, “awareness
presumes consciousness, but consciousness does not imply awareness:
consciousness is a necessary but not sufficient condition of awareness.”

In regards to the relationship attention and awareness, Schmidt (2001)
proposes that attention entails awareness since both are the flip-side of the same
coin (Schmidt, 1990, 2001). The same can be said for Tulving (1993), who
considers that consciousness is not directed at anything, while awareness is
directed at something; implying a bond between awareness and attention.

The most recent connectionist models try to integrate consciousness into a
model that works in relationship to selective attention, working memory, and
cognitive control. Maia and Cleeremans (2005) based their analysis on the notion
of biased competition which consists in the “competition between representations
that are widely distributed in the brain, with top-down influences”. Although
these connectionist models do not intend to explain consciousness, they have
implications for the understanding of consciousness.

An important distinction is between states of consciousness and the
contents of consciousness. In a clear cut distinction between both states, the
former refers to being awake, asleep, in coma, etc., while the second refers to
being conscious about the scene or state in which one is. If we compared it with
Tulving’s (1983) distinction between consciousness and awareness, the former
refers to the states of consciousness and the latter to awareness.
The model distinguishes between two types of representations. The first of them explains the implicit knowledge we have in the case of, for example, knowing a language. The knowledge drives behavior but it is not accessible to direct inspection. The second representation permits explanation of the conscious representations. Those connections are briefer but more active and take the form of firing patterns. These ‘outputs’ can potentially be conscious.

Global workspace theory is a theory that also helps to explain consciousness. This theory predicts that consciousness depends on access to a “global workspace.” An important characteristic of this model is that it works on biased top-down competition. Therefore, the primary postulation is that the winning coalition of neurons determines conscious experience at a particular time.

An important finding of the global competition model is that it responds to the relation between attention and consciousness. A stimulus enters in the consciousness only if its representation is part of the winning coalition.

A major distinction between Dahaene, Baars and their collaborators and Maia and Cleeremans’ proposal is that for the former the brain consists of specialized, modular processors and a global workspace that connects those processors, whereas for the latter computation is more distributed and interactive at a global scale.

Another proposal based on the workspace framework is Dahaene and Naccache’s (2001). They state that the dilemma of the research of the
consciousness is that the object of study is an introspective phenomenon and not an objectively measurable response. The solution to this is to compare the subjective reports of consciousness with neuronal information. Subjective reports in the form of verbal reports must be considered serious data. Based on three empirical findings in consciousness 1) cognitive processing is possible without consciousness, 2) attention is a prerequisite of consciousness, and 3) consciousness is required for specific mental operations (i.e., the ability to maintain representations in an active state for a durable period of time in the absence of stimulation seems to require consciousness, the ability to combine mental operations or to perform a novel or unusual task requires consciousness, the possibility of reporting the mental state requires consciousness as well), they postulate the “hypothesis of a global neuronal workspace.”

This hypothesis synthesizes the idea of the modularity of the mind with the opposite idea of the non-modularity of the mind. Therefore, besides specialized processors, the mind consists of a distributed neural system or “workspace” with long distance connectivity that can, in case it is necessary, interconnect multiple specialized areas.

Unconscious processing can take place at either low level or high level (i.e., face processing, word reading, and postural control) as long as “they are associated with neural functional pathways established at either by evolution, laid down during development, or automatized learning” (p. 8). Considering that a
significant amount of mental processing can occur unconsciously, the logical exploration is to ask whether there are any benefits associated with the consciousness. Based on Sperling’s (1960) study and on their own studies, they suggest that one of the benefits of consciousness is that consciousness is needed for the ability to maintain representations in an active state for a durable period of time in the absence of stimulation.

The issue of consciousness is intrinsically related with the issue of its measurement. Measuring consciousness entails saying something about conscious levels. Conscious level is the state of any organism, which can be total unconsciousness (e.g., death or coma) to vividly wakefulness. The instrument used to measure consciousness is strongly related to the theories of consciousness. Seth, Dienes, Cleerenans, Overgaard, & Pessoa (2008) present a review of three current approaches to the unconsciousness. Worldly discrimination theory considers that any mental state that expresses its content in behavior is conscious. This theory prefers signal-detection theory (SDT) as its favorite statistical design method to quantify the discrimination of stimuli. A mental state is conscious if it offers enough informative discrimination among an extend repertoire of possible states. High-order theories (HOT) propose that a mental state is conscious when a person is actually aware or being disposed to being in that state.
Empirical studies in regards to (un)awareness in the non-SLA field

Studies about (un)awareness in psychology and cognition field have a long tradition of more than a hundred years (Brown & Besner, 2004). Comparing the vast number of empirical studies in this regard, the following synthesis represents a small percentage of the total; however, it encompasses some of the current interests. Studies can be divided in three groups: a) the difficulties to measure (un)awareness, b) the possibility of perceiving without awareness, and c) whether abstract learning can develop with the absence of awareness.

Valid and effective empirical measurements of awareness are the core of the studies of (un)awareness. The questions whether “how can we tell if a person is aware of being in a mental state? Or how can we tell if a person is aware of knowing?” (Dienes & Seth, 2009) represent the crucial questions for those researchers interested in the studies of consciousness. Measuring awareness of being in a mental state is a key issue of whatever a priori hypothesis about awareness one embraces (Seth, Dienes, Cleerenans, Overgaard, & Pessoa, 2008).

Tunney and Shanks (2003), Gross and Greene (2007) and Dienes and Seth (2009) reflect about the methods and instruments employed to measure (un)awareness. Measurement of awareness lies at the center due to the fact that the method used will be responsible for classifying as (un)aware either the person or the resulting knowledge. In this regard, Tunney and Shanks (2007) consider that a method called opposition logic procedure, which consists in proposing a
dual mechanism (remember-know), to classify conscious and unconscious processes in artificial grammar learning results as being inadequate. In fact, according to the researchers, those two assumed processes can be the result of a single process due to the fact that participants lack of a response criterion for saying “know” and a stricter one for saying “remember.”

Gross and Greene (2007) explore in an analogical inference task whether it is possible without awareness to transfer the knowledge of a certain pattern set to an unrelated set. Participants are exposed to a set of patterns that contains the combinations A, B, C and they have to transfer that knowledge to the novel combinations X, Y, Z. Based on a post-experimental questionnaire to assess awareness, they conclude that “exhibiting a higher degree of abstraction and flexibility is possible without deliberative processes or awareness.” The concluding remarks may be too pompous if we cautiously review the flaws in the tool used to measure awareness. Due to the fact that the questionnaire used measures awareness only after having completed the task, there is a great chance that this measurement suffers from memory decay. Leow (1997, 2000) has raised concerns about the tools used to measure awareness in those studies that support dissociation between awareness and learning. Postexposure questionnaires have internal validity issues given that they are unable to report “what actually learners paid attention to or became aware of during the experimental procedure.” (Leow, 2000, p. 559)
Since perception without awareness seems to be a consensus (Merikle & Joordens, 1997, Brown & Besner, 2003 and Sung & Tang, 2006), new studies have tried to test the limits of this account. Merikle and Joordens (1997) explore whether studies of perception without awareness and studies of perception without attention address a similar concept of conscious awareness. Using three phenomena (e.g., stroop priming, false recognition, and exclusion failure) that have been used to show that stimuli perceived with awareness can lead to different consequences than stimuli perceived without awareness, they conclude that “studies investigating perception without awareness and studies investigating perception without attention may be addressing the same underlying concept of conscious awareness.” (p. 232).

Boyer, Destrebecqz, and Cleeremans (2005) and Smith and Squire (2005) are two studies that investigate the resulting knowledge of being exposed to sequence learning. Boyer, Destrebecqz, and Cleeremans (2005) present a study based on the serial reaction time model, a tool used to address resulting knowledge of sequentially structured stimulus material. Based on the results of two experiments, they conclude that in fact participants do not learn anything in these sorts of tasks, and the results represent the core mechanism involved in sequence learning, which is statistical in nature. Performance in this experiment merely reflects continued reinforcement of knowledge that participants already possess before being exposed to the task (p. 395).
Attention and awareness in L2 acquisition

There is some consensus on the role of attention in L2 acquisition. Different attentional models of SLA include attention as a crucial component for learning to take place (Robinson, 1995b; Schmidt, 1990, 2001; Tomlin & Villa, 1994; VanPatten, 1994, 2004). These presuppositions are supported by empirical data (e.g., Leow, 1997, 2000, Rosa & O’Neill, 1999; Rosa & Leow, 2004). For Schmidt (2001, p. 5), attention is “the mechanism that controls access to awareness.” Attention for him entails awareness since both are the flip sides of the same coin. Although Tomlin and Villa (1994) critique what they consider is a “coarse grained” definition of attention, they acknowledge the critical role of attention as that mechanism that permits “to sort out the massive information.”

However, the role that awareness plays in subsequent L2 processing is a debatable issue. Theoretical models that agree on the role of attention disagree in this particular domain (Robinson, 1995b; Schmidt, 1990, 2001; Tomlin & Villa, 1994). In the case of empirical studies, most of the studies present results that support the thesis that awareness is important for learning to take place and that more awareness results in more learning (Leow, 1997, 2000, 2001; Rosa & Leow, 2004; Hama & Leow, 2010). Only a handful of studies support the idea that learning can take place in the absence of awareness (Leung & Williams, 2011, 2012; Williams, 2004, 2005). As stated by Hama and Leow (2010) a debate inexistent in the SLA field is developing. Awareness has been classified as “a
particular state of mind in which an individual has undergone a specific subjective experience of some cognitive content or external stimulus” (Tomlin & Villa, 1994, p. 193).

Models of attention and awareness in SLA

Schmidt’s noticing hypothesis

Schmidt (1990, and elsewhere) supports the proposal that awareness is necessary for intake and potential learning to take place. Schmidt (1990) draws attention to the terminological vagueness and disagreement that affects the term “consciousness” and proposes a clear distinction of the term in three different categorizations: a) consciousness as awareness (alludes to the different levels of awareness that lead to more or less learning: perception, noticing, and understanding), b) consciousness as intention (related to intentional versus incidental learning), and c) conscious as knowledge (related to the distinction between declarative and procedural knowledge).

The hypothesis (Schmidt, 1990 and elsewhere), modified to incorporate the increasing findings from psychology, cognition, as well as SLA, distinguishes between two levels of awareness: awareness at the level of ‘noticing’ and a higher level, awareness at the level of ‘understanding.’ Intake, according to the hypothesis, takes place at the lowest level of awareness: noticing. Noticing is necessary and sufficient for intake, while a higher level of awareness, awareness at the level of understanding, which encompasses the possibility of understanding
a rule, is not necessary, although it can be beneficial (Schmidt, 1995). He opposes the idea that there can be any substantial learning without awareness, and he is against what is called the ‘magical realist view’, in which unconscious processes do everything. According to Schmidt, “Learning in the sense of establishing new or modified knowledge, memory skills, and routines is therefore largely, and perhaps exclusively, a side effect of attended processing” (Schmidt, 2001, p. 29).

Although there can be some form of learning without awareness (i.e., semantic priming; analogical inference, cf. Gross & Greene, 2007), this seems limited in scope and relevance for SLA. What is important is not if awareness is necessary or not (to measure its total absence is a fallacy), but rather the fact that more awareness results in more learning. Due to the fact that not all aspects of the L2 are similar (i.e., syntax, pragmatics, phonology), attention must be directed to whatever evidence is relevant for that particular language domain; attention must be specifically focused and not just global. However, how learners process the information attended and acquire knowledge of lexical categories, constructions and rules is “a question related to the contrast between implicit and explicit learning, rather to the contrast between attended and unattended input” (Schmidt, 2001, p. 32).

For Schmidt, evidence of awareness is the ability for verbal report during or after exposure to the input. He admits the intricacy of reporting awareness due to the fact that verbal reports do not necessarily reflect the absence of noticing
since individuals may differ in their ability to verbalize awareness. With respect to retroactive measures, recognition tests may be more accurate than recall tests because memory delay can affect to a great extent the recall of forms previously noticed. According to Schmidt, “failure to achieve above-chance performance in forced-choice recognition test is a much better indication that the subjective threshold of perception has not been exceeded and noticing did not take place [compared to failure on recall tests]” (Schmidt, 2001, p. 20).

Schmidt advocates for a relationship between noticing and recognition memory although this measure is also controversial because it may be possible to notice data during processing but failing to do so in a subsequent recognition test.

**Tomlin and Villa’s Fine-grained model of attention**

Tomlin and Villa (1994) view attention as a process dissociated from awareness. The closest construct to noticing is their attentional function detection that is within selective attention, although it does not necessarily involve awareness. They recognize a difference between detection with or without awareness for those cases that “do not implicate awareness or for processing of non-target items (Tomlin & Villa, 1994, p. 200).

Tomlin and Villa consider that SLA research has been influenced by four main conceptions of attention that they considered coarse-grained. Those conceptions are: 1) the view of attention as a limited capacity processor. It is not
possible to process all the input received, and therefore, attention must focus on certain specific aspects. 2) Automatic versus controlled processing, which means that “automatic processes require little or no attention and thus not interfere with other activities,” whereas controlled processes need attention, which is limited, and as a result these processes interfere with the processes that demand attention. 3) The notion of control of information and action that considers that control of linguistic processing depends on the individual’s ability to intentionally focus attention or not on relevant information. 4) The notion of selective attention that states that attention only selects critical information for further processing.

Due to vagueness of definitions previously used, and to the necessity to incorporate those four key ideas for cognitive sciences, they consider that a fine-grained attentional analysis is needed. Based on Posner and Petersen (1990), attention should be examined in terms of three interrelated networks: a) alertness (overall readiness to deal with incoming input, b) orientation (direction of attentional resources to a certain type of stimuli), and c) detection (cognitive registration of the stimuli). Detection is the attentional function at which information is encoded in memory for intake to take place.

A crucial claim of this model is that there is no need for awareness in detection. The conscious apprehension of stimuli (i.e., awareness) is a separate mechanism that is not required for intake to take place. According to their fine-grained model, it is necessary to make a distinction that previous models have
failed to make. This vital distinction is between the cognitive processes involved in learning (i.e., attention as detection), and the subjective experience of what is being attended to during input processing (i.e., awareness). This view is contradictory to Schmidt’s, who believes that there cannot be learning without awareness (i.e., conscious attention).

Tomlin and Villa (1994) proposed a set of criteria to consider the presence of awareness. Attentional studies should adopt Allport’s (1988) criteria to consider someone as being aware. These criteria states that in order to consider someone “aware” individuals must: a) show some behavioral or cognitive change as a result of being aware (e.g., learning a sequence); b) report the experience of being aware; c) describe the subjective experience.

Robinson’s Model (1995b) of attention and memory

Robinson (1995b) offers a model that brings together Schmidt’s and Tomlin and Villa’s contradictory perspectives by defining noticing as “detection plus rehearsal in short-term memory, prior to encoding in long term-memory” (Robinson, 1995b, p. 296). He considers that Tomlin and Villa’s detection plays a role in earlier stages before noticing that last only milliseconds. Consequently, Robinson (1995b, 2003) supports the noticing hypothesis, claiming that no learning can take place without attention. After reviewing different models of short and long-term memory in cognitive psychology, he proposes a model of
attention and memory that is consistent with Schmidt’s noticing hypothesis. According to him, noticing requires rehearsal in working memory prior to encoding in long term memory: “It is possible to briefly notice and permanently or temporarily forget, and to notice and remember over time. More permanent encoding in long-term memory is a consequence of the level of activation of the information, itself the result of rehearsal and elaboration” (Robinson, 1995b, p. 298). Although there is, according to him, some evidence in cognitive psychology of learning following detection, those effects are usually temporary, and they are not enough evidence to discard Schmidt’s claims in regards to the key role of awareness for learning to happen.

Robinson’s model’s main contribution is the inclusion of a detailed analysis of the role of short and long-term memory in attention and awareness. In this model, short-term/working memory is seen as a subset of long-term memory in a currently activated state, and the place where noticing takes place. In addition, rehearsal, short-term memory may depend on the processing demands of the tasks performed.

Following Best (1992), he proposes two different types of processing strategies, both of which involved awareness: data driven (bottom-up processing), which is rehearsal and maintenance in memory of isolated instances; and conceptually driven (top-down processing), which is a rehearsal that distributes instances in abstract configurations. This latter process involves the elaboration of
rehearsal and the activation of schemata (i.e., prior knowledge) or higher-order relations with long term-memory. In the case of search for rules underlying a sequence of stimuli, for example, or to apply previously learned rules to new exemplars.

In sum, the disagreement between both theoretical perspectives is that Schmidt considers that noticing is isomorphic with awareness, and therefore both complement each other. He then distinguishes between two levels of awareness: awareness at the level of noticing (low level), and awareness at the level of understanding (high level). For Schmidt, noticing is necessary for intake while the higher level of awareness might be beneficial but it is not imperative. Tomlin and Villa, on the contrary, view attention from three functions: a) alertness, b) orientation, and, c) detection (the cognitive registration of the stimuli). According to Tomlin and Villa, solely detection is necessary for further processing of input and further learning to take place. Nevertheless, detection does not need or does not imply awareness; learning can be dissociated from awareness.

Studies in SLA conducted within the awareness strand of research There are up to now a series of empirical studies within the attentional framework in SLA (Leow, 1997, 1998a, 1998b, 1999; Rosa & O’Neill, 1999; Rosa & Leow, 2004). They have incorporated definitions of noticing, attention, and awareness. Due to its proven effectiveness, in order to measure and operationalize those notions, they have incorporated verbal reports, mostly concurrent verbal reports
(e.g., think-alouds) as the tool to study learners’ cognitive processes while they interact with the L2 input.

The seminal article in this line of research was Leow (1997). The article, intended to test Schmidt’s noticing hypothesis and based on the results of 28 participants that solved a crossword puzzle, found enough evidence in support of the claim that awareness was a sufficient condition for intake. Through the solving of the crossword puzzle, participants were exposed to the third person singular and plural irregular Spanish preterit form –ir. The task intended that participants produced the incorrect stem vowel of the irregular form first; later the correct irregular stem vowel was provided from other clues. This permitted data to be gathered with regard to learners’ allocation of attention. The experiment measured the performance on two post exposure assessment tasks (recognition and written production) after having completed the crossword puzzle. The conclusions were a) meta-awareness appeared to correlate with an increase usage of hypothesis testing and morphological rule formation, b) an absence of meta-awareness correlated with an absence of such process, and c) learners that demonstrated a higher-level of awareness performed significantly better than those with a lower level in both post assessment tasks. Although awareness proved to have a facilitative role, as advocated by Schmidt (1990, and elsewhere) and Robinson (1995b), the claim for the essential role of awareness remained unresolved due to the nature of the task.
Rosa and O’Neill (1999) investigated how intake is affected by awareness and by two conditions of exposure to the input ( [+/- formal instruction] and [+/- explicit directions to search for rules]). 67 intermediate learners of Spanish took part in the experiment that found statistical effects for both learning conditions and awareness. In the former case, analysis revealed a positive effect of explicit rule explanations on intake of the grammatical forms, which means that formal instruction – regardless of rule search instruction – was significantly better than any of the implicit conditions. This could have been the result of explicit conditions of exposure to the input that “may have contributed to raising learners’ level of awareness” (p. 547). In the case of awareness, it was found that aware participants classified under the level of understanding made significant gains between the pre and the posttest. In this regard, it was concluded that the results supported “claims in the cognitive psychology and SLA fields that associate higher levels of awareness with a structural, conceptually driven way of processing input” (p. 543).

Rosa and Leow (2004) is another study that measured awareness in relation to different degrees of explicitness. In relation to specifically levels of awareness, the results showed that higher levels of awareness correlate with more complex and refined input processing that allowed for extraction of patterns from specific instances of a given L2 structure. Although it was found that participants with a level of awareness at the level of “understanding” had significantly higher
scores than those at the level of “noticing” or with those that failed to produce any report, participants at the level of “noticing” produced a significant increase in scores than those that did not report any level of awareness. Noticing had a significant effect on learners’ performances.

Learning without awareness

In order to address the issue of awareness, especially in comparison with those participants classified as unaware, Leow (2000) replicated Leow (1997) with some diverse methodological features intended to address the issue of awareness or lack thereof in relation to learning. 32 beginning learners of Spanish took part in this experiment that investigated the same target form under a task design that permitted data to be gathered from unaware learners. Although Leow recognized the difficulty to operationalize a complete absence of awareness in foreign language behavior, he solved this fallacy by establishing awareness (operationalized as the report of being aware) as the baseline and compared the performance of participants reporting being aware with those that failed to report that low-level of awareness. The crossword puzzle also underwent some modifications. Instead of providing the clues that would direct learners’ attention to the correct vowel shift, in this study the vowels were retained in the stems, and participants’ attention was diverted from the stems by asking them to fill in the morphological verb endings. After having analyzed the data elicitation measures
of awareness from post exposure questionnaires (offline) and think-alouds (online), the conclusion was that aware participants of the targeted morphological form recognized and produced more of those forms “when compared to with the group that demonstrated a lack of such awareness” (p. 568). Results from ANOVAS showed that there was a significant difference in performance between the aware and the unaware group, between the pretest and the posttest, and a significant interaction between Group and Time. As the statistical results indicated, no dissociation between learning and awareness was found, and therefore, awareness appears to play an important role in language learning.

While awareness appears to play an important role in the learning process recent findings from Williams (2004, 2005) and Leung and Williams (2011, 2012) appear to provide evidence of implicit learning. In the first two studies it is shown that learning without awareness seems to be viable at least at a percentage above chance. In Williams (2004, 2005) an artificial determiner system that encoded animacy and distance as the features to be acquired was tested.

Williams (2004) carried out two experiments that comprised an artificial determiner system. Despite not being able to articulate some sort of rule, 30 out of 37 participants performed above chance. An interesting correlation found was between L1 language that encoded gender and test performance. Experiment 2 tried to reduce the weight of the correlation between L1. None of the participants became aware of the animacy distinction, which was not mentioned. Scores of the
first subset of the assessment section were not significant above chance for the
generalization items (52%). In the second phase of this section, the scores for the
generalization items increased to 60%, significant above chance. After finishing
the testing section, three participants reported that they were aware of the key
distinction, and that they became aware of it during the training section. It is also
impossible to determine whether awareness developed in the testing task only or
whether implicit knowledge from the training became available in the testing
section.

Williams (2005) uses the same determiner system developed in
experiment two but with some additional features that corrected some
inefficiencies detected in experiment 2. In first place, to make sure that
participants’ attention was drawn to the animacy distinction they were asked to
repeat each sentence after its presentation. In those cases in which it was difficult
for some participants to do it, especially the non-native speakers of English,
participants had to repeat carefully the determiner and the noun. They were told
that they were going to be tested for memory. The test session was divided in two
phases: the first one requested participants to base their decisions on what seemed
more familiar or a better option. Immediately after this subsection, they were
asked to verbalize their reasons for making their decisions, which would be
classified as awareness at the level of understanding (Schmidt, 1990). In the
second part of the testing phase, participants were told that the selection of the
determiners responded to a rule. They were asked to find the rule while performing the task. This was designed to investigate whether implicit learning could facilitate explicit learning (Mathews, Buss, Stanley, Blanchard-Fields, Cho, Druhan, 1989). Awareness was measured using post exposure questionnaires.

In the first experiment, eight out of 41 participants reported being aware of the rule when they were asked after the first section of the testing phase. Their score was almost perfect. Of the remaining 33 participants, six reported some sort of rule but with low levels of accuracy. This group of participants actually performed poorly with 49% on generalization items. The remaining 27 participants reported basing their decisions on familiarity or intuition; nevertheless, they performed above chance with a mean score of 63% in the generalization items.

In experiment 2 exactly the same procedure was taken with the exception that each noun appeared with only one determiner. Seven out of 24 participants became aware of the animacy rule. The score of this group was 85%, significantly above chance. The seventeen participants that reported not being aware of any rule in the first part of the testing phase, got scores higher above chance, 60%, on the trained and generalization items. Three out of seventeen participants became aware of the rule after the second section. However, they were grouped as partially aware, since they were aware of the animate rule for only one of the determiners (e.g., near or far). Their performance was lower than chance. In the
case of the remaining eleven participants, their score was above chance although they failed to report any awareness of the rule.

According to the researcher, the experiment proves that form-meaning connections can take place at a level below noticing (Schmidt, 1990, 1994, 2001). Therefore, the findings support Tomlin and Villa’s (1994) assumption that stimuli processed at a level below awareness can contribute to learning. Awareness at the level of noticing is the lowest level of awareness proposed by Schmidt (2001) for learning to take place. Tomlin and Villa (1994), on the contrary, proposed that representations can be activated even without the individual been aware of them: “Detection is the most related to awareness, but there is considerable evidence indicating that the information can be cognitively detected, even though the individual is not aware of its having occurred” (p. 197).

Hama and Leow (2010) extended Williams (2005) by adding some crucial methodological elements intended to investigate the role of unawareness simultaneously while encoding the incoming information. Those adjustments comprised 1) online (treatment and testing phases) and offline (at the end of the experiment) measurements of awareness, or its absence, 2) increase in the number of possible options (from two to four) in the multiple-choice test, and 3) the inclusion of a production test. Using the same artificial determiners put up by Williams (2005), the final results contradicted the original study’s main findings, and participants classified as unaware were unable to generate the correct
determiner significantly above chance. One major difference in both studies that could explain the discrepancy of the results is that the assessment tasks measured the capacity of generating the correct answer according to both the animacy and distance rules. Williams (2005) asked participants to choose a correct answer between two possible options that differed only in one variable, (animacy or distance). In Hama and Leow (2010) study participants needed to choose a correct response among four possible answers that included all the variables (animacy and distance). This increase in the number of possible answers lowered chance as a factor (from 50% to 25%) in the experimental condition, as well as accurately measured production. In order to statistically compare the two studies, Hama and Leow employed a chance correction formula (p. 488).

The hybrid design (previously and effectively used in earlier studies, e.g., Leow, 2000) permitted the measurement of awareness at a lower level, that is, at the level of noticing, while Williams (2005) measured awareness at a higher level, that is, at the level of understanding. Reporting a level of awareness similar to understanding is a clear proof of being aware, while the opposite cannot be claimed.

Ninety-six students participated initially in the study, of which thirty-four were selected in the final pool of the study according to the following conditions: they had learned the distance rule, and were reportedly unaware of the animacy rule. Results from paired sample t-tests revealed that unaware participants were
unable to learn significantly above chance the critical distinction of animacy of
novel words in either selection or production of the new or trained determiner-
noun combinations.

Table 1

Studies addressing awareness and unawareness

<table>
<thead>
<tr>
<th>Study</th>
<th>Part</th>
<th>Tasks and target forms</th>
<th>Method to measure awareness / unawareness</th>
<th>Meas. of awareness</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leow (2000)</td>
<td>32 college students L2 Spanish</td>
<td>Crossword puzzle Third person preterit form (verbs ending –er and –ir) Recognition task Written production task</td>
<td>Online: TA (treatment and post exposure tasks). Offline: 1) Answer two probe questions after completing the crossword puzzle. 2) After probe question after completing the post exposure tasks. 3) Unaware participants interview three weeks aprox. after completing the task</td>
<td>TA (online) Quest. (offline)</td>
<td>Recognition and written production: Significant difference in performance in groups (aware produced more target forms than unaware) Significant interaction between group and time. Aware group vs. unaware group: For both tasks, indicate that 75% of the variance in performances between the two groups is the result of being aware of those forms during exposure. Aware group (pretest vs. posttest): 86%-88% of the variance in the ability to recognize and produce in writing in the posttest is the result of being aware of the irregular forms during exposure. Unaware group (pretest vs. posttest): No difference in means for both tasks.</td>
</tr>
<tr>
<td>Williams (2004)</td>
<td>Exp. 1 37 adult participants</td>
<td>Treatment : Listening presentation of the target items: Post exposure questionnaire (Forced decision task)</td>
<td>Quest. offline (only)</td>
<td>7 participants became aware. 30 unaware participants based their</td>
<td></td>
</tr>
<tr>
<td>Exp. 2</td>
<td>Target structure: A simpler version of the artificial determiner system 12 different nouns: 6 (living things) and 6 (non-living things) The language responded to the distinction between living/non-living and far/near. Computer mediated Visually</td>
<td>Post exposure questionnaire (Forced decision task)</td>
<td>Questionnaire offline (only)</td>
<td>None became aware of the relationship between near/far and animacy. Scores: a) first set of generalization: 52%, b) trained items: 84%, c) second set of generalization items: 60% (above chance p &lt; 0.05). Some little evidence of implicit learning in the case of unaware participants. However, the improvement in scores suggests that some</td>
<td>17 university students</td>
</tr>
</tbody>
</table>
presentation of the combination determiner and noun. Choose the correct answer: between near and far. Produced a sentence 32 training items x 4 and 5 cycles (128 and 160 trials) Testing phase a) two trained items, b) eight generalization items, c) eight trained items, 8) generalization items. 

| Williams (2005) | 41 adult participants L1 various languages L2 various languages | Similar determiner system to Williams (2004) Twice as nouns per class Novel determiner words are: gi, ul, ro, ne Determiner system responded to the distinction animacy (living/non-living) and distance (near/far) Training phase: a) listening a sentence, b) indicate if it meant near or far, c) repeat the sentence, d) create a mental image of it. Testing phase: Written selection test. Choose in between two possible answers that complete | Offline questionnaire After the completion of the first test phase and the second test phase First testing phase: Aware participants: Self-reported aware participants obtained 91.4% on generalization items. Unaware participants: Six reported some sort of erroneous rule, but were included in the unaware participants due to the fact that it was low level of confidence and when all the unaware participants were told the rule they claimed not to have it considered it. This group of participants that reported a rule obtained 49% on generalization items and 68% on trained items. Unaware group plus group with some sort of rule: 60.8% on generalization items Unaware group |
the sentence. Criteria of selection “more familiar, better, or more appropriate”. Test 2 trained items plus followed by 8 generalization items. 8 generalization items plus 8 trained items. Second test phase: 10 trained items, first group of generalization items, and second group of generalization items.

<p>| Hama &amp; Leow (2011) | 34 college students | Same artificial determiner system used in the second experiment in Williams (2005) Training task Same as Williams (2005) Assessment tasks: Similar to Williams (2005) but the options were increased to four instead of two. Listening assessment tasks: MC listening assessment task (select among four possible options) Production assessment task listening (produce the correct determiner) | Hybrid design Online by TA and offline by questionnaire | TA Training and assessment tasks Three levels of awareness: “Understanding,” “noticing” and “no report” | Results for unaware participants MC assessment tasks Unaware participants Main effects for distance in new and old items. Results according to CACD (correct animacy and correct distance) vs IACD (Incorrect animacy and correct distance). Bias in distance. IACD was the prefer response for new items. Production assessment task Main effects for distance in both trained and new items. No main effects for animacy in either new or trained items. CACD vs IACD No significant differences between those categories. A bias in |</p>
<table>
<thead>
<tr>
<th>Experiment 1:</th>
<th>Same experiment as Leung &amp; Williams (2011a).</th>
<th>Offline questionnaire</th>
<th>Classified as aware at the end of the experiment</th>
</tr>
</thead>
<tbody>
<tr>
<td>33 L1 English speakers (undergraduate and graduate students)</td>
<td>Training phase: 272 slides &gt; transition phase (8 slides) &gt; control (32 slides) &gt; violation (32 slides)</td>
<td>Experiment 1: Distinction between near and far (disclosed) and between animate versus inanimate (undisclosed).</td>
<td>Experiment 1: 20 out of 33 participants were classified as unaware; 13 were classified as aware. The difference between the control and violation blocks was significant for the aware and unaware participants.</td>
</tr>
</tbody>
</table>

Experiment 2: 26 L1 English speakers (undergraduate and graduate students) | Experiment 1: Distinction between animate and inanimate (disclosed feature) and size: big and small objects (undisclosed feature). | Experiment 2: 20 of the 26 participants were unaware after the end of the experiment. Of those 20, 9 participants mentioned relative size as a possible correlation at the end of the experiment when they were asked to guess, but only two could guess which articles corresponded to which relative size values. | Unaware participants: The difference in reaction time between the control and violation block was not significant. The control block reaction was slower than the violation block. Aware participants: The violation block reaction time was slower than the control block but the difference was not significant either. |

Leung & Williams (2011b) | Same experiment as Leung & Williams (2011a). | Offline questionnaire | Classified as aware at the end of the experiment |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiment 1:</td>
<td>Reaction time</td>
<td>Experiment 1: 20 out of 33 participants were classified as unaware; 13 were classified as aware. The difference between the control and violation blocks was significant for the aware and unaware participants.</td>
<td></td>
</tr>
</tbody>
</table>

| Experiment 2: | Reaction time | Experiment 2: 20 of the 26 participants were unaware after the end of the experiment. Of those 20, 9 participants mentioned relative size as a possible correlation at the end of the experiment when they were asked to guess, but only two could guess which articles corresponded to which relative size values. | |

Unaware participants: The difference in reaction time between the control and violation block was not significant. The control block reaction was slower than the violation block. Aware participants: The violation block reaction time was slower than the control block but the difference was not significant either.
Faretta-Stutenberg and Morgan-Short (2011) replicated Williams (2005) in order to deeply investigate implicit learning by (1) examining learners with a specific language background, and (2) using a fine grained method to analyze awareness by using think-aloud protocols.
Thirty participants took place in the experiment that followed the similar methodological design without changes and even used the same awareness classification system as Williams (2005). Nevertheless, the study failed to discover above chance performance for the unaware participants on the critical items: the generalization items. Researchers concluded that the discrepancy in results seem to be attributable to individual differences, e.g., linguistic and educational differences. In Faretta-Stutenberg and Morgan-Short (2011) study all participants were L1 English undergraduate students. In Williams’ study there were graduate and undergraduate students with a diverse of L1 backgrounds. While individual differences require further study, it seems that the inconsistency between these two studies can be the consequence of the finer method to measure awareness (e.g., think-aloud protocols) as these results are similar to Hama and Leow (2010) who found no evidence of learning without awareness.

Leung and Williams’ (2011) experiment follows previous Williams’ investigations employing an artificial determiner system. Using a novel reaction time procedure, evidence for implicit learning of form-meaning connections was found. An artificial determiner system that responded to the distinction between age (adult vs. children) and thematic roles (agent vs. patient) was designed. Only the former distinction, and not the latter, was disclosed to the participants. Gi and Ro were used referring to adults, and Ul and Ne, to children. However, the articles Gi and Ul were only used with nouns working as agents and Ro and Ne working
as patients. For example, *Kiss ul Mary a boy on the face* (*ul* child and agent) and *Kiss ne David a girl on the face* (*ne* child and patient). Each experimental trial consisted of a picture and an accompanying audio description. The participants had to choose as quickly as possible whether the named individual appeared on the left or right side of the picture after they had heard a description of the picture they had in front. It was expected that after a serial of trials, if participants learned the connection between articles and thematic roles, then their attention could be immediately directed to the appropriate individual in the picture even before hearing the proper name, thus easing their responses (e.g., on hearing the *Kiss ul*… they would know that the named individual is the agent of the action, while hearing *Kiss ne*… they would know that the named individual is a patient). It was assumed that if the mapping was subsequently reversed (in what it was called the violation block), then reaction times should be slowed and errors should increase. The main assumption was that “if the mapping between articles and thematic roles is learned implicitly, then these effects should be obtained even if participants have no awareness that there is an association between articles and thematic roles or that if it is reversed in the violation block”. (p. 42)

One hundred and twenty pictures clearly depicting a relationship between an agent and a patient were adopted. 88 of them were used for the training session and 32 were used for the testing phase. An important contribution of this study is that there was no distinction between the training and the testing phase from the
participants’ perspective. The participants had to complete the experiment without knowing that the last two blocks (32 trials) were the control and violation blocks, respectively. All the pictures used in the last two trials depicted two adults or two children. Therefore, according to the researchers “the articles could have a predictive value only if participants knew the correct mapping from articles to thematic roles.” (p. 43). In the control block, the sentences respected the same system used in training, while in the violation block (the last 16 trails), the relationship between the articles was reversed. The articles previously used as agents ($Gi$ and $Ul$) were used as patients, and the articles used as patients ($Ro$ and $Ne$) were used as agents.

The experiment was veiled as a reaction time study, and participants had to click to proceed to the next task. For each picture, participants had to complete three tasks: 1) picture description. Participants had to describe the picture in their own words to make sure that they paid attention to the action and the direction of action as presented in the picture. 2) Reaction time test. Participants had to choose left or right as quickly as possible as they listened to an audio description of the picture they had in front, as soon as they could decide on which side of the picture the name individual appeared. Only those responses given before the end of the auditory description were counted; those responses beyond this limit were not recorded. 3) Sentence reformulation. Participants had to put in the correct English order that sentence they just heard.
To measure whether participants were aware or unaware of the way the system worked they were asked at the end of the whole experiment if they had any feelings about when the determiners had been used. They were again presented with some training items and they were invited to make as many guesses as possible.

25 English speakers participated in the study. None of them were advanced learners of any language with a developed case assigned system, i.e., German or Latin. At the end of the experiment, 20 out of the 25 participants (80%) were classified as unaware participants. The remaining five participants were classified as aware since they verbalized some sort of a rule that linked the usage of articles with concepts of active and receptive people. The aware participants were classified as aware regardless of whether they said that they became aware during the experiment or after the experiment. The unaware participants showed a clearly trend to decrease in reaction times during the whole experiment until there was a sudden increase in reaction times at the violation block. A repeated measure ANOVA showed that there was a significant difference between the reaction time of the control block and the violation block, $F(1,19) = 4.64, p < .05$. However, in the case of the aware participants there was not a significant difference between the control and the violation blocks.

Twenty participants, classified as unaware, claimed to have no awareness of the connections between thematic roles and age. Nevertheless, their reaction
times increased significantly compared with the control block when the mapping
between the articles and the thematic roles was reversed in the final block of the
experiment. This was considered the evidence that the participants have learned
implicitly the relationship between thematic roles and age in the usage of
determiners. A demonstration that contextually derived form-meaning
connections might also be implicitly learned.

Leung and Williams (2012) employed the same artificial determiner
system used in Leung and Williams (2011), nonetheless different features
assigned to the determiners. Following a similar reaction time procedure
employed in Leung and Williams (2011), it addresses implicit learning of
grammatical form-meaning connections while at the same time studying whether
some meanings are more amenable to implicit learning than others. In the first
experiment, participants had to learn four determiners. They were told that the
distinction responded only to near and far, but they were not told that that animate
versus inanimate distinction resulted crucial to decipher the meaning of each of
the four determiners used in the experiment. In this regard, gi and ul were
employed when referring to near objects and ro and ne in the case of far objects;
participants were not told that gi and ro were used solely with animate nouns and
ul and ne with inanimate objects.

Participants were exposed to a training phase (272 slides), followed by a
transition phase (8 slides), and two last and crucial phases: a control and a
violation blocks (32 slides each). Only grammatical items were used in the first three phases (training, transition, and control), whereas ungrammatical items were used in the last phase (violation). The Control block was the baseline for comparison with the Violation block that contained ungrammatical items. Since participants had to indicate as quickly as possible which of the two pictures presented was being referred by an article-noun combination, an evidence of learning was whether response times would slow down when ungrammatical items were presented instead of solely grammatical. Response time and accuracy was in this final phase was compared with the immediate previous phase (control) in which the animacy system was respected.

Participants’ awareness was measured using an offline questionnaire that asked participants their feelings in regard to the determiners used: gi versus ro and un versus ne. If they failed to mention any reference to the undisclosed feature, they were classified as unaware.

Thirty-three native speakers of English, enrolled in a university, took part in the experiment. 20 participants remained unaware of the animacy undisclosed feature at the end of the experiment. Of these 20 participants, seven mentioned an animacy correlation when they were reviewing training items but they were not able to mention which articles corresponded to the animacy values. Another participant was able to correctly assign the determiners with the two features: distance and animacy. Those 20 participants were classified as unaware because
they claimed they were not aware of the animacy during the experiment (they were further subdivided according to if they were able to guess animacy at the end).

In the case of the participants classified as unaware it was found that the 68 milliseconds mean reaction time difference between the Control and the Violation blocks was statistically significant $p < .05$. In the case of the participants classified as aware, it was found that the 232-ms increase in the Violation block was significant $p < 0.01$.

The second experiment was almost identical to the first one with the exception that the participants were told that the determiner responded to a noun animacy distinction. Participants were not told that the use of the articles also responded to the relative size of the object denoted to by the associated noun. Participants were told that the *gi* and *ro* were used with animate objects and *ul* an *ne* with inanimate objects. What they were not told is that *gi* and *ul* referred to bigger objects and *ro* and *ne* with smaller objects. The objects in Experiment 2 were simple adopted from Experiment 1 but were adapted to form the new material. The objects were combined such that the size of the two objects on the screen should be unambiguous. Participants would be asked to judge the size of two objects that were completely different in size, e.g., a pair of scissors and a cow, but not between a cow and a bull.
Twenty-six native-speakers of English took part in the experiment. As in Experiment 1, participants were asked to repeat and translate the phrase, for example, “ne scissors, the inanimate scissors,” in which they had to translate the article into animate or inanimate (the disclosed feature). 20 of the 26 participants (77%) mentioned that they were not aware of the relationship between articles and relative size during the experiment or to have noticed anything unusual toward the end. Of those 20 participants, 9 mentioned a relative size correlation as one of their guesses when they were asked, but only 2 of them could guess which articles corresponded to which relative size values.

For those participants classified as unaware there was an increase in time between the mean reaction time in the Control block (2,206 ms.), and that of the Violation block (2,158 ms.). The 48 ms. difference was not significant $p > .1$. For those participants classified as aware, there was a slow-down in the mean time between the Control (1,520 ms.) and the Violation (2,227 ms.) blocks. As in the case of the unaware participants, this difference was not significant $p > .1$ either. Out of those six participants classified as aware, there were three participants that were responsible for the slowdown in reaction times. These three participants were able to correctly report the mapping between articles and relative size values.

According to the researchers, Experiment 1 demonstrated implicit learning of form-meaning connections in regard to animacy, but this effect was not
observed in the case of Experiment 2. Researchers elucidated that two possible interpretations for this are that the people in general – regardless of their native language – are able to detect animate entities with little or no focal attention. Another possible interpretation is that animacy is information that is available for forming associations with grammatical morphemes in the cases of participants’ L1 language; for example, in the case of the English pronoun selection (“he, she” vs. “it”). Animacy is information that “typically interacts with grammatical processes, whereas information about the relative size does not.” (Leung & Williams, 2012, p. 24). They conclude with the affirmation that it is not possible to make definite claims about why some meaning are more amenable to implicit learning until more types or meaning are tested.

In summary, four studies reported some sort of learning without awareness (Williams, 2004, 2005), and Leung and Williams (2011, 2012). Williams fails to measure awareness due to the fact that he only measures awareness at a high level (e.g., understanding). A more refined method might have carried different consequences, as can be seen in Hama and Leow (2010) that extends Williams (2005) with a hybrid measurement of awareness: online and offline. This study, intended mainly to screen carefully aware and unaware participants, reports that none of the unaware participants was able to demonstrate any sort of unconscious knowledge. Leung and Williams (2011; 2012) present an innovative methodology that improves the data collection (e.g. from the participants’ perspective, there
was no division between the training and testing phases but rather just one task consisting of a continuous set of trials); nevertheless, relies only in off-line verbal reports to detect the presence or lack of awareness. Awareness that is able to be verbalized by the participants is the highest level of awareness, a lower level of awareness seems to be oblivious to participants (Leow, 1997, 2000; Hama and Leow, 2010; Faretta-Stutenberg and Morgan-Short, 2011). It is also problematic, at least, to rely on the performance of only five participants despite the fact that their performance appears to be relatively different than the group classified as unaware by an offline verbal report (unaware = 20; aware = 5).

**Measuring attention / awareness in SLA**

As the importance of the roles of attention and awareness in L2 studies has increased, empirical studies have tried to operationalize and measure these constructs inside their methodological designs. The theoretical base of these empirical studies is the important role of attention for processing in L2 development (Robinson, 1995b; Schmidt, 1995; Tomlin & Villa, 1994). Measures of attention and awareness include offline methods, such as diaries, questionnaires, free recall, stimulated recall, immediate reports, and immediate retrospective verbal reports (Egi, 2004). The main limitation of offline verbalizations is that they are subject to potential reconstructive processes as well as memory decay.
Online methods of measuring attention includes note taking, underlining or circling the target form in the text, online uptake charts, and think-aloud protocols. Although think-alouds have limitations, they let the consciousness be observed in much the same way that one can examine external events in the real world, and therefore, one can gain access to processes that otherwise would be unavailable by other means (Leow, 2000).

**Verbal reports**

According to Schmidt (2001) “The clearest evidence that something has exceeded the subjective threshold and been consciously perceived or noticed is a concurrent verbal report, since nothing can be verbally reported other than the concurrent level of awareness” (p. 20). From the beginning of psychology as a science, investigators, needing to access information not available under external observation, have asked subjects about their experiences, thought processes, and strategies (Ericsson & Simon, 1993).

According to Ericsson and Simon (1993), think-alouds can be classified in three levels of verbalizations. Level 1 and 2 Verbalizations involve vocalizations of thought per se. Level 1 is simply the vocalization of covert articulatory or oral encodings. At this level, there are no intermediate processes and as a result the subject does not need to make any especial effort to express his thoughts. The instructions given are crucial since verbalizations will be the result of the
interpretation of the instructions. Level 2 entails description or explication of the thought content that is not of the same nature as the language (e.g. information about odors). Level 3 includes the rational and explanations for participants’ thoughts and actions. This last type of verbal reports has been noted as responsible for inducing alterations in cognitive processes, as it will be reviewed in detail below.

Verbal reports to measure attention, awareness, and noticing in SLA

Concurrent think-alouds have been used predominantly as a fruitful and accurate tool to measure attention and awareness. They were used to establish that participants did notice the target vocabulary (Bowles & Leow, 2005), the target linguistic forms that were enhanced (Leow, 2001), to assess levels of awareness (Leow, 1997; Rosa & Leow, 2004; Rosa & O’Neill, 1999), and to address the issue of unawareness (Hama & Leow, 2010; Leow, 2000).

Think-alouds criticism

Despite the effectiveness of verbal protocols, they have been criticized in psychology and the cognitive sciences as well as in the SLA field because there are some doubts that they can provide a complete and valid account of learners’ cognitive processes. In other words, there are doubts regarding the possible effects of completing a task while concurrently verbalizing the cognitive process.
This is the potential effect of verbalization on the primary task. Two specific areas have been identified: reactivity and latency. The former considers that carrying out two tasks concurrently creates a synergy between both tasks that increases participants’ cognitive load. Therefore, the task results of one participant that carries out the task while thinking-aloud should be different to the results of a different participant that carries out the same task without thinking aloud. Latency, on the other, refers to the increase in time due to the effects of think-aloud.

Ericsson and Simon (1993) support the effectiveness of verbal protocols as a useful research tool that provides researchers rich information that they could not get otherwise. Leow has been extensively using think-aloud protocols as an effective way to obtain concurrent data to operationalize attention and awareness in SLA. In one of his early studies, Leow (1997) showed the effectiveness of think-alouds to efficiently draw the line between aware and unaware learners while completing a task. As a result, think-alouds in SLA have emerged as a robust tool in the awareness and attention strand of research to elicit data on learners’ internal processes (Leow 1997, 2000, 2001; Rosa & O’Neill, 1999).
Further research

It is thorny to extrapolate results from non-SLA to SLA fields; however, the role of awareness is an unsettled issue that has become a much researched topic in the cognitive science and cognitive psychology as well as an in the SLA field. Regardless of the importance given to (un)awareness and its role in learning, the ineffective way to operationalize and measure it has been the main limitation. This has consequences because studies based their findings about the role of (un)awareness in learning depending on their own conception of awareness and the instrument used to measure (un)conscious state of knowledge. In cognitive science, Gross and Greene (2007), and in the SLA field, Williams (2004, 2005) and Leung and Williams (2011, 2012) measure awareness using offline verbal reports intended to measure (un)awareness, when in fact they were measuring awareness at a higher level of understanding. Indeed, offline measurements of attention, much less awareness, have been questioned due to their impossibility to capture the briefness in which attention can occur. To solve this limitation and to increase the effectiveness of measuring attention and awareness, hybrid designs that include online and offline measurements of awareness have been proved to be more effective (Bowles, 2004; Leow, 2000; Hama & Leow, 2010; Rosa & Leow, 2004).

Possible levels of awareness and their different implications in learning should continue to be addressed, especially in studies intended to examine the
prospect of learning second languages in unaware conditions (Hama & Leow, 2010, Leow, 2000; Williams, 2004, 2005;). It is also important to explore participants’ strategies while they complete the different phases of the experiment.

This is essential since some proposals ponder that learning a second language at an adult life is done by using native language knowledge and problem-solving systems. Bley-Vroman’s (1988, 2009) Fundamental Difference Hypothesis proposes that children use Universal Grammar and domain-specific learning procedures, contrary to what adults do. Adults are cognitively mature individuals with a fully developed system: their native language. Adults also have better attention span and superior working memory.

Williams (2004, 2005) and Hama and Leow (2010) studied a syntactic determiner system, and different features need to be investigated. Recent minimalistic syntactic theory presents an opportunity to test the acquisition of (un)aware learning.
CHAPTER 2: UNIVERSAL GRAMMAR AND SECOND LANGUAGE ACQUISITION

Introduction

A major concern in L2 acquisition research is whether or not adult L2 learners can ultimately reach native-like attainment. Some L2 researchers have adopted what is called a deficit hypothesis and their results support the idea that only in rare cases does L2 interlanguage converge with native grammars (Bley-Vroman, 2009). These studies have been contested by more recent studies that proposed that “nativelikeness in late L2 is not typical, but neither is it exceedingly rare” (Birdsong, 2006, p. 20). From the early studies in L2, it has been corroborated that the language of the second language learners is systematic and that errors are not random mistakes. From this finding developed the concept of “interlanguage,” a construct that explains that L2 learners have an internalized mental grammar which is a natural language that follows its rules and principles (Selinker, 1972; White, 2003). L2 researchers under the generative approach consider that interlanguage grammars, like native grammars, entail unconscious mental representations constrained by the UG and influenced in different degrees by the L1 (White, 1999, 2003).

Generativist linguistic theory seeks to provide a description of the linguistic competence of the native speaker of a language and to explain how it is
possible to achieve that competence. UG is an explanation of the way L1 learners come to know (at least implicitly) properties of the grammar that go far beyond the input, how they can distinguish which things are or are not possible (Herschensohn, 1999). The central claim of this theory is motivated by what is called the **logical problem of language acquisition or poverty of stimulus** (White, 2003). The innate predisposition explains the asymmetry between the input received (insufficient and imperfect) and the fairly well established mental representation of the mother tongue derived mostly from positive evidence taken from the input; negative evidence, therefore, is considered unnecessary.

Due to the fact that generativist grammar offers a model of the linguistic competence of the L1 speaker, the assumption of L2 researchers working under the same theoretical framework is identical. Although there are differences in L1 and L2 acquisition at the level of final attainment, there are also similarities that justify the study of L2 acquisition under a generativist approach. White (1989) points out the same **poverty of stimulus** in the case of L2 acquisition: *undertermination, degeneracy, and lack of negative evidence*.

**Background**

The conception of a limited set of features from which a learner selects is closely related to the Principles and Parameters framework.
Parameters with an associated cluster of properties seemed to have explanatory adequacy. Due to the brief span of time in which the process develops, the theory proposed that the number of parameters should be ideally small and that each parameter should be related to a cluster of syntactic properties (Hornstein, Nunes, & Grohmann, 2005; Radford, 2010).

For example, the pro-drop parameter explained not only the appearance of Null subjects, but a series of three seemingly unrelated properties that were triggered under the instantiation of null subjects in the input. These were allowing subject verb inversion (as in 1), null subjects in those cases in which [-pro-drop] need an overt syntactic expletive (as in 2), and that-trace effect sequences (as in 3) (Ayoun, 2003; Wakabayashi, 2002).

(1) a. Peter arrived / *Arrived Peter.
   b. Pedro llegó / Llegó Pedro.
(2) a. There / *Ø was a book.
   b. Había un libro.
(3) a. Who did you say /[* that opened the door.
   b. ¿Quién dijiste [ que ] abrió la puerta

In general, despite the enormous work devoted to characterizing a parameter there has been little consensus about what constitutes a parameter (Lardiere, 2009a). What should be a finite number of elements associated with a cluster of consequences that constrains and helps the learnability of the languages has instead resulted in an increasing fragmentation. Under the current minimalist
approach, variation on languages is reducible to properties of functional
categories that force syntactic derivation (Radford, 2009).

**Theoretical introduction: Current Minimalist Theory and Universal Grammar**

Under the recent minimalist syntactic theory, L1 and L2 language
acquisition can be understood as “a function from set of lexical items to meaning-
sound pairs” (Adger, 2003). This function responds to a small number of
homogeneous syntactic operations in all languages. The variability in languages is
the result of the specification of lexical items, which are thought to be collections
of morpho-phonological, syntactic, and semantic features. Minimalist proposals
draw a distinction between interpretable and uninterpretable features.
Interpretability is related to the intrinsic nature of features and bears on whether a
feature makes a semantic contribution to interpretation (Pesetsky & Torrego,
2001). Interpretable features interface with the semantic-conceptual system of the
mind. In the derivation of a sentence, those features are checked but not
eliminated, while uninterpretable features making no semantic contribution need
to be both checked and eliminated. By erasing uninterpretable features in the
process of the derivation, they become invisible to the LF system, whereas they
remain visible to the PF system.
Table 2

<table>
<thead>
<tr>
<th>Type of constituent</th>
<th>Interpretable features include</th>
<th>Uninterpretable features include</th>
</tr>
</thead>
<tbody>
<tr>
<td>T- constituent</td>
<td>tense, aspect, mood</td>
<td>person and number</td>
</tr>
<tr>
<td>Noun expression or</td>
<td>person, number and gender</td>
<td>case</td>
</tr>
<tr>
<td>Pronoun</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Interpretable and uninterpretable features can be exemplified in the case of English with the finite T constituents in the examples below (Radford, 2009).

(1) (a) She *is* working/ She *was* working

(b) He *is* writing/ He *has* written the assignment

(c) He insists that she *is* respected / He insists that she *be* respected

(d) They *are* working.

In the case of (1a) the interpretable feature Tense defines whether the sentence is in present or past. In (1b) interpretable aspect determines whether the sentence refers to an action that is taking place or one that is a completed action. Interpretable mood in (1c) defines whether a sentence is describing a *realis* or *irrealis* state of affairs. Interpretable tense features in (1d) define that the sentence is in present but the phi-features (person and number) are uninterpretable and they mark agreement with a particular nominal.

Another occurrence of uninterpretable feature is case, which in English can be instantiated in nominative, accusative or genitive (Radford, 2009).
Although it affects nouns and pronouns, it is clearly distinguishable in its realization only in the latter.

(2) (a) It seems [they] were arrested

(b) He expected [them] to be arrested

(c) He was shocked [at their being arrested]

In the case of L1 acquisition, interpretable and uninterpretable features are available and, therefore, language acquisition is the accommodation of features from a universal inventory of available features. For L2, the difference in interpretability becomes crucial. According to some proposals, interpretable features are predicted to be acquired even if the L2 differs from the native language, while the acquisition of uninterpretable features (syntactic features) might be problematic. One proposal, The Interpretability Hypothesis (Franceschina, 2001; Hawkins, 2001; Tsimili and Dimitrakopoulou, 2007; Tsimili & Mastropavlo, 2008) claims that uninterpretable features are subject to a critical period, which leads to their unavailability (Tsimili & Dimitrakopoulou, 2007).

Another proposal, Full Transfer Full Access hypothesis (FTFA) (White, 2003) considers that both interpretable and uninterpretable features are available for L2 acquisition, and L2 speakers can have the same underlying grammatical representations as native speakers.
Empirical studies under the UG strand of research have been centered on presenting data that supports one or the other perspective (in favor of the Interpretability Hypothesis, Franceschina, 2001; Hawkins and Sattori, 2006, Tsimpili, 2003; Tsimpili and Dimitrakopoulou, 2007; Tsimpili & Mastropavlou, 2008; and in favor of the FTFA, Lardiere, 1998, 2007; Schwartz and Sprouse, 1996; White, 2003).

However, some researchers have mentioned the necessity to incorporate novel perspectives to explain L2 variability. Lardiere (2009a), for example, affirms that feature acquisition – interpretable and uninterpretable – depends in part on detecting feature contrasts in meaningful contexts. Assuming that the “basis for detectability is the observation of any formal contrast, such as the difference between student ~ students, or xuesheng ~ xueshengmen” (p. 214), the observation of these contrasts will lead the learner to associate them with some difference in meaning or grammatical function, and hence, construct a certain type of representation for it.

She rejects any intrinsic distinction between acquiring interpretable and uninterpretable features: “My own view is that any feature contrast that is detectable is, in principle, ultimately acquirable (although it might not be actually acquired in any given particular case for independent reasons) (Lardiere, 2009a, p. 214).
According to Lardiere (2009a), a series of questions emerge: Are language learners able to derive syntactic representations solely through exposure to the relevant morphological forms of the target language? Or, on the contrary, the fact of having already some L1 feature knowledge constrains their hypotheses in order to match up the features with the forms? Is it possible to assemble the right combination of features into the right lexical items for each language especially in those cases in which the same features are present in both languages but they are arranged differently? What happens in those cases in which the features in the L1 do not correspond to those of the L2?

Given the contradictions in the predictions at the theoretical level (Interpretability Hypothesis vs. FTFA), as well as dissimilar results at the empirical level (Franceschina, 2001; Hawkins, 2001; White et al, 2004; Lardiere, 1998, 2007, among others), L2 feature acquisition appears as an ideal candidate to be studied under the attentional and awareness strand of research.

The roles of attention and awareness in L2 acquisition have grown as theoretical and empirical studies have proved its importance (Leow, 2001; Robinson, 2003). Attention, therefore, becomes the central point in which internal (e.g., aptitude, motivation, current L2 knowledge, processing abilities etc.) and external (e.g., the complexity and distribution of input, discoursal, instructional treatments, task characteristics, etc.) factors converge (Schmidt, 2001).
The three most comprehensive and elaborated theoretical models of attention in SLA include attention as a crucial aspect for learning to take place (Schmidt, 1990; Tomlin & Villa, 1994, Robinson, 1995b); however, they disagree with regards to the role of awareness in subsequent processing of the L2 data. The number of empirical studies in SLA that focuses on attention and awareness as the primary goal of research or as the interaction with other factors has grown. While these studies appear to support the important role awareness plays in input processing (e.g., Leow, 1997, 2000, Rosa & Leow, 2004; Rosa & O’Neill, 1999), whether awareness is crucial for learning to take place is still debatable (Hama & Leow, 2010; Leow, 2000; Williams, 2004, 2005).

The benefits of incorporating the attentional strand of research to L2 studies under the UG strand are twofold: on the one hand, it is possible to test the attentional studies using a completely different formal framework to address how salient and non-salient cues are detected and analyzed in learners’ cognitive systems, while at the same time, it is possible to address learners’ attention and (un)awareness and not use posttest scores as the only indicator of attention paid, under the false assumption that improved scores were mainly due to attention directed to the targeted forms.

The present study will address the issue of the acquisition of the features [+/- animate], [+/- count], and [+/- feminine] gender in Spanish by using an artificial determiner system that is set up following Williams’ (2004, 2005)
artificial syntactic determiner system. The determiner system will test the acquisition and assemble of a bundle of features while students read a text for meaning. The features under examination are animacy and count (interpretable features) and their acquisition or lack thereof.

The use of an artificial determiner system permits control for previous exposure to the target structure and modifies the bundle of features while at the same time diminishes the prominent role of L1. The issue of learnability of features is not solely explained as a selection of individual features, but, in various cases, as a re-assembly of feature clusters (Lardiere 2008, 2009a; Montrul & Yoon, 2009).

The purpose will be to test if there is any distinction between the availability of the features [+/- animate] and [+/- count] when they interact with gender to test awareness’ role in L2 acquisition. This research conducted under the cognitive and attentional strand measures attention and levels of awareness using online and offline protocols to identify the possible relationship between levels of (un)awareness and the acquisition of the targeted features.

In the particular case of L2, interlanguage variability also emerges for three main reasons: 1) the inventory of features selected in the L1 and the L2 can be different. That is to say, the L2 learner approaches the acquisition of a L2 with a set of features already established in the L1, which may possibly not be the same between the L1 and the L2. 2) The bundle of features can be dissimilar
between languages. Since in many occasions the cluster of features in the L1 and the L2 are dissimilar, L2 acquisition involves figuring out the correct assemblage of features in the L2; process referred as feature assembly and re-assembly (Lardiere, 2008, 2009a; Montrul and Yoon, 2009. Cf. Theoretical background: Determiner features below section). 3) A particular feature can be present in both the L1 and the L2, however, how that feature is encoded in both languages – L1 and L2 – can be different.

In this last case, for example, gender in nouns is marked by languages such as Spanish or French. All of them mark nouns for gender, and have gender agreement with other nominal elements of the DP, such as adjectives or articles (4-5). English has natural gender, which shows up typically in the selection of few pronouns, e.g., he vs. she in the singular, but not in the plural; contrary to what happens in Spanish in singular: él ‘he’ vs. ella ‘she, but also in singular ellos ‘they masculine’ vs. ellas ‘they feminine.’ English lacks noun classes based on grammatical gender; therefore, determiners and adjectives do not show gender agreement with head noun within the determiner phrase (DP).

\[
\begin{align*}
4) & \text{ la } \text{ casa } \text{ blanca} \\
& \text{ the-FS house-FS white-FS} \\
& \text{ ‘the white house’}
\end{align*}
\]

\[
\begin{align*}
5) & \text{ la } \text{ maison } \text{ blanche} \\
& \text{ the-FS house-FS white-FS} \\
& \text{ ‘the white house’}
\end{align*}
\]
Still, those languages that encode gender in nouns do it in different ways: while Spanish and French have two noun genders (masculine and feminine), Dutch, Russian, and German have three noun genders (masculine, feminine, and neuter). Therefore, gender acquisition does not result unproblematic for those L1 learners of gender-marking languages learning another gender-marking language. In this case, learners need to learn how gender is represented in the target language (Montrul and Yoon, 2009). As mentioned above, some languages mark two genders, while others mark three or more; furthermore, words that are feminine in one language can be masculine in the other (e.g. the masculine word “the voyage” in Spanish (‘el viaje’) and French (‘le voyage’) is feminine in Portuguese (‘a viagem’); the feminine word “milk” in Spanish (‘la leche’) becomes masculine in French (‘le lait’) and Portuguese (‘o leite’).

L1 and L2 feature acquisition

The acquisition of interpretable/uninterpretable features is a fairly recent line of research in both L1 and L2; still, in the case of L2 the feasibility of

---

1 In the evolution of Latin to Spanish, gender in nouns was simplified from a three gender system (feminine, masculine, and neuter) to a two-gender system (masculine and feminine) (Penny, 2004). Still, Spanish has a neuter article ‘lo’, which is a remnant of the Latin three-gender system. Since there are no neuter nouns in Spanish, we will not find the sequence lo + noun, except when an adjective is nominalized (i.e. lo bonito ‘the beautiful (things)’) . ‘Lo’ can be followed by adjectives (‘lo bueno y lo malo’ the good and the bad), certain pronouns (‘lo mío’, ‘lo otro’ the other), certain clauses (‘lo que tú digas’ whatever you say) (Alvar, 2000).
acquiring one type or the other carries wider implications beyond the acquisition of lexical or formal categories. In L1, the acquisition of both type of features – uninterpretable and interpretable features –, being a core part of the system, are available and are acquired from the early stages of acquisition (Liceras, Zobl, & Goodluck, 2008). For L2, the distinction between both kinds of features turns out to be critical.

Theoretical as well as empirical evidence differ as to what causes such variability in feature acquisition. An issue that has been extensively debated and that is a fundamental part of the L2 generative research is linked to the issues of the extent to which the first language (L1) influences and shapes L2 initial-state knowledge and development, and whether UG is available (if so, to what extent) in second language acquisition.

Two theoretical proposals at the opposite end try to articulate a response to this. The Interpretability Hypothesis (Tsimpili and Dimitrakopoulou, 2007; Tsimpili & Mastropavlou, 2008), a minimalist account of variability in L2 grammars, proposes that, due to maturational constraints, it is unfeasible to properly represent and analyze uninterpretable features that have not been selected by the L1. Uninterpretable features not selected from the inventory of features given by the Universal Grammar (UG) in the L1 acquisition process are not accessible in L2 acquisition, leading to representation deficits in L2 interlanguage grammars.
The other position, the Full Transfer Full Access hypothesis (FTFA, Schwartz and Sprouse, 1996; White, 2003) considers that new parameters can be reset to settings of the target language or new features (interpretable or uninterpretable) can be acquired regardless of being present or not in the L1 (whether or not there is a critical period), and that L2 speakers have the same underlying grammatical representations as native speakers. What are vulnerable to maturational constraints are the more peripheral components of the language faculty, such as the morphological module (Cf. the Missing Surface Inflection Hypothesis (MSIH, Haznedar & Schwartz, 1997; Lardiere, 2000, 2007; Prevost and White, 2000). 

In order to illustrate the predictions of each theoretical model, one empirical study that supports each model will be reviewed: Hawkins and Hattori (2006), in favor of the Interpretability Hypothesis; Lardiere (1998, 2007), in favor of the FTFA.

Hawkins and Hattori (2006) conducted an empirical study that tested the theoretical assumptions of the Interpretability Hypothesis. As it was mentioned before, this model proposes that uninterpretable features are subject to a critical period. As a result, those uninterpretable features not selected in the L1 are no longer accessible by adult language learners. All other aspects of the grammar are assumed to remain available: the computational devices, their associated
operating principles, all interpretable features, and (only) those uninterpretable features that have been selected in the L1.

Hawkins and Hattori (2006) study the acquisition of the uninterpretable feature that forces wh-movement in English biclausal interrogative sentences. The performance of 19 highly proficient Japanese speakers of English (JSE) was compared to that of 11 Native speakers of English (NSE). JSE and NSE that participated in the study were selected according to a strict criterion that required them to pass a syntax exam that excluded 16 JSE and seven NSE. The final pool of participants had to answer a 38-item test that consisted of a small story and a question related to the text. The question was followed by three answers. Only one of those was both syntactically and pragmatically possible. One was pragmatically but not syntactically possible, and the other was not possible at all. Participants were asked to select all possible answers. The test was designed to investigate the sensitivity of high proficiency JSE to the ‘Attract Closest’ principle. According to the researchers, this must apply if the interlanguage grammars of the JSE have acquired the uninterpretable feature of English that drives movement of a wh word/phrase to interrogative C (i.e. \[uwh^*:\]), leading them to reject what are called ‘superiority’ and ‘subjacency’ violations. A superiority violation is illustrated in (9b), and a subjacency violation in (9c).

9) a. When did Sophie’s brother warn [Sophie would phone who _when_]?
b. * Who did Sophie’s brother warn [Sophie would phone _who_ when]?

c. * When did Sophie’s brother warn [who Sophie would phone _who_ _when_]?

Statistical results showed that there was a significant difference between the mean scores of the JSE and the NSE, in question type (c) violation of superiority, question type (d) violation of subjacency, and question type (e), violation of superiority and subjacency. According to the researchers, highly proficient Japanese learners of English (JSE) are significantly less sensitive than native English (NSE) controls to superiority or subjacency effects associated with constraints on *wh*-movement in multiple *wh*-questions. The results suggested that JSE were missing an uninterpretable *wh*-feature (or EPP feature) that forces *wh*-movement, which is present in English but not in Japanese.

At the other side is the FTFA hypothesis that argues against any syntactic deficit in L2 grammars, assuming a full access to the UG in SLA (whether or not there is critical period). The hypothesis known as Missing Surface Inflection Hypothesis (MSIH) assumes the same theoretical assumptions of the FTFA in regards to full access to the UG, but accounts for the variation in the instantiations of morphological forms. The MSIH suggests that variable production can be attributed to difficulties learners experience in mapping abstract syntax to the
morphological forms due to processing pressure (Prévost & White, 2000), or to problems in pronouncing certain phonological combinations (Lardiere, 1998a, 1998b). According to this hypothesis, interpretable as well as uninterpretable features can be acquired regardless of whether or not they are part of the repertoire of L1 features. Interpretable features (morphosyntactic features) such as [+/- past] or [+/- definiteness]) can be acquired, along with uninterpretable features such as case.

Empirical evidence comes from a longitudinal study carried out by Lardiere (1998, 2007). This study describes the interlanguage of a Chinese speaker, Patty, living in the United States, whose English idiolect has achieved a steady-state. Patty was recorded in three sessions using a naturalistic data collection. The first recording was conducted when she was 32 years old and had lived for ten years in the country in 1986. The second and the third recordings (conducted two months apart) were conducted when she was 41 years old and had lived for 18 years fully immersed in the target language. There was a fourth recording in 2002 besides a series of elicited written samples in the form of e-mails. The results demonstrated that features considered uninterpretable, and therefore impossible to acquire (according to the Interpretability Hypothesis), such as case marking, were properly acquired by Patty, a L2 English learner, while plural marking on nouns, considered to be interpretable, and therefore unproblematic, are less consistently provided. Lardiere (1998) demonstrates that
Patty is perfectly accurate on nominal and accusative pronominal case assignment (100%) in both cases, regardless that case is not overtly realized in Patty’s L1. At the same time, past-tense marking, assumed to be interpretable, is less consistently presented (34%), contrary to the uninterpretable feature case. These results seem to demonstrate that there is no absence of abstract syntactic features, regardless or not being part of the repertoire of L1 available features, but, instead, a mapping problem.

**DP internal agreement in Spanish**

The acquisition of DP’s, especially in the case of English-speaking learners of Spanish, has been well documented. Although both languages have determiners and a functional category DP, English does not have gender and number agreement between determiners (Det), nouns (N), and adjectives (Adj.). A crucial difference between DP’s in both languages is the strength of Num feature: It is weak in the case of English, whereas in the case of Spanish, it is strong. DP strong nominal features in Agr/Num forces overt raising of Nouns, as it is the case in Spanish but not in English. This variation in the strong/weak feature in both languages causes the difference in Spanish between positions of the adjective, with respect to the head, N, pronominal in the case of English (15) and postnominal in the case of Spanish (16).

15) The **black** cat
16) *El gato negro*
Following Carstens (2000), it is assumed that concord within the DP proceeds as follows. As the N rises, it enters into a relationship (head/ head or specifier/head) with the Adj and the Det for its interpretable features to check the corresponding uninterpretable features (17).

Figure 1.

*L2 acquisition of features: Gender, Animate/ inanimate, mass/ count*

![Diagram showing the structure of a DP with features for gender, animacy, and number]

**Gender feature acquisition**

Due to the distinctions in gender agreement within the DP, it is hypothesized that the inability to fully acquire control agreement is due to the impossibility to learn uninterpretable feature properties of the L2 not present in the L1. This current position has received the name of Interpretability Hypothesis (mentioned above). Hawkins (2001) conducted a study of English speakers learning French that appeared to support the Failed Functional Features
Hypothesis assumptions (which the current version is the Interpretability Hypothesis). Bruhn de Garavito and White (2000) (study 1) investigated the acquisition of Spanish DP’s by French-speaking learners to test Beck’s Local Impairment Hypothesis, which predicted variability/impaired representation despite of whether feature values were similar or different between the L1 and L2. The results were compared with those of Hawkins (1998) (study 2) in which he studied English-speakers learning French. Bruhn de Garavito and White (2000) studied two languages that share the availability of the gender feature: French and Spanish mark gender; Hawkins (1998) intention, on the other hand, was to test gender acquisition in the case of one language that marks gender, Spanish, against another that does not mark gender, English. The idea in comparing these two studies and Hawkins (1998) that employ languages that share the same gender features in the L1 and the L2 (French and Spanish), i.e., Bruhn de Garavito and White (2000) with two languages that diverge in the gender feature (English and Spanish), i.e., Hawkins (1998) is to test predictions that assume that L2 errors are attributable to L1. The intention is to test uninterpretable gender features since they are present in some languages but not in others. Both studies are presented in comparison.

In the particular case of gender, it was observed in both studies that the accuracy in gender agreements was high, and that there were more errors with the usage of indefinite than with definite NPs (2% and 7%, study 1 and 2,
respectively). Gender errors were more common in the case of feminine nouns; that is, masculine articles were used more often in the case of feminine nouns, which show that masculine was the default gender. The main point is that both studies had similar findings which could not be due to the lack of gender in the L1, since Bruhn de Garavito and White’s study included languages that have gender both in the L1 and in the L2. An interesting finding is that learners were more accurate in grammatical gender – uninterpretable gender – (14.5% and 9% study 1 and 2, respectively) than in natural gender (i.e., inherent) – interpretable gender – (23% and 14% study 1 and 2, respectively). One explanation, according to Bruhn de Garavito and White (2000) for these results may be that learners pay more attention to morphosyntactic cues, rather to non-linguistic cues assigning gender.

Rates of errors with determiners were lower than with other modifiers, i.e. adjectives. In the case of determiners, the masculine determiner (el) was the less marked option for most of the learners due to the fact that was overused with feminine nouns; although there were some minor cases in which the learners adopted the feminine determiner (la) as the default.

The results of the studies were interpreted by Bruhn de Garavito and White (2000) as evidence for the acquisition of strong gender features after a critical period, against the Interpretability Hypothesis that argues against the probability to happen so, which leads to representation deficits. Considering that
the population targeted is intermediate, the authors propose a morphological explanation for the errors observed. The researchers asserted that L2 learners had acquired gender as a feature, but it was the actual mapping of features to morphology where the problems resided. According to this account, gender errors are a production issue and not a representational problem at the level of formal features. It seems that learners use the default gender, masculine, when they are uncertain about the gender of the noun, and this explains why there is more masculine agreement in the case of feminine nouns than the opposite case. Similar results were found in the L2 French gender agreement study conducted by Renaud (2010). Participants (2nd, and 4th semester, advanced, and native speakers) had to rate the acceptance of sentences that appeared segmented on a computer. The acceptance rates as well as the reading times were analyzed. In the case of the interpretable gender as well as adjectives in superlative constructions, participants inserted a masculine pronoun in feminine contexts rather than feminine pronouns in masculine contexts.

Franceschina (2001), based on a longitudinal study conducted with Martin, a native English speaker learning Spanish in his ultimate level of attainment, disagrees with the morphological explanation. Martin, who started learning Spanish at age 17, has lived 24 years immersed in the Spanish language. Consequently, he has reached the end state of acquisition; nevertheless, his
performance in gender agreement compared with that of a native speaker, who participated in the study, does not reach 100% accuracy.

After analyzing 94 minutes of free speech, results show that he is perfectly accurate (100%) on noun endings – the (interpretable) ones that trigger agreement –, but he is less accurate on adjectives, articles, pronouns, and demonstratives – the targets for (uninterpretable) agreement –. Martin also produced more errors of overgeneralization of masculine determiners with feminine nouns (83.3%), rather than masculine nouns with feminine determiners (16.7%); like previous studies (Bruhn de Garavito and White, 2000; Franceschina, 2001; Hawkins, 1998).

Franceschina gives an explanation to the occurrence that nouns do not present any problems although they are marked for gender. An important distinction between nouns, on the one hand, and determiners, adjectives, pronouns and demonstratives, on the other, is that the former group is triggers of gender, whereas the latter group is targets. Triggers carry interpretable gender features and targets have uninterpretable gender features, which needs to be valued and deleted by the corresponding trigger.

Based on the results, Franceschina argues that the greater percentage of errors in gender (93%) compared to the small percentage in number agreement (7%) can be explained by a distinction between interpretable and uninterpretable features available in the L1 compared to those available in the L2. Martin produces more errors in gender agreement than in number agreement, still the
gender agreement rate was high (85-92%). Martin produces a small percentage of errors in number agreement due to interpretable and uninterpretable number features being available in Martin’s L1 (this book vs. these books). She concludes – but her conclusion is not unproblematic – that the greater number of errors in gender agreement could be an indication that uninterpretable gender L2 features are either not corrected or fully specified or even altogether absent from the subject’s L2 syntactic representations. However, it should be mentioned that considering the high percentage of correct gender agreement (85-92%) in her data, which is significantly above chance, it could also concluded that the uninterpretable features were acquired.

White, et al. (2004) is a study that addresses the acquisition of Spanish by speakers of L1 language that also has gender (French), as well as an L1 that does not (English). Number, as a feature present in all three languages, is also investigated. Subjects were adults, at three levels of proficiency and a control group of native speakers. The intention of the study was to test both theories as the possibility of parameter resetting in nonnative acquisition: The Interpretability Hypothesis and the Full Transfer Full Access hypothesis (FTFA) (Cf. above the predictions of both theories).

48 L1 French speakers and 68 L1 English speakers took part in the study divided in three proficiency levels of Spanish: basic, intermediate, and advanced, according to a reading and vocabulary section of the MLA, and a cloze test from
the Diploma de Español como Lengua Extranjera (Spanish Embassy in Washington, DC). There were four tasks designed for the study: two elicited production tasks, a vocabulary test, and a picture identification task.

The results were highly consistent across the production tasks and the picture identification tasks. Number was unproblematic for all learners regardless of proficiency level or L1. In the case of gender, lower proficiency participants were more accurate on number than in gender and more accurate when the noun was masculine than in the case of feminine nouns. The intermediate and advanced participants did not differ significantly from the native speakers. There were significant effects for proficiency level but not for L1.

The results were contrary to the Interpretability Hypothesis predictions in two aspects: 1) L1 English speakers performed as well as on gender and on number as the L1 French speakers. It is important to remember that French marks gender, whereas English does not. The findings also proved the feasibility of acquiring interpretable gender features on nouns and uninterpretable features on adjectives and determiners. Learners are able to acquire gender agreement regardless of the status of the L1, contra the Interpretability Hypothesis. 2) According to the Interpretability Hypothesis, no proficiency effects should be found in the case of the L1 French speakers since the L1 and the L2 mark gender and there is nothing to learn than the gender of the nouns themselves: once the gender of the noun is known, the uninterpretable features on adjectives and
determiners are checked by mechanisms already in place because of the L1 grammar. The study found that the masculine gender is less problematic than the feminine gender, as other studies have found (Hawkins, 1998; Franceschina, 2001; Bruhn de Garavito and White, 2000).

It is remarkable that this study did not find any L1 effects at the low proficiency level: L1 French and L1 English speakers performed at the same level of competence. However, both the Interpretability Hypothesis and the FTFA, predicted differences in performance. No effects were found in the intermediate or the advanced levels regardless of the language. One possible reason for these results is that the tasks used in the experiment were not that demanding, which might have caused a “ceiling effect.” Unfortunately, the tasks are not included in an appendix, but one of the production tasks that consisted of an adaptation of the “Guess Who” game required what can be described as simplistic descriptions, especially for the intermediate and the advanced levels, such as: ¿Es un hombre viejo? (Is it an old guy?) Or ¿Tiene barba roja? (Does he have a red beard?). Three sample items from the picture identification task presented in the article with objects like suitcase, book, and socks were not demanding either.

In summary, gender has been studied from a diversity of perspectives. With regards to acquisition there are no conclusive findings. For some, gender seems to be present with a high accuracy (Bruhn de Garavito and White, 2000; White et al., 2004), even at intermediate levels of proficiency (White et al, 2004);
for others, despite having reached a high level of attainment, gender assignment continues to be non-target like although quite high (Franceschina, 2001).

However, the conflicting results might be due to two reasons: task differences and individual differences. In the case of White et al. (2004), as it was mentioned, some of the tasks were not demanding, which might have caused “ceiling effects.” Franceschina (2001) based her results on a single participant. While longitudinal studies are essential, extrapolating their findings to the whole population is problematic, especially when there is a single participant in the study and individual differences are studied.

Some of the consistent results across the studies are: the masculine gender seems to be the default for most of the participants; the number of errors in gender assignment is higher in adjectives than in determiners; and the availability of gender features in the L1 might be beneficial; nevertheless, there are similar problems for L1 speakers whose language lacks gender assignment and for those whose language presents gender assignment, which suggests that the absence of gender in the L1 is not the only factor affecting L2 gender acquisition (Bruhn de Garavito and White, 2000; Franceschina, 2001; Hawkins, 1998; White et al, 2004).

A general critique pertinent to all the studies reviewed is that the theoretical interpretation of the findings supports the theoretical perspective that frames the study. There is a clear cut distinction between those studies positing
that adult L2 learners do not have access to those aspects of the UG available in L1 acquisition and those studies supporting the opposite view. Franceschina (2001) and Hawkins (1998) results are interpreted as evidence that uninterpretable features are not available. Bruhn de Garavito and White (2002) and White et al. (2004) findings are consistent with the MSIH: both types of features – interpretable and uninterpretable – are available and variability is due to a deficiency in production.

**Count/ Mass and animacy feature acquisition**

The distinction between mass and count entities is considered to be widespread across languages. Gavruseva (2008) presents data in regards to Russian and Azerbaijani L1 children learning L2 English; Armon-Lotem, Crain, & Varlokoarta (2004), on the other hand, presents data of comparability of children and adults production of genitive and relative clauses in relation with the semantic feature [+/- count]. The distinction between mass and count nouns relates to the possibility or impossibility of counting the entities, for example, count nouns such as *bag* can exist in the plural, whereas *sugar* cannot. Another distinction is that numbers (numeral determiners) can be used to modify count nouns but not mass nouns.

10) a. Mary bought *three bags* of sugar.

b. *Mary bought three sugars.*
A semantic distinction between mass and count nouns is a) a singular count noun that applies to an object does not apply to its parts. For example, the word *table* applies to the whole noun but does not apply to its parts, that is to say, one of its legs. b) On the contrary, the word *oil*, for example, applies to all of its parts, for example a drop of oil can receive the name *oil*. Due to this, count nouns are not divisible and thus known as atomic, whereas a mass noun is divisible and thus non-atomic (Sera & Goodrich, 2010).

To my knowledge there are two articles that have investigated the acquisition of the feature [+/-count] under the minimalist account of research. However, in both cases the feature is not studied in isolation but in relation to other features. Gavruseva (2008) studies the L2 acquisition of the features Q[quantifier] and [+/- definite] and their relationship with the finiteness clauses. Armon-Lotem, Crain, & Varlokosta (2004) studies the relations between certain syntactic structures, including genitives and relative clauses, and semantic relations, including part-whole relations, possession, location, and the semantic features [+/- animate] or [+/- count].

Gavruseva (2008) explores the contingency in child L2 child acquisition of DP (specifically the functional feature Q [Quantifier], which quantifies DPs) and [+/- definite], and finiteness in clauses. The hypothesis is that clauses will lack finiteness, as long as the DP functional layer (especially the features Q and Definiteness) has not been acquired. In English the feature Q (quantification) in
the DP is important for the aspectual interpretation, as in the case of sentences as

*Mary counted beads vs. Mary counted five beads| the beads.*

The data, collected in spontaneous play sessions, is drawn from a longitudinal study of four children learning English naturally. The recordings comprise different time lengths that vary from 4 months till 9 months of language development. Participants’ ages are 6-9 years old. Their L1 lacks of overt determiners. The results, although not completely relevant to the current study, demonstrates that D and Q are underspecified in the earliest developmental stages, conclusion that proves the difficulty in its acquisition.

Armon-Lotem, Crain, & Varlokosta (2004) is a comparative study of children and adults in regards to the relationship between semantic and syntactic features. The experiment was conducted in English, Hebrew, and Greek. The participants were 25 speaking-children ranging from 3.3 to 6.3 years old. The range covers a period in which language is still undergoing change. Despite the age differences, the researchers decided to treat the entire age range as one group because the homogeneous results did not reveal any correlation with age. A group of 30 adult participants (10 for each language) was tested as a control group. The results from adult participants were compared to those of the children’s.

The semantic relations under study, which were concerned with the (in) alienability distinction, are the lexical distinctions between part-whole relations
and control relations due to the distinction between mass /count and animate/inanimate nouns.

<table>
<thead>
<tr>
<th>Part–Whole</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>(11) a. the man’s arm</td>
<td>b. the man’s tree</td>
</tr>
<tr>
<td>(12) a. the zebra’s neck</td>
<td>b. the zebra’s bushes</td>
</tr>
<tr>
<td>(13) a. the tractor’s wheels</td>
<td>b. *the tractor’s rock</td>
</tr>
<tr>
<td>(14) a. *the rice’s grains</td>
<td>b. *the rice’s nuts</td>
</tr>
</tbody>
</table>

For all [+count] nouns, part–whole relations between the whole and an autonomous part can be expressed syntactically using the genitive marker -’s, as in (11a) to (14a). The possessive marker ’s can also be used to express control relations (both possession and location) for [+animate] nouns, as shown in (11b) and (12b), but it cannot be used with [-animate] nouns, as illustrated in (13b). On the other hand, the possessive marker’s cannot be used with mass nouns to express either autonomous part-to-whole relations (14a) or control relations, as shown in (14b).

The results showed evidence that quite young children make the relevant semantic distinctions by age 3. According to the researchers, the cross linguistic results proved that the semantic distinctions, along as the interface distinctions – measured according to the plausibility of the part-whole and control lexical distinctions –, are not necessarily learned on the basis of the linguistic data. This was interpreted as an innate specification of the UG.
The feature [+/- animate] was tested in English in those relationships that can express, both possession and location only in the case of [+ animate] nouns, as shown in (11b) and (12b), but it cannot be used with [-animate] noun as exemplified in (13b). The conclusion, with regards to the English language, is that children were able to distinguish [+ animate] from [-animate] by using over 90% of the time the part-whole relation with a [+animate] possessor, such as the cowboy’s arm, but less than 50% of the time in the case of [-animate] possessor as in the tractor’s wheel. They clearly distinguish the semantic properties of the head noun (the possessed) and the subject (the possessor).

Although participants’ age is dissimilar to the population that will be used in this study, as well as the design and purpose of the study, it can be mentioned that the results of both studies are contradictory. In the first study, the [+/- count] feature appears to be underspecified and difficult to acquire in the case of L2 children acquisition; on the contrary, in the case of L1 acquisition, it is shown how in those three languages (e.g. English, Hebrew, and Greek) L1 children’s production makes semantic distinctions, at an early age of three, similar to the ones that adult production presents. In a weak conclusion based on only two studies, it seems that the semantic feature is difficult to acquire in the case of L2 acquisition, whereas in the case of L1 acquisition, participants are able to make relevant semantic distinctions that comprise the features [+/- count] and [+/-
animacy] even at a younger age (3 years old) than their counterparts in the first experiment (ages 6-9).

Predictions

Both hypotheses, FTFA and Interpretability Hypothesis foresee feature reassignment from the L1 in L2 acquisition; however, only the FTFA hypotheses predicts change in the developing interlanguage grammar in regards to uninterpretable features present in the L2 but not in the L1. According to the Interpretability Hypothesis, uninterpretable features not present in the L1 cannot be acquired. Since the population in the study will be English native learners learning Spanish, it is assumed that some differences will exist.

Summarizing the predictions, they are as follows:

1. Learners will perform more accurately on interpretable gender features than on uninterpretable gender features, according to the Interpretability Hypothesis.

2. Learners will perform equally on interpretable and on uninterpretable gender features, according to the FTFA.

3. Both hypotheses predict that learners will be able to acquire the semantic features [+/- count] and [+/- animate] since they are semantic (interpretable) features.
Addressing the acquisition of features within a Minimalist Theory and, more specifically, testing whether there is a distinction in the acquisition of interpretable and uninterpretable features in an effort to contribute to the debate on the causes of variability in L2 production is clearly warranted. Given that several researchers have called for the necessity of incorporating theoretical constructs (e.g., attention, awareness, noticing) to shed light in the complex process of understanding SLA under the generative strand of research (Birdsong, 2009; Lardiere, 2009a, 2009b), it would be of interest to incorporate the current theoretical models that address these issues in SLA (Tomlin & Villa’s model, 1994; Schmidt’s noticing hypothesis, 1990 and elsewhere; Robinson’s model, 1995b), and employ experimental designs that attempt to capture concurrent cognitive processes (e.g., Leow, 1997, 2000, Rosa & Leow, 2004; Rosa & O’Neill, 1999; Hama & Leow, 2010).

For example, an artificial determiner system that encodes the following features can be created: [+/-gender], [+ animacy], and [+/- count]. The use of an artificial determiner system permits control for previous exposure of the target structure while at the same time diminishes the prominent role of L1. The issue of learnability of features is not solely explained as a selection of individual features, but, in various cases, as a re-assembly of feature clusters and differing conditioning factors for their expression (Lardiere 2008; 2009a).
The benefits of incorporating the attentional strand of research to L2 studies under the UG is twofold: on the one hand, it is possible to test the attentional studies using a completely different formal framework to address how salient cues (gender, especially in the case of natural gender) reacts in relation to abstract features such as count and animacy. On the other, the more robust research design from the attentional strand of research will test more efficiently the assumptions that some researchers inside the UG field have in regards to the possibility of acquiring features from the presence of meaningful contrasts. (Lardiere, 2009a). For example, Carroll (2001) assumes that comprehension involves the extraction of meaning to form conceptual representations, and as a result learning a language is learning to parse the signal so that the sound stream is perceived as sequence of sounds. In the case of gender, for example, a recurrent finding in Bruhn de Garavito and White (2002) and in Hawkins (1998) study is that learners were more accurate in grammatical gender than in natural gender. A possible explanation offered is that learners rely on syntactic cues rather to non-linguistic cues that assign gender; however, the use of think-alouds protocols can reveal with certainty how participants process the input and also what and how deep they process the target items. In regards to Lardiere’s (2009a, p. 214), assumptions that “any feature contrast that is detectable is, in principle, ultimately acquirable”, the use of think-aloud protocols deals with the participants’
performance in experimental groups by helping to illuminate the actual learners’ processes while interacting with the input.

**Purpose of the study**

It is important to mention that results from studies under the UG strand of research are not straightforward comparable to those studies under the cognitive attentional strand of research. A series of theoretical and empirical differences exist. Generative Grammar assumes that language acquisition is the interaction between the innate endowed mechanism (UG), which shapes grammars by constraining the extent to which languages can vary, and the input received. The UG enables the construction of abstract (unconscious) knowledge that is formed with relative ease and speed based only on positive evidence (Herschensohn, 1999; White, 2003). Since UG is mainly concerned with the study of the representation of grammatical abstract knowledge, it usually does not address how the input is attended from a cognitive perspective. An interesting perspective within the UG is the one proposed by Carroll’s (2001) Autonomous Induction Model. Carroll’s complex proposal considers that the language faculty consists of several autonomous representational systems in which the linguistic information can flow in a limited way from the conceptual system to the grammar. As a result, the linguistic faculty involves a chain of levels of representation with different levels that interact with the physical stimuli at the lowest level and with the
conceptual representations at the highest. Two types of processors work at each level: a) the integrative processor, combining smaller representations into larger units, and b) a processor, which moves the representations from one level to the next. Each level of the modules has a form in which the representations are categorized and combined according to UG constraints and primitives or to rules learned. Interestingly, Carroll’s model leaves room for the role of awareness. According to her, the process just outlined is relevant for implicit and unconscious learning since “the lower levels of parsing seem to be unavailable to reflection or awareness” (p. 129), but inductive learning can consist of implicit and explicit learning. Feedback, correction, and metalinguistic instruction, which depend on awareness, play a role in the model, but only can influence the morphosyntactic information stored in lexical entries. SLA research interested in explaining the information processing operations – based on cognitive studies – has studied the essential role of attention for attended language input (Robinson, 2003). As the importance of attention and awareness increases and it is recognized how these two concepts help to explain several aspects of SLA (Schmidt, 2001; Leow, 1997, 2000, 2001), it is important to incorporate those concepts to the UG SLA studies.
Research questions

The present study seeks to answer the following research questions:

1) Do aware advanced L2 learners perform significantly better than unaware learners, as measured on recognition and production assessment tasks? If so, will the effects be maintained over a one week period?

2) Is it possible to learn form-meaning connections without awareness (at the encoding stage) above chance level? If so, do unaware and aware participants show learning of the undisclosed features besides gender?

3) Does type of features [+/- animacy] and [+/- count] have a differential effect on the level of awareness reported by advanced L2 learners during the reading task?

4) Is type of noun ending (–o and –a) and (–e and –n) errors related to the features [+/- animacy] and [+/- count]?

5) Is type of gender [masculine] and [feminine] errors related to the features [+/- count] and [+/- animacy]?

6) Does the type of ending canonical (–o or –a) and arbitrary (–e and –n) differ on (a) the recognition and (b) production task?
CHAPTER 3: RESEARCH DESIGN AND METHODOLOGY

Introduction

This chapter is divided into two sections. The first one begins with the pilot test and the corresponding qualitative and quantitative analyses. It concludes with a report of the adaptations that the pilot study underwent till the final version of the experimental design. The second, and main section, presents a detailed account of the actual experimental design, the instructional package, and the instruments. It concludes with the research questions and the statistical analyses needed to answer those questions.

Pilot study

The main purpose of the pilot test was to examine the length of the sessions, and to test the components of the experiment.

Participants

A pilot study was conducted with nine advanced learners of English who were enrolled at the Academic Writing class at the Georgetown University in the fall 2010, in which the materials and assessment tasks were tested.

All of those participants were native speakers of English. The original design of the experiment was a within-group design and, as a result, all the
participants had to complete both treatments: the animate vs. inanimate and the count vs. mass features. Participants completed the experiment in two sessions with a week difference between them. All participants completed the two sessions under the experimental condition [+ TA] since the main intention of the pilot study was to assess the treatment and the tests. The information gathered in the think-alouds was useful not only to detect awareness, but to discover how the participants approach the actual completion of the tasks.

Two participants were eliminated; in one case, this person did not complete the tasks while thinking-aloud while the other one produced an inaudible think-aloud. After having eliminated those participants, the final sample size was nine participants in the first session and seven participants in the second session. Those participants included in the final pool a) completed all the sessions, b) followed the instructions, c) produced an audible think-aloud, and d) were not a linguistic student.

In order to improve the experimental design, participants also completed a feedback questionnaire, on which they were asked to comment on the experiment (testing conditions, difficulty to read the texts, complete the tasks, etc.). The think-alouds were also used for this purpose since some comments in this regard were detected.
Training phase

All participants were exposed to the same determiner system each session. They were exposed to the animate vs. non-animate distinction in the first session, while in the second one they were exposed to the count vs. mass distinction. To avoid any order effect, in the actual study the treatment phase was randomized.

In this pilot test the training phase consisted of the same four texts (300-350 words each) for each session. The same four texts were used in the animate vs. non-animate session and the count vs. mass session, which has the inconsistency that participants encountered twice the same texts but with a different determiner system and, as a result, they also had to answer twice the same comprehension questions. To avoid the priming effect in the actual study, texts 1 and 2 contained the distinction animate vs. inanimate and texts 3 and 4 the distinctions count vs. mass.

Testing procedure

In the first session, participants completed a background questionnaire. The instructional package consisted of a) experimental task (2 texts and their corresponding comprehension questions), b) immediate posttest (MC and fill-in-the-blank), and c) post-exposure questionnaire (an explicit questionnaire exploring participants’ answers). In the first session they were exposed to the animate vs. inanimate distinction, while in the second session they were exposed
to the count vs. mass distinction. In the second and last session they had to complete the gender questionnaire aimed to assess their gender knowledge. 60 nouns in Spanish were presented and participants had to assign whether masculine or feminine gender using the article *el* (masculine) and *la* (feminine).

**Assessment tests: Fill-in-the-blank and GJT**

The pilot study consisted of two assessment tasks [grammatical judgment task (GJT) and fill-in-the-blank], in which the maximum score was 40 points in each one. The 40 items were divided in 20 old and 20 new. As a result of the pilot test, the assessment tasks were reduced to include 12 new and 12 old items. It was not possible to have 20 old target words that appeared 3 times in each text. This would have increased the length of the texts that would have made it difficult to complete in one session.

**Quantitative analysis**

**Training phase**

The training phase that consisted of reading the four texts and answering the comprehension questions for each text lasted approximately 30 to 45 minutes. One participant commented on the difficulty he experienced reading the texts with the changes in the determiners; however, he also mentioned that after some time he got accustomed to the texts. Another participant also pointed out that it was
odd that the texts were in the L2 while the questions were in the L1. However, as it was mentioned above, having the questions written in Spanish would have exposed participants to the L2 using the correct Spanish determiner system, which would cause a mismatch between the input they had received and the Spanish canonical usage of determiners.

**Testing phase**

Participants completed the fill-in-the-blank and the GJT while thinking aloud. This phase lasted approximately 30 - 45 minutes in most cases. Since the pilot study involved so few participants, it was not possible to conduct statistical analyses on the results; nevertheless, the descriptive statistics are presented.

**Aware vs. unaware participants**

The participants were classified as aware or unaware based on the think-alouds and the post-exposure questionnaire. In the first session there were three aware participants and six unaware participants for a total of nine participants. In the second session, there were seven participants, five of them classified as aware and two as unaware. The fact that more participants were classified as aware in the second session compared with the first was the result of having completed the first session previously. Before completing the training phase and the testing phase in the second session, participants already knew that they were going to be
tested not only on reading comprehension but on the semi-artificial determiner system as well; therefore, instead of reading the texts for meaning, participants were attempting to construct explicit rules regarding the determiner system ever since they were confronted to it in the very beginning of the experiment. The most appropriate solution was to modify the experimental design from a between-subjects design to a within-subjects design. Participants in the final experiment were randomized and took part in only one experimental condition: either animate vs. inanimate or count vs. mass instead of having to complete both experimental conditions.

As can be seen in Tables 3, the scores of participants classified as aware and those classified as unaware are relatively similar (45.41 vs. 45.83) in the first session on the fill-in-the-blank for the aware and unaware group, respectively, and slightly higher on the GJT (54.16% vs. 50%, aware and unaware, respectively).

In the second session (Table 4), nevertheless, there is a clear difference between the aware and unaware participants. In both assessment tasks, the aware participants scored higher than the unaware participants. Aware participants scored 59% and 65% compared with the 47% and 52.5% of the unaware participants on the GJT and the fill-in-the-blank.
Quantitative Results First Session

Table 3

*Percentage of correct answers in the GJ and fill-in-the-blank by aware vs. unaware participants in the first session*

<table>
<thead>
<tr>
<th></th>
<th>Aware participants (N = 3)</th>
<th>Unaware participants (N = 6)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M SD</td>
<td>M SD</td>
</tr>
<tr>
<td>GJT</td>
<td>54.16 6.29</td>
<td>GJT 50 5.7</td>
</tr>
<tr>
<td>Fill in</td>
<td>45.41 11.44</td>
<td>Fill in 45.83 10.40</td>
</tr>
</tbody>
</table>

Quantitative Results Second Session

Table 4

*Percentage of correct answers in the GJ and fill-in-the-blank by aware vs. unaware participants in the second session*

<table>
<thead>
<tr>
<th></th>
<th>Aware participants (N = 5)</th>
<th>Unaware participants (N = 2)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M SD</td>
<td>M SD</td>
</tr>
<tr>
<td>GJT</td>
<td>59 24.87</td>
<td>GJT 47.5 0</td>
</tr>
<tr>
<td>Fill in</td>
<td>65.0 11.18</td>
<td>Fill in 52.5 7.07</td>
</tr>
</tbody>
</table>

These results may be explained according to the number difference that exists between the aware and unaware participants in each session and the level of awareness reached for those classified as aware students. For the first case, it can be observed that the number of unaware participants was higher than the aware participants. In the second session, this trend was reversed; the number of aware participants was higher when compared to that of the unaware participants.

A second explanation may be the level of awareness reached for those classified as aware. In the first session, there was just one participant who
reported the highest level of awareness (namely, at the level of understanding), while in the second session there were three. Still, the number of participants was too small to draw any definitive conclusion (Table 5).

Table 5

<table>
<thead>
<tr>
<th>Level of awareness of those participants classified as aware</th>
<th>Awareness at level of noticing</th>
<th>Awareness at level of understanding</th>
</tr>
</thead>
<tbody>
<tr>
<td>First session</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Second session</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

This difference may be clearly seen when the results of the percentage of new and old items endorsed by the aware and unaware participants are compared. New items are the ones that mirror learning (Williams, 2004). The results obtained on the fill-in-the-blank (production task), the most demanding assessment task, in the two sessions are presented (Table 6 and Table 7). Aware and unaware participants endorsed a higher number of old items than new items in both sessions; nevertheless, in the case of the second session, the one with the larger group of aware participants who reached awareness at the level of understanding, the percentage of new items endorsed was higher.

Table 6.

Percentage of new and old items from the fill-in-the-blank endorsed by aware and unaware learners in the first session
Table 7

Percentage of new and old items from the fill-in-the-blank endorsed by aware and unaware learners in the second session

<table>
<thead>
<tr>
<th></th>
<th>Aware participants (N = 3)</th>
<th>Unaware participants (N = 6)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>New items</td>
<td>Old items</td>
</tr>
<tr>
<td>$M$</td>
<td>54.16</td>
<td>53.33</td>
</tr>
<tr>
<td>$SD$</td>
<td>7.21</td>
<td>14.71</td>
</tr>
</tbody>
</table>

Qualitative analysis

Think-alouds

All the think-alouds from the two sessions were analyzed to classify participants as aware and unaware and to detect any possible difficulty participants encountered as they completed the training phase or the assessment tasks. They were particularly useful to categorize participants as aware and unaware in the case of the second session that encompass the distinction between count and mass nouns. This distinction was more difficult to be verbalized on a post-exposure questionnaire and none of the participants were able to verbalize it at the end of the experiment. Having analyzed the verbal reports, it is clear that
some participants engaged the task trying to find a rule and adopted a rule-based approach. In some cases it was a rule that did not actually match with the one that underlies the system: “Mu is for persons that are female. Ci is for persons that female or male.” But in others there was some approximation to it: “Ru is for objects, ideas and concepts. Ci is for female concepts and things.” “Physical things and people should have Ci and not Ru.” “It seems like Ci is with individuals and Ru with ideas or abstracts concepts.” In either case, this effort categorizes all these participants as aware (Hama & Leow, 2010; Leung & Williams, 2011; Williams, 2005).

**Major changes**

There were some important changes made based on the feedback gleaned from the pilot test with respect to the final version of the experiment.

The training phase. The current experimental design incorporated a between-group design and, as a result, there was only one session instead of two. This reduced the effect of one experiment over the other as clearly observed in the increased number of aware participants in the second session of the experiment when compared to that in the first.

The materials. The four original texts were divided in two sessions and the length was increased to 700-750 words each. The participants would read two texts,
according to whether they were exposed to the animate vs. inanimate or the count vs. mass distinction.

Target items. Participants were exposed to the same amount of target items: 100 from each feature (animate vs. inanimate and count vs. mass) divided in 50 masculine and 50 feminine. Careful effort was made to ensure an even distribution not only in the text but in the paragraphs.

Assessment tasks. The GJT was replaced by a multiple choice to better test participants’ ability to recognize the target items. The recognition task presents the four possible combinations at the same time and assesses in a better fashion the recognition effects. Participants chose the best option according to what he or she encountered in the text. The GJT did not present the four possible options at the same time, but a single option that had to be evaluated as correct or incorrect. The number of items in both tests, recognition (multiple-choice) and production (fill-in-the-blank) was reduced to 24 instead of 40. These 24 items were divided in 12 new and 12 old, of which six were feminine and six were masculine. Another improvement was the fact that all the old items being tested were evenly distributed in the texts and appear an equal number of times (6).

Introduction to present study
The following section and the rest of the chapter present information about the purpose of the study, participants, the target items, the instructional package, the experimental tasks, assessment tests, the background questionnaire, testing procedure, scoring procedure, and statistical analysis used to assess the research questions presented in the previous chapter.

The purpose of the study was to test whether unaware and aware participants perform significantly different on a series of assessment tasks after they read a text for meaning. The purpose of the study was also to assess whether there was a significant difference in the results of both features being tested (animate vs. inanimate and count vs. mass) in relation to gender.

**Participants**

Participants in the study were 62 English-native speakers adults enrolled at a Northeastern university in the United States studying advanced Spanish as a foreign language. In this language program that follows a communicative approach in the classroom, the textbook *Repase y escriba. Curso avanzado de gramática y composición* (Wiley) is used. The advanced proficiency level was required as part of this experiment because participants took a gender test that will be explained below.

Participants were selected from an original pool of 117 participants. 55 participants were excluded from the final sample if they failed to fulfill the
following criteria: (1) Failing to attend all sessions; (2) Not following the instructions; (3) Failing to produce audible think-aloud protocols, or not complying with the instructions to think-aloud; and (4) Being a linguistic student or having taken some linguistic courses. All these criteria ensured that participants were as uniform as possible before completing the experiment. Additionally, all participants completed a Spanish gender test at the end of the second session.

Participants were also administered a background questionnaire at the end of the second session designed to elicit any prior language background, that is, (a) whether they had any L2 formal language instruction in other languages, (b) what other language(s), and (c) for how long, (d) and whether any of those languages coded gender.

Since the target structure was made up by the researcher, it was not necessary to assess participants’ prior knowledge of the structure or exposition of the target item outside the experiment.

To guarantee homogeneity of the sample, all the Spanish classes from which participants were recruited were from the same college. The classes have a strong emphasis on teaching the four skills (listening, speaking, reading, and writing) based on communicative functions and the use of the target language in a variety of contexts.

Target items

Description of the structure
The target items were four artificial determiners that were created *ex profeso* for the experiment. The selection of the artificial determiners was created in such a way that it did not resemble any of the articles that exist in Spanish. Therefore, the vowels *a, o, or e* were not used since they typically code masculine or feminine gender in Spanish. As it is the case in Spanish, in which a vowel usually marks gender, the vowels *i* and *u*, which never code gender, were chosen to mark gender in this experiment. The vowel *i* was chosen to be used in combination with all the masculine nouns used in the experiment, whereas *u* was used with all the feminine nouns. For instance, in experimental condition A, *ji* marked masculine animate, and *ju* marked feminine animate. *Ri* is masculine non-animate, and *ru* is feminine inanimate (Table 8). In experimental condition B, the same vowels *i* and *u* were used for masculine and feminine, respectively, but in a combination with a different consonant: *pi* and *pu* corresponded to mass, masculine and feminine, respectively, and *fi* and *fu* correspond to count, masculine and feminine, respectively (Table 9).

As it can be seen, a specific consonant was also used as a marker of that particular feature; therefore, the consonant *j*- was used for animate and *r*- for inanimate *f*- was used for count, and *p*- was used for mass.

Table 8

*Target items according to the distinctions +/- animate and gender*
A combination of a consonant and a vowel was chosen because it is the most common syllabic combination in Spanish: consonant + vowel. Any other combination resulted artificial and excessively salient since the purpose was to direct learners’ attention to the critical form and meaning mapping while reading a text for meaning. Since salience is the prominence of an item based on criteria of form and form-meaning mappings (DeKeyser, 2005; Gass & Svetic, 2003), it was extremely important to control this variable and use a combination similar to the one used in Spanish.
All the nouns presented in the reading materials, as it will be explained below, were used in the singular form since one of the distinctions to be presented to the learners was the count vs. mass distinction. Having some nouns in the plural form and some in the singular form would have provided a hint about the nature of the noun in this regard. If count nouns could appear in singular or in plural, while mass words only in singular, by reading the word padres (‘parents’) in plural, for example, the participants could find out that this noun is count since it appeared in the plural form. By having all the nouns in the singular form this possibility was ruled out. The determiners were also only definite and they did not distinguish between definite and indefinite determiners. Indefinite determiners appeared in the reading texts as they are used in Spanish (e.g., un ideal ‘an ideal’, una nueva relación ‘a new relationship’).

All the nouns that appeared in the tests were divided accordingly to the distinction [+/- animate], [+/- count], and [+/- masculine], in the case of gender. The words that belonged to each group were carefully selected in order to include words that were representative of each feature. This was done by inter-rater reliability. The list of words was revised by two different Spanish speaker raters who approved the final group of words included in the experiment. This was crucial in the case of the [+/- count] words. The words that belong to each category (count or noun) are not the same in English and Spanish, e.g., furniture or problem are noncount in English whereas in Spanish they are count ‘muebles’
and ‘problemas’. The final pool of words used in the experiment was representative of that category in Spanish. The words to be tested were also carefully chosen according to their endings in order to have equal number of masculine and feminine words ending in the canonical endings (–o or –a) as well as arbitrary (–e and –n).

**Independent variables**

The study investigated the effects of the following four independent variables:

1. Aware versus unaware participants
2. type of grammatical feature [+/- animacy]
3. type of grammatical feature [+/- count]
4. gender [masculine] and [feminine]

**Dependent variables**

The four dependent variables in the study were: text comprehension (multiple-choice comprehension questionnaire), recognition task (multiple-choice test), production task (fill-in-the blank test) of the target items and a Spanish gender questionnaire (fill-in-the blank gender test).

**Materials**

**Introduction**
All components of the study were computer-based which allowed for a tighter control of the variables, and more accurate measurement of time and accuracy scores. The experiment was delivered through a custom-built web application using the Blackboard pedagogical platform as a medium to deliver the tasks and the tests. The Blackboard site was created by Georgetown University’s Center for New Designs in Learning and Scholarship (CNDLS).

Since participants had to verbalize their thoughts, they wore a serial bus (USB) headset and talked into an attached microphone. The think-alouds were recorded in the CNDLS Language and Technology computer lab using Macintosh computers through the Quick Time Pro. Recordings were compressed and made available to download using a USB external portable hard-drive.

**Instructional package**

To ensure that all participants had the same baseline information prior to starting the experiment, an instructional package was given when they arrived at the language laboratory. The instructional package consisted of a) an IRB consent form, and b) a series of instructions to complete the different tasks and tests that comprised the experiment according to their experimental treatment condition.

While completing the experiment, they had to download other materials (i.e., the verbalization training materials, instructions for the texts, a vocabulary pre-training session, and the reading texts). They also had to complete a series of
tests (i.e., a vocabulary pre-training test, a comprehension test, a recognition test and a production, as well as a gender test) and a series of questionnaires (i.e., debriefing questionnaire, and post-exposure questionnaire).

**Verbalization training materials**

This study employed think-aloud protocols as a measure of attention and noticing. Therefore, all participants had to complete the training session and the experimental tasks while thinking-aloud. They received the following instructions just prior to engaging in the experimental task. These instructions are modeled from those of Bowles (2005), Ericsson and Simon (1984; 1993), Leow (2000), and Sanz et al. (2009).

In this experiment, I am interested in what your thoughts are as you perform some tasks that I give you. In order to do this, I am going to ask you to THINK ALOUD as you do the different tasks. What I mean by “think-aloud” is that I want you to say aloud everything you would silently say to yourself. When you are instructed to begin, I would like you to begin thinking aloud, that is, verbalize your thoughts, and continue to do so the entire time as you are doing the tasks. Speak in ENGLISH.

After these instructions, participants were exposed to one example of what a think aloud protocol might look like, which was usually either arithmetic or language practicing tasks. An arithmetic task was chosen following Bowles (2005), and Sanz et al. (2009). Sanz et al (2009, p. 67) preferred this sort of task since “a Math problem was deemed to be more neutral than a language problem;
in this way, we avoided to put words in participants’ mouths.” Participants read
the following example adopted from (Bowles, 2005):

If I were working through a math problem, it might sound like this:

\[ 24 + (7 \times 3) = \_\_\_\_\_\_\_ \]

“OK. I have to multiply first … so seven times three … um, that’s 21… Add
24 to that… OK. Four plus one is five … two plus two is four … so, uh, 45
yeah, 45.”

After the practice, participants were given the opportunity to ask questions
about thinking aloud. To ensure that participants did think aloud, throughout the
experiment they were reminded to think aloud while completing the training tasks
as well as the testing phase.

Treatment / Experimental tasks

Reading Materials

Four texts on the Latin American family, taken from the book Tradición y
Cambio. Lecturas sobre la cultura latinoamericana contemporánea, were
modified. The textbook is used in a second year language and culture course of
Spanish as L2 at the same university. The criteria for using these texts were: (a)
text length, (b) text genre, (c) readability, (d) topic, (e) target items. The length of
each text was adapted so that each one had 700-750 words each. A fairly similar
length in all the texts controls for a balanced exposure to the texts across the two
experimental groups. The modified text, descriptive and non-fiction, was similar to any of the assigned texts participants had to read as part of their reading materials about Spanish or Latin American culture. Having one single topic for all the texts also allowed to frame the readings under a common subject that avoids an abrupt transition and the activation of new cognitive schemas for each text.

Readability was also a criterion that was carefully controlled. The descriptive text used is clearly organized around topical paragraphs that included simple sentences, clear transitions, and explicit pronouns. It was extremely important to have a text with a low readability index since participants were using the text to be exposed to the semi artificial determiner system. A difficult text would have been distracting by increasing the time participants spent trying to decipher the message, and, as a result, lowering the time they spend making the essential form-meaning connections.

The four reading texts were submitted to a test of readability to ensure that all the texts have the same readability index in order to rule out any confounding role readability could have played in the results. The well-known text readability index used was Fernández-Huerta (1959). The resulting readability index was relatively similar in all the four texts, 80 and 78 for the animate vs. inanimate condition, and 79 and 80 for the mass vs. count condition. According to the readability index (Fernández-Huerta, 1959), texts in the 70-80 range are considered to be “fairly easy” according to the scale.
The topic was also chosen because it offered enough instances of animate and inanimate nouns. The intention was to balance attention between the meaning of the text and the necessary form and meaning mappings needed to decipher the rules that underlie the determiner system. More importantly, all the definite determiners in each text were modified according to the determiner systems described above.

A series of drafts were modified to include ample possible target items. After numerous trials and extensive piloting of the texts, a final version was produced containing all the final targeted items. In order to address the potential impact of the placement of the target items in the text, as pointed out by Leow, Hsieh and Moreno (2008), careful attention was paid to the number and distribution of the items throughout the texts. An equivalent number of target items for each feature was included in the texts. A total of 100 animate and inanimate and 100 count and mass nouns, equally divided in masculine and feminine were carefully chosen and included in the texts (Tables 10, 11, 12, and 13). As special attention was paid to the distribution of the items, they were evenly distributed within the paragraphs. In none of the paragraphs was there a gap difference of three features (Table 14 and Table 15). Due to these two methodological improvements, participants were exposed to an equal number of target items that were evenly distributed.
Total distribution of the target items by feature: Animate vs. Inanimate and texts

<table>
<thead>
<tr>
<th></th>
<th>Text 1 Masc.</th>
<th>Text 2 Masc.</th>
<th>Total</th>
<th>Text 1 Fem.</th>
<th>Text 2 Fem.</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Animate</td>
<td>27</td>
<td>23</td>
<td>50</td>
<td>24</td>
<td>26</td>
<td>50</td>
</tr>
<tr>
<td>Inanimate</td>
<td>25</td>
<td>25</td>
<td>50</td>
<td>24</td>
<td>26</td>
<td>50</td>
</tr>
</tbody>
</table>

Table 11

Total distribution of the target items by gender and feature: Animate vs. Inanimate

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>Total</td>
<td>100 Animate</td>
<td></td>
<td>100 Inanimate</td>
<td></td>
</tr>
</tbody>
</table>

Table 12

Total distribution of the target items by gender and feature: Count vs. Mass and texts

<table>
<thead>
<tr>
<th></th>
<th>Text 1 Masc.</th>
<th>Text 2 Masc.</th>
<th>Total</th>
<th>Text 1 Fem.</th>
<th>Text 2 Fem.</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count</td>
<td>24</td>
<td>26</td>
<td>50</td>
<td>27</td>
<td>23</td>
<td>50</td>
</tr>
<tr>
<td>Mass</td>
<td>25</td>
<td>25</td>
<td>50</td>
<td>24</td>
<td>26</td>
<td>50</td>
</tr>
</tbody>
</table>

Table 13

Total distribution of the target items by gender and feature: Count vs. Mass

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>Total</td>
<td>100 Count</td>
<td></td>
<td>100 Mass</td>
<td></td>
</tr>
</tbody>
</table>

Table 14

Distribution of the target items in the paragraphs: Animate and Inanimate

<table>
<thead>
<tr>
<th></th>
<th>Text 3</th>
<th>Total</th>
<th>Text 4</th>
<th>Total</th>
<th>Total Text 3 and Text 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>JI</td>
<td>8</td>
<td>10</td>
<td>10</td>
<td>28</td>
<td>22</td>
</tr>
</tbody>
</table>

121
As mentioned above, there were two experimental conditions to which the participants were randomly assigned. The difference in each group was whether they participated in the experimental condition A (animate vs. animate) or experimental condition B (count vs. mass).

All the participants were asked to think aloud while completing the training, treatment, and tests. Their recordings were given an identification number and saved for subsequent transcription and coding to identify awareness or unawareness as well as possible levels of awareness.

The Blackboard Grade Center was used to gather all the information regarding questionnaires and tests for successive coding, analysis, and scoring.

### Assessment Tests

### Introduction
There were four assessment tests, two of which were critical for the study since they measured the result of the treatment task. These two critical tests were the multiple-choice (recognition task) and the fill-in-the-blank (production task). The other tests are a comprehension reading questionnaire, and a gender test, which consisted of the same version for all the participants. The first one measured reading comprehension and the questions followed the order of the text; the gender test, on the other hand, measured participants’ knowledge of gender assignment in Spanish.

Critical tests

In this study there were two assessment tests: a recognition test in the form of a multiple-choice, and a production test in the form of a fill-in-the-blank. In this study, learning was operationalized as the ability to demonstrate knowledge of the animacy feature at the production level of the old items – trained items – or at the level of generalization to new items.

The recognition and production tests were administered in an immediate and post-delayed design, in the order of production and recognition for all the participants. It is important to mention the rationale for having this fixed order of presentation. Methodology textbooks recommend counterbalancing, which refers to have different order of presentation of the tests to reduce the overall influence of one test on the performance of the other (Mackey & Gass, 2005). In this case,
the order of the tests was always the same to avoid participants receiving extra input from performing the multiple choice test first. The multiple-choice test had the four possible answers present on the screen, while the fill-in-the-blank asked participants to write the correct answer according to what they had previously seen without having them present on the screen.

The same questions were presented in the immediate post and delayed tests, but they were randomized by the computer; as a result, the version of the tests was different for each participant.

Although there is confusion in the operationalization of the terms recognition and production, and taking into account the processing stages of learning (e.g., intake and L2 development), intake was operationalized as the capacity to recognize the target items, which was measured using a multiple-choice task, while learning was operationalized as the ability to produce the target items, which was measured using a fill-in-the-blank test.

The recognition and the production tests were both administered in an immediate and delayed posttest design, with the following order of presentation: production and recognition, as explained above. The questions were presented in random order each time the test was administered.

Each test consisted of 12 old items and 12 new items for a total of 24 items divided in an equal number of masculine and feminine nouns. All the old items tested appeared in the texts three times in each text for a total of six times
for each session since participants received two texts in each session. The number of exposure of the items (six times) was selected as some studies addressing the impact of exposure frequency on incidental vocabulary growth have found that six encounters resulted in significantly more receptive as well as more productive word knowledge compared with two and four exposures (Rott, 1999; Rott, 2007). The target words were also evenly distributed throughout the paragraphs (Table 14 and Table 15). Having an equal number of exposures and being evenly distributed controls tightly some possible methodological issues.

**Recognition test: Multiple-choice (MC)**

The purpose of this task was to elicit whether the participants were able to recognize above chance the targeted items. The MC comprised 24 items, divided into 12 items that assessed [+count] or [+ animate] and 12 that assessed [-count] or [-animate], for the first or second systems, respectively. 12 of them were old items and 12 were new items and an equal number of masculine and feminine words. In this test participants had to mark those sentences that were correct according to the logic of the determiner system. An example is seen in Figure 2.

**Figure 2**

*Multiple-choice (recognition test)*

<table>
<thead>
<tr>
<th>Please choose the correct answer.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. ___________ madrina.</td>
</tr>
</tbody>
</table>
Production test: Fill-in-the-blank

The purpose of this task was to test whether participants were able to produce the target items in a fill-in-the-blank task immediately after exposure to the texts. This test elicited controlled productive knowledge of the target items.

The fill-in-the-blank consisted of 24 items, divided into 12 items that measured [+count] or [+animate] and 12 that measured [-count] or [-animate], for the first and second determiner systems, respectively. 12 of them were old items and 12 new and an equal number of masculine and feminine words. In this test participants had to produce one of four determiners according to the variables that rule the determiner system: masculine/feminine and [+/- count], for the system A, and masculine/feminine [+/- animate], for the system B. An example is provided in Figure 3.

Figure 3

*Fill-in-the-blank (production test)*

Please fill in the blanks with the correct word. Use what you found in the texts as a
**Reading test: Multiple-choice comprehension assessment task**

Reading comprehension has been measured using different techniques. None of them is completely satisfactory and, as a result, “there is no ‘best method’ to test reading” (Alderson, 2000, p. 203). Each method presents advantages and disadvantages that need to be known to the developer of the test. Although many methods have been used for testing reading (i.e., (a) cloze test and gap-filling tests, (b) free recall tests, (c) think-aloud procedures, (d) reading comprehension questionnaires, (e) matching techniques, (f) ordering tasks, (g) short-answer tests, among others), multiple-choice comprehension questionnaires are the most common method to test reading comprehension (Alderson, 2000).

The multiple-choice comprehension questionnaire measures reading by asking the reader to choose the best answer that responds to the comprehension question. This type of reading measurement allows testers to regulate the range of possible answers to the comprehension question, and, as a result, can be scored automatically by a machine. However, researchers have questioned whether it is a reliable measure of understanding. Some have argued that deliberately tricking

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>model:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. __________________ niño.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. __________________ felicidad.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. __________________ máquina.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. __________________ familia.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. __________________ madre</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

...
students may result in a false measure of understanding. It is also mentioned that the ability to answer multiple-choice questions is a distinct skill, completely different from the reading ability. A severe difficulty with multiple-choice questions is that the tester does not know why the participant responded in that fashion. The tester might have guessed or used test-taking strategies (Alderson, 2000).

Despite these critiques, a multiple-choice comprehension questionnaire, an objective method, is the most commonly used measure of global understanding (Grabe, 2009). Since the purpose of the study was not to measure understanding per se, this method, which adopts a top-down perspective, resulted adequate. This assessment technique was selected for the following criteria: (a) as it was mentioned previously none of the techniques is the ‘best method’ to measure this skill, (b) questionnaires in the L1 are commonly used to measure text comprehension at the college level, and (c) having at least three options reduces the possibility that participants’ correct answer is the result of chance.

The 10-item multiple-choice assessment task designed to test comprehension had three possible answers for each question. Participants received one point for each correct item, and zero point for an incorrect item.

In this study, reading comprehension is operationalized as the ability to understand the meaning encoded in the texts while reading the L2 for meaning. The questions adopted a top-down perspective. The main purpose of the
comprehension questionnaire was to engage participants in the exposition task. By telling the students that the rationale of the study was to measure their reading comprehension, they were distracted as much as possible from adopting an explicit strategy that would potentially trigger an aware state regarding the semi artificial determiner system.

The comprehension questionnaire was delivered in the L1 despite the fact that the texts were written in Spanish. The intention was not to prompt the participants with the L2 that did not respond to the determiner system used in the experiment.

**Gender test**

An important part of the experiment was the recognition and production gender test. Participants completed a gender test in which they had to prove that they were able to assign gender in Spanish. As mentioned above, nouns were either the natural designation of gender (i.e., el hijo ‘the son’) or the discretionary usage of gender (i.e., la mesa ‘the table’). However, nouns with different endings than those that generally marked gender (–o and –a, for masculine and feminine, respectively) were also used, i.e., ‘la felicidad’ (the happiness, feminine), ‘la canción’ (the song, feminine), ‘la vejez’ (the old age, feminine), ‘la costumbre’ (the custom, feminine); as well as exceptions, i.e., ‘la mano’ (the hand, masculine), ‘el clima’ (the weather, masculine), ‘el planeta’ (the planet,
masculine). The baseline for inclusion was 90%. Previous studies in Spanish gender assignment report a production of correct gender assignment in the range of 91.07% (Franceschina, 2001) and 81.5% and 89% (Bruhn de Garavito & White, 2002).

By testing gender production and recognition, any differences in the posttests will be due to the interaction of the bundle of features [+/- animate] and [+/- masculine] and [+/- count] and [+/- masculine].

**Measurement of awareness**

Awareness was measured in two ways, offline and online. Offline measurement was conducted after finishing the first session using a post-exposure questionnaire while the online measurement was the use of think-aloud protocols.

Post-exposure questionnaires are one of the most common methods to collect information in second language research (Dörnyei, 2010). Depending on the type of questionnaire (i.e., open-ended questionnaire, closed-item questionnaire, interview, or a combination of different types), the data gathered can be categorized as either qualitative or quantitative. The open-ended questionnaire design follows the recommendations mentioned in Gass and Mackey (2007): it was displayed in a user-friendly format, presented through a computer interface, and its items were formulated as clearly and unambiguously as possible avoiding any researcher bias.
Think-aloud protocols have proved to be the most effective method to gather data regarding awareness and levels of awareness at the concurrent stage (Leow, 1997, 2000; Rosa & O’Neill, 1999). However, by employing the post-exposure questionnaire, it was possible to complement and triangulate data collected on the think-aloud protocols regarding the treatment session as well as awareness and possible levels of awareness (cf. Hama & Leow, 2010 for a similar design). Still, post-exposure questionnaires, as a tool to measure awareness and levels of awareness, have been criticized on the basis of memory decay (Ericsson & Simon, 1984, 1993; Bialystok, 1979; Leow, 2000).

The post-exposure questionnaire had to be completed by the participants at the end of each session. The first question surveyed whether participants could identify what the purpose of the study had been. The following four questions, framed to measure awareness, asked participants’ rationales for the usage of each one of the four determiners used in the text. The subsequent question asked participants to group the four determiners into groups of two related determiners. The following question asked participants to justify why they grouped the determiners in that particular order. Subsequently, they were asked if they could explain how the target items worked and if so, when they became aware of how the articles worked. The next three questions asked participants’ opinions about the think-aloud procedure. The following two questions asked participants
whether they had taken any linguistic class, and the last question requested participants’ overall opinions about the whole experiment.

**Procedure**

The University Internal Review Board approved all materials and procedures (Reference IRB # 2010-567). To recruit participants for the experiment, the researcher visited seven Advanced Spanish classes in the non-intensive program. These students received three classes per week for a total of 150 minutes. (50 minutes each day). The researcher invited participants by briefly explaining some details about the experiment. Since the actual purpose of the study could not be disclosed, participants were told that the purpose of the study was a comprehension reading exercise that would be delivered computer-based. Those willing to participate were asked to offer their contact information and to choose and sign up for the two sessions that comprised the experiment on a particular date. Participants received extra points in an exam as compensation for their time. They received a written invitation to participate and a document explaining the different sessions of the experiment.

The experiment had two sessions (Figure 4). On the day of the treatment, participants reported to the language laboratory, where they received the IRB consent form, instructions explaining how to complete the session, and were randomly assigned to one of the two treatment conditions. The study was
computer-based and had been designed using the power point software and the Blackboard system. Two different instructions sheets were created for each of the two experimental conditions: animate vs. inanimate and count vs. mass distinction.

Participants also received writing instructions to think aloud during the training phase, the reading phase, and while completing the assessment tasks. In the second session, taking part one week after the treatment session, participants were assigned to the same experimental condition, and completed the assessment tests – while thinking aloud –, a post briefing questionnaire and two gender tests.

Since it was essential to the experiment to detect any instance of awareness, participants’ awareness was measured using a hybrid method. Awareness was measured on-line, using think-aloud protocols, and offline, completing a post treatment questionnaire. After having completed a think-aloud training session, participants were instructed to use the headset to begin the experiment. Upon completing the training session, participants read the treatment texts and then answered the comprehension questions, and completed the assessment tests. The assessment tests were delivered in the order of production (fill-in-the-blank) and recognition (multiple-choice) to avoid additional exposure to the target items by using the multiple-choice test before the fill-in-the-blank. At the end of the first session participants had to complete a post-exposure
questionnaire. The second and last session of the experiment took part a week after the first session.

Think-aloud training task

After receiving the instructional package, participants received instructions on how to verbalize aloud. As previously explained, participants read instructions of what a think-aloud might sound like, using a mathematical example. After reading the think-aloud example, participants were asked to practice verbalizing aloud. The practice consisted of four mathematical tasks, completely different to the language task they had to verbalize when they took the experiment. Participants were told to ask the researcher any questions they had regarding the think-aloud procedure.

Vocabulary Pre-training session

Upon completion of the think-aloud training task and as a preparation for the vocabulary pre-training session, participants received the following instructions that informed them about the purpose of the study as well as the changes in the determiners. Instructions were the same despite the experimental condition with the only difference being the articles used, ji, ju and ri, ru, for the animate vs. nonanimate condition, and pi, pu and fi, fu for the mass vs. count distinction.

Instructions: Texts 1 and 2
This is an experiment about how people formulate sentences in different languages. Different languages have different ways of saying things, which raises the question of whether this forces people to think slightly differently in order to speak and understand in different languages. Here I am investigating this issue in a situation where the sentences are almost entirely in Spanish, apart from the articles.

When you read the text below, please note that the Spanish articles *el* or *la* are not used but instead…

Masculine articles correspond to
JI and RI and
Feminine articles correspond to
JU and RU

Please keep this in mind as you read the texts for the cultural information reported in them. You will be asked to answer questions based on the content of the texts.

The vocabulary pre-training session used was delivered using the Blackboard platform. It consisted of 6 power point presentations with 20 target items each for a total of 120 expositions to the target items. The 20 items were equally distributed in ten masculine and ten feminine as well as ten animate and ten inanimate or ten mass and ten count, depending on the experimental condition (Tables 16 and Table 17).

Table 16.

*Distribution of the target items Animate vs. Inanimate condition in each power point presentation*

<table>
<thead>
<tr>
<th></th>
<th>Masculine</th>
<th>Feminine</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>N = 10</td>
<td>N = 10</td>
<td>Total N = 20</td>
</tr>
</tbody>
</table>

Table 17
Participants were presented with the four elements of the determiner system by explaining that they were going to read two texts in Spanish and would have to answer a series of comprehension questions for each text. However, it was mentioned that all the articles (el ‘masculine’ y la ‘feminine’ English, “the”) in the texts had been modified and a different determiner system was used instead of the one used in Spanish. As it is customary in this sort of experiments, only one of the two features that make up the determiner system was disclosed, whereas the other was undisclosed. Participants were told that the distinction responded only to the gender distinction: masculine and feminine even though there were four determiners. More specifically, the gender distinction was disclosed, but the feature distinctions [+/- animacy] and [+/- mass] were not disclosed to test whether or not participants were going to become aware of the animacy or the mass feature (according to each experimental condition) after being exposed to the experimental tasks (cf. Table 18 and Table 19).
Table 18

The artificial determiner system used according to the disclosed and undisclosed features: Animate vs. inanimate

<table>
<thead>
<tr>
<th>Participants were not told:</th>
<th>Animate</th>
<th>Inanimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants were told:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Masculine</td>
<td>ji</td>
<td>ri</td>
</tr>
<tr>
<td>Feminine</td>
<td>ju</td>
<td>ru</td>
</tr>
</tbody>
</table>

Table 19

The artificial determiner system used according to the disclosed and undisclosed features: Count vs. mass

<table>
<thead>
<tr>
<th>Participants were not told:</th>
<th>Count</th>
<th>Mass</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants were told:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Masculine</td>
<td>pi</td>
<td>fi</td>
</tr>
<tr>
<td>Feminine</td>
<td>pu</td>
<td>fu</td>
</tr>
</tbody>
</table>

Participants were asked to complete a training phase to familiarize them with the novel determiner system they were going to encounter in the reading phase of the experiment. In order to do so, they were presented with a combination of a noun (not presented in the reading texts) and an instance of the novel determiner system. They were asked to complete the following instructions: (a) repeat aloud the determiner and the noun, to make a mental image of the combination, (b) repeat aloud the determiner and the gender as it appeared on the screen (i.e., ju feminine), (c) repeat again aloud the determiner and the noun, (d) form a mental image of the combination of the determiner and the noun, and
finally, in order to continue with the following slide, (e) they had to correctly choose whether the noun presented was ‘masculine’ or ‘feminine’ – once it had disappeared of the screen – while thinking aloud. If the selection was correct, they could move forward and continue with the following slide. They were presented with 120 target items equally distributed in gender (masculine and feminine), as well as type feature (animate vs. inanimate and mass vs. count nouns).

Participants completed the pre-training session while thinking aloud to measure any potential level of awareness. After having finished the pre-training phase, participants had to complete a post-training test to confirm that they had learned the masculine and feminine distinction, which in both treatment conditions corresponded to –i to mark masculine and –u to mark feminine. The training session lasted between 15 and 22 minutes.

Training task

After having received the instructions, the participants read their texts according to their experimental condition. They were prompted to read the two texts for meaning. They were told that in order to assess their comprehension, they were going to answer a ten-question comprehension questionnaire after finishing reading each text, without being able to refer back to the text. Each question appeared on the screen and it was not possible to backtrack to the previous questions. As was mentioned before, this training task was the main
exposure task. Participants encountered in the texts, for both treatment conditions, a total of 100 target items, evenly distributed within the text, and an equal number of target items, 50 and 50 according to the treatment condition [+/- animate] and [+/- mass], and 50 masculine and 50 feminine. The training task lasted between 20 and 34 minutes. The two assessments tasks (fill-in-the-blank and multiple-choice) were completed after the reading and comprehension questionnaire tasks. One week later, participants completed the delayed posttest (Figure 4).

Research Design Overview: Experimental Design

First session

<table>
<thead>
<tr>
<th>IRB Consent Form</th>
<th>Treatment A</th>
<th>Treatment B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Animate vs. Inanimate</td>
<td>Count vs. Mass</td>
</tr>
<tr>
<td>Training phase</td>
<td>(120 target items)</td>
<td>Training phase</td>
</tr>
<tr>
<td></td>
<td>(120 target items)</td>
<td>(120 target items)</td>
</tr>
<tr>
<td>Testing phase (MC 1 and MC2)</td>
<td>Testing phase (MC 1 and MC2)</td>
<td></td>
</tr>
<tr>
<td>Two texts/ Exposure treatment</td>
<td>Two texts/ Exposure treatment</td>
<td></td>
</tr>
<tr>
<td>Animate vs. Inanimate</td>
<td>Count vs. Mass</td>
<td></td>
</tr>
<tr>
<td>Think aloud</td>
<td>Think aloud</td>
<td></td>
</tr>
<tr>
<td>Two comprehension questionnaires</td>
<td>Two comprehension questionnaires</td>
<td></td>
</tr>
<tr>
<td>Testing phase: fill-in-the-blank (production test) and MC (recognition test)</td>
<td>Testing phase: fill-in-the-blank (production test) and MC (recognition test)</td>
<td></td>
</tr>
<tr>
<td>Think aloud</td>
<td>Think aloud</td>
<td></td>
</tr>
<tr>
<td>Post-exposure questionnaire (Measure Awareness)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Second Session (A week later)
<table>
<thead>
<tr>
<th>Treatment A</th>
<th>Treatment B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Animate vs. Inanimate</td>
<td>Count vs. Mass</td>
</tr>
<tr>
<td>Testing phase: fill-in-the-blank (production test) and MC (recognition test)</td>
<td>Testing phase: fill-in-the-blank (production test) and MC (recognition test)</td>
</tr>
<tr>
<td>Background Questionnaire</td>
<td></td>
</tr>
<tr>
<td>Gender Questionnaire 1 and 2</td>
<td></td>
</tr>
</tbody>
</table>

**Scoring and coding**

Each correct target item was awarded one point if the answer was correct and zero point if the answer was incorrect. The grade center on Blackboard saved the answers for subsequent quantitative and qualitative analysis. The post-training phase that followed the pre-training phase consisted of two multiple-choice tests with 12 items each that measured participants’ intake of the gender assignment rules. The total score for each test is 12. The comprehension questionnaire presented ten multiple choice questions for each text for a maximum score of ten points. The fill-in-the-blank and the multiple-choice post-treatment tests consisted of 24 target items divided into 12 masculine and 12 feminine, as well as 12 animate and 12 inanimate or 12 mass and 12 non count nouns, depending on the experimental condition. The total score for each task – recognition and production – is 24 points.

The think-aloud protocols of all the participants were transcribed for further analysis.
Qualitative analyses

Think-aloud protocols

This first section of the qualitative analysis presents common findings in the reading tasks for both types of features (1) animate vs. inanimate and (2) count and mass. As mentioned in the Methodology section – Chapter 3 –, in order to analyze in depth and detect any instance of awareness, this hybrid-design relies on post-exposure questionnaires and think-aloud protocols to classify participants as Unaware and Aware. Two raters analyzed the post-exposure questionnaires and, essentially, the think-aloud protocols to classify as aware any participant who provided a report of being aware of the targeted forms (e.g., animate vs. non-animate or count vs. mass) or provided a metalinguistic description of the rule underlying the determiner systems. The intention was to triangulate the data and detect even low levels of awareness. Participants that failed to do so were classified as unaware. Two thirds of the think-aloud protocols were analyzed by the researcher and a different expertise researcher to verify the coding. Interrater reliability was 100%.

Most of the participants in both groups read the text in a linear manner; either reading aloud in Spanish, or reading aloud a translation of the text in English. Although most of the participants read the text from the beginning to the end, some participants approached the text making pauses in order to summarize or paraphrase ideas every time a paragraph was finished.
Few participants commented upon the semi artificial determiner system or mentioned anything about the difficulty of reading the text with a different determiner system, a pattern found in the pilot tests as well. One participant, for example, commented about the difficulties of reading the text with a different determiner system, but before starting the second text the same participant said that she had got accustomed to the new artificial determiner system. Another participant made a comment during the comprehension questionnaire mentioning that it could have been easier to answer the comprehension questions if the text had not included that semi-artificial determiner system. Regardless of these comments, a baseline for inclusion as part of the final pool of participants was a score of sixty percent of correct answers on both comprehension questionnaires.

**Animate versus Inanimate**

As mentioned previously, the think aloud protocols were the main source for measuring awareness or lack thereof during the experiment. Throughout the training phase of the experiment, which took no less than 20-25 minutes, there was no reported knowledge of the animacy rule underlying the determiner system. During the reading task none of the participants mentioned any example of awareness of animacy or showed a rule formation approach.

The think aloud protocols revealed instances of awareness of the animacy rule during the testing phase of the experiment. Six out of the seven participants
classified as aware were detected in this section, and only one participant reported becoming aware during the training session although no such report was made in the think-alouds but in the post-exposure questionnaire. Having to choose a correct option, some of the participants adopted a successful rule formation strategy. The fact that most of the aware participants deciphered the rule while completing the recognition or production tasks explains why the score of the aware participants was not perfect for this entire group of participants. Being aware or becoming aware – the analysis of these think-aloud protocols seems to show – is a process and not a fixed state.

Although results were analyzed by groups aware vs. unaware, as in previous studies in this strand of research (Williams, 2004, 2005; Hama & Leow, 2010) and not by level of awareness, the analysis followed Leow (1997, 2000) and intended to detect different levels of awareness. At the level of noticing, there were three participants who made a comment on some aspect of the animacy rule either by a partial report or an incorrect rule, as illustrated below in (1):

(1)  
   a) Respeto (‘respect’), I’m going to put ri. I think that j is for people.
       Ciudadana? I don’t know if it’s ru or ju. Ju? (Choosing correctly)
       The same person a few items after:
       El profesor. Ji is a person. So, ji.
At the level of understanding—the rest of the four participants—, there was only one person that provided the correct rule immediately after the beginning the assessment phase (2):

(2) a. Juventud is not a person. Miembro is ji miembro. It’s a person.

For the majority of the participants who either provided a partial or an incorrect rule becoming aware was a process that could be followed as the think-aloud protocols were analyzed; a conscious effort to learn something that classified participants as aware. An example of an incorrect rule formation is the following, in which the participant verbalized a partial rule that s/he did not use since the participant ended up employing j for animate and not only for family and r for non-animate words (3):

(3) a. I don’t know still how to distinguish between ju and ru and ji and ri.
   The same participant later
   a. Honestly, I’m going to have family with J and all the rest with r.
      Ru vecina.
      Ji profesor.
      Ru casa.
      Ri rio.
      Ji alumno.
      Ju ciudadana.
      Ru ciudad. Let’s go with ru ciudad.
This participant adopted conscious and engaging approaches that led her to find the correct rule formation even though at the beginning of the task the participant mentioned uncertainty. During the recognition task (4):

(4)  
  a. Profesor? I don’t remember this one but I’m going to go with ji. …  
      Respeto? No, probably ri….  
      Ciudadana? I don’t remember this one but I’m going to go with ju.  
      *People and living things are going to be J’s. I'm positive*  
      Nieto is definitely Ji.  
      Factor is ri factor (…)  
      Novio is Ji novio.  
      Hija it was definitely ju hija.

Some of these participants even at the end of the experiment and after having completed the recognition and the production tasks, verbalized a rule with uncertainty. The following examples are taken from the post-exposure questionnaire (5):

(5)  
  a. Ri and ru are paired while ju and ji are paired together to refer to things pertaining to people. This was evident in words such as ji esposo and ju esposa. Although I am still not certain about how the articles work entirely. I became more aware when taking the quizzes after the training sessions.  
  b. Ji is also used for masculine. During the tests I used ji instead of ri when the article was for a person or living being. I may have made this up completely.’”
Mass versus Count

Analysis of the count vs. mass distinction offers a completely different perspective. The participants failed to comment on the count or mass rule that underlay the determiner system. Only one participant reported a partial rule at the concurrent stage while completing the post exposure assessment tasks. Despite having followed a rule formation approach – as some of the participants that became aware in the animate vs. non-animate group – the comments made at the concurrent stage reflected the dismay of not being able to discover the rule; as a result, most of the remarks commented on those words that had previously appeared in the reading texts. Participants exposed to the count vs. mass distinction were classified according to (a) those participants whose comments consisted of gender as the only rationale for selecting the determiners, and (b) those that commented upon a certain rule that they could not grasp or consistently evaluated the items in at least two occasions. The former group consists of the Unaware participants, while the latter represents the Aware participants. As it was mentioned previously, this study does not address levels of unaware participants in this Aware group; examples for the former group are in (6) and for the latter in (7):

(6) a. Escuela is feminine, so I guess pu.
   Familia is feminine, so it gets pu.
   Presente? Pi.
   Amiga; pu.
Solidaridad; feminine.
b. El jugo; let’s say pi jugo.
   Esposo; fi esposo.
   El carro; fi carro.
   La riqueza; fu riqueza.
   I’m just making up articles.
Another participant:
   I’m going to say that it’s fu educación because it’s feminine.

(7) a. Esposo. I remember reading it.
   Leche? Feminine; pu or fu. I can’t remember.
   Padre; Fi. I remember seeing this one.
   Riqueza; fu, perhaps? I don’t remember it.
b. That was masculine so…
   That was masculine too.
   La mujer. I remember this one. Feminine.
   I think that there’s a reason why pi or fi but I don’t know the rule, so
   I’m going to keep going.
   Infancia. I saw this one.
   Vaso? It’s new, isn’t?
To address the research questions, which are repeated below, the following statistical analyses were conducted. To answer RQ # 1, that is,

1) Do aware advanced L2 learners perform significantly better than unaware learners, as measured on recognition and production assessment tasks? If so, will the effects maintained over a one week period?

To test whether aware participants performed better than unaware participants and whether those effects could be maintained over a one week period the scores were submitted to a $2 \times 2$ repeated measures ANOVA, with group (aware vs. unaware) as the between subject variable and time (immediate posttest and delayed posttest) as the within-subject variable.

To answer RQ # (2), that is,

2) Is it possible to learn form-meaning connections without awareness (at the encoding stage) above chance level? If so, do unaware and aware participants show learning of the undisclosed features besides gender?

To address the first half of the research question, a series of one-sample $t$ test were run with the test value set at 50% for the recognition and production assessment tasks. The analysis was also conducted between old and new items.
To answer the second part of the research question, the scores obtained on the recognition and the production tasks were submitted independently for old and new items to individual 2 × 2 repeated-measures ANOVAs with type of feature animacy – for the animate vs. inanimate – or count – for the count vs. mass – (correct or incorrect) and gender (correct and incorrect) as within-subject factors.

To answer RQ # (3), that is,

3) Does type of features [+/- animacy] and [+/- count] have a differential effect on the level of awareness reported by advanced L2 learners during the reading task?

The scores of the recognition and the production tasks of the aware and unaware participants analyzed by type of feature were submitted to a two different 2 × 2 repeated measures ANOVA, with group (aware vs. unaware) as the between-subject variable, and type of feature: one for (animate vs. inanimate) and another for (count vs. mass) as the within-subject variable.

To answer RQ # (4), that is,

4) Is type of noun ending (–o and –a) and (–e and –n) errors related to the features [+/- animacy] and [+/-count]?

A series of two-way group-independence chi-square tests were conducted, one for aware and unaware participants as a group (composite score), and
another two that measured that relationship according to (un)awareness: one for aware participants and a different one for unaware participants.

To answer RQ # (5), that is,

5) Is type of gender feature [masculine] and [feminine] errors related to the features [+/- count] and [+/- animacy]?

A series of two-way group-independence chi-square tests were conducted, one for aware and unaware participants as a group (composite score), and another two that measured that relationship according to (un)awareness: one for aware participants and a different one for unaware participants.

To answer RQ # (6), that is,

6) Does the type of ending canonical (–o or –a) and arbitrary (–e and –n) differ on (a) the recognition and (b) production task?

Paired-sample t tests were used to analyze the results of the recognition and the production tests according to type of endings canonical ending (–o or –a) and arbitrary ending (–e or –n) to test whether those scores differed statistically.
CHAPTER 4: RESULTS

Introduction

This chapter presents the results of the statistical analyses computed to answer the research questions proposed in the previous chapter. The results will be presented in the order of the experimental conditions, namely, (1) animate vs. inanimate, and (2) count vs. mass; and in the order of the assessment tasks: recognition test (multiple-choice) and controlled production test (fill-in-the-blank).

The statistical analyses were conducted using the Statistical Package for Social Sciences (SPSS) software version 20.0. The exact statistical test used is mentioned at the beginning of each research question.

Reliability analysis

Reliability analyses were performed on all the materials and tests prior to conducting statistical analyses to address the research questions. The reading materials, which were part of the treatment condition, were submitted to a complexity test to make sure that all the texts had a similar level of complexity in order to control for text complexity. The results were similar for all the four texts, 80 and 78 for the animate vs. inanimate condition, and 79 and 80 for the mass vs. count condition (Fernández-Huerta, 1959). The controlled tests (comprehension questionnaire, multiple choice, fill-in-the blank, and gender questionnaire) were
also analyzed to verify the internal consistency of the assessment measures. The reliability coefficients were computed using Cronbach’s alpha, which measures how well items can be considered to be measuring the same construct (Larson-Hall, 2011). The idea is that if a test is reliable any difference in the scores is due to the person that is taking the test. An $\alpha$ value of .70 or greater is considered to be an acceptable reliability coefficient.

Cronbach’s $\alpha$ was found to be acceptable on all the tests. (1) Animate vs. inanimate: comprehension questionnaire 1, .74; comprehension questionnaire 2, .77; multiple-choice (recognition test), .73; fill-in-the-blank (controlled production test) test, .84; and (2) count vs. mass: comprehension questionnaire 1, .72; comprehension questionnaire 2, .78; multiple-choice (recognition test), .82; fill-in-the-blank (controlled production test) test, .79.

Before the results of the statistical analyses are presented, a table summarizing the number of participants per cell of all the two experimental conditions is presented in Table 20.

Table 20

<table>
<thead>
<tr>
<th>Number of participants per experimental group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Determiner system</td>
</tr>
<tr>
<td>-------------------</td>
</tr>
<tr>
<td>Unaware participants</td>
</tr>
<tr>
<td>$n = 26$</td>
</tr>
</tbody>
</table>
Quantitative analyses

Research Question 1

In order to address the research question one (Do aware advanced L2 learners perform significantly better than unaware learners, as measured on a recognition and a production assessment tests? If so, will the effects maintained over a one week period?), the scores were submitted to a $2 \times 2$ repeated measures ANOVA, with Group (aware vs. unaware) as the between-subject variable and Time (immediate posttest vs. delayed posttest) as the within-subject variable. Two separate repeated measures ANOVAs were conducted, one for old items and another for new items.

The results are presented in the following order: (1) animate vs. inanimate experimental group and (2) count vs. mass experimental group. The order of the assessment tasks is (1) recognition test (multiple-choice) and (2) controlled production test (fill-in-the blank).

Animate vs. Inanimate

Multiple-choice (as recognition task)

To test the distribution of the data a boxplot was produced according to the results of the post-test and delayed post-test (Figure 5). Both groups (aware and unaware) display non-normality because there are outliers, most of the boxes are not symmetrical around their medians, and there are not equal-length tails on
both ends of the boxes. We see that the aware group performed better than the unaware group although there are outliers in the unaware group that performed similarly to the aware group.

Figure 5

Boxplot of recognition task across time: Animacy

Third boxplot: MC Old items Fourth boxplot: MC 2nd Sess. Old items
Fifth boxplot: MC New items Sixth boxplot MC 2nd Sess New items

The descriptive statistics and the results from the repeated measures ANOVA are shown in Table 21 and Table 22, respectively.
Table 21

Descriptive statistics for recognition task across time: Animacy

<table>
<thead>
<tr>
<th>Variable</th>
<th>Immediate Unaware</th>
<th>Posttest Aware</th>
<th>Delayed Unaware</th>
<th>Posttest Aware</th>
</tr>
</thead>
<tbody>
<tr>
<td>Old items</td>
<td>57.7 (15.97)</td>
<td>88.09 (10.6)</td>
<td>50.64 (19)</td>
<td>94.04 (6.3)</td>
</tr>
<tr>
<td>New items</td>
<td>41.67 (11.05)</td>
<td>77.38 (25.32)</td>
<td>46.8 (16.68)</td>
<td>84.52 (11.2)</td>
</tr>
</tbody>
</table>

*M (SD) Unaware N = 26, Aware N = 7

Table 22

Repeated measures ANOVA results for the recognition task: Animacy

<table>
<thead>
<tr>
<th>Source of variability</th>
<th>d</th>
<th>Sums of squares</th>
<th>Mean Squares</th>
<th>F</th>
<th>P</th>
<th>η²</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>MC Old Time</td>
<td>1</td>
<td>3.33</td>
<td>3.33</td>
<td>.013</td>
<td>.910</td>
<td>.00</td>
<td>.051</td>
</tr>
<tr>
<td>Group</td>
<td>1</td>
<td>15023.04</td>
<td>15023.04</td>
<td>56.15</td>
<td>.000</td>
<td>.644</td>
<td>1</td>
</tr>
<tr>
<td>Time × Group</td>
<td>1</td>
<td>466.26</td>
<td>466.26</td>
<td>1.80</td>
<td>.189</td>
<td>.055</td>
<td>.256</td>
</tr>
<tr>
<td>MC New Time</td>
<td>1</td>
<td>415.27</td>
<td>415.27</td>
<td>1.43</td>
<td>.241</td>
<td>.044</td>
<td>.213</td>
</tr>
<tr>
<td>Group</td>
<td>1</td>
<td>14873.91</td>
<td>14873.91</td>
<td>82.04</td>
<td>.000</td>
<td>.726</td>
<td>1</td>
</tr>
<tr>
<td>Time × Group</td>
<td>1</td>
<td>11.20</td>
<td>11.20</td>
<td>.039</td>
<td>.846</td>
<td>.001</td>
<td>.054</td>
</tr>
</tbody>
</table>

Old items
Descriptive statistics of old items – both on the immediate and the delayed post-tests – are provided in Table 21 (cf. above) and presented graphically in Figure 6.

Figure 6

*Plot of the recognition task across time by groups (old items): Animacy*

A main effect was found for Group in the case of old items (F 1, 31 = 56.15, p < .001, partial η² = .644, power = 1), but not for Time (F 1, 31 = .013, p > .91, partial η² = .00, power = .051) nor was there a significant interaction of Group and Time (F 1, 31 = 1.8, p > .189, partial η² = .055, power = .256). The lack of main effect for Time indicates that the scores did not vary across time. To further investigate the main effects for Group, independent sample t-tests were run on both the immediate and the delayed post-tests. Results from the independent sample t-test revealed a significant difference between the Aware and the Unaware group on the immediate post-test (t(31) = -5.98, p < .001), and the delayed post-tests (t(14.08) = -5.61, p < .001) (Table 22).

New items
Descriptive statistics of new items – both on the immediate and the delayed post-tests – are provided in Table 21 (cf. above) and presented graphically in Figure 7.

Figure 7

*Plot of the recognition task across time by groups (new items): Animacy*

![Plot of the recognition task across time by groups (new items): Animacy](image)

A main effect was found for Group in the case of the new items (F 1, 31 = 82.04, p > .241, partial \( \eta^2 = .044 \), power = .213), but not for Time (F 1, 31 = 1.43, \( p < .001 \), partial \( \eta^2 = .79 \), power = 1) nor was there a significant interaction of Group and Time (F 1, 31 = 0.39, p > .846, partial \( \eta^2 = .001 \), power = .054). The no main effect for time indicates that the scores did not vary across time. To further investigate the Group difference, independent sample \( t \)-tests were run on both the immediate and the delayed post-tests. Results revealed a significant difference between the Aware and the Unaware group on both the immediate post-test (\( t(31) = -5.62, p < .001 \)), and the delayed post-tests (\( t(14.08) = -5.61, p < .001 \)) (Table 22).

**Fill-in-the-blank (as controlled production task)**
The descriptive statistics are found in Table 23 and the ANOVA results in Table 24.

Table 23

*Descriptive statistics for production task across time: Animacy*

<table>
<thead>
<tr>
<th>Variable</th>
<th>Immediate Unaware</th>
<th>Posttest Aware</th>
<th>Delayed Unaware</th>
<th>Posttest Aware</th>
</tr>
</thead>
<tbody>
<tr>
<td>Old items</td>
<td>54.8 (13.78)</td>
<td>73.8 (26.54)</td>
<td>53.84 (14.57)</td>
<td>83.33 (16.67)</td>
</tr>
<tr>
<td>New items</td>
<td>48.07 (16.56)</td>
<td>79.76 (17.9)</td>
<td>45.19 (16.01)</td>
<td>84.52 (18.9)</td>
</tr>
</tbody>
</table>

*M (SD) Unaware N = 26, Aware N = 7*
Repeated measures ANOVA results for the production task: Animacy

<table>
<thead>
<tr>
<th>Source of variability</th>
<th>df</th>
<th>Sums of squares</th>
<th>Mean Squares</th>
<th>F</th>
<th>P</th>
<th>η²</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>FIB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>1</td>
<td>99.97</td>
<td>99.97</td>
<td>.485</td>
<td>.491</td>
<td>.015</td>
<td>.104</td>
</tr>
<tr>
<td>Group</td>
<td>1</td>
<td>10004.7</td>
<td>10004.7</td>
<td>63.31</td>
<td>.000</td>
<td>.671</td>
<td>1</td>
</tr>
<tr>
<td>Time × Group</td>
<td>1</td>
<td>188.35</td>
<td>133.37</td>
<td>.913</td>
<td>.347</td>
<td>.029</td>
<td>.153</td>
</tr>
<tr>
<td>FIB Old</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>1</td>
<td>202.19</td>
<td>202.19</td>
<td>.637</td>
<td>.431</td>
<td>.020</td>
<td>.121</td>
</tr>
<tr>
<td>Group</td>
<td>1</td>
<td>6483.44</td>
<td>6483.44</td>
<td>32.92</td>
<td>.000</td>
<td>.515</td>
<td>1</td>
</tr>
<tr>
<td>Time × Group</td>
<td>1</td>
<td>303.202</td>
<td>303.202</td>
<td>.955</td>
<td>.336</td>
<td>.03</td>
<td>.157</td>
</tr>
<tr>
<td>FIB New</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>1</td>
<td>9.71</td>
<td>9.71</td>
<td>.032</td>
<td>.858</td>
<td>.001</td>
<td>.053</td>
</tr>
<tr>
<td>Group</td>
<td>1</td>
<td>13907.58</td>
<td>13907.58</td>
<td>53.64</td>
<td>.000</td>
<td>.634</td>
<td>1</td>
</tr>
<tr>
<td>Time × Group</td>
<td>1</td>
<td>161.21</td>
<td>161.21</td>
<td>.538</td>
<td>.469</td>
<td>.017</td>
<td>.11</td>
</tr>
</tbody>
</table>

Old items

Descriptive statistics of old items – both on the immediate and the delayed post-tests – are provided in Table 23 (cf. above) and presented graphically in Figure 8.
A main effect was found for Group in the case of old items (F 1, 31 = 32.92, p < .001, partial η2 = .51, power = 1), but not for Time (F 1, 31 = .637, p > .431, partial η2 = .02, power = .121) nor was there a significant interaction of Time and Group (F 1, 31 = .955, p > .336, partial η2 = .79, power = .157). The lack of a main effect for Time indicates that the scores did not significantly vary across time. To further analyze the main effect for Group, independent sample t tests were run on both the immediate post-test and the delayed post-test. Results from the independent sample t test showed a significant difference between the Aware and the Unaware group on the immediate post-test (t(31) = -2.62, p = .013), and the delayed post-test (t(31) = -4.61, p < .001) (Table 24).
Descriptive statistics of new items – both in the immediate and the delayed post-tests – are provided in the Table 23 (cf. above) and presented graphically in Figure 9.

**Figure 9**
*Plot of the production task across time by groups (new items): Animacy*

A main effect was found for Group in the case of new items (F 1, 31 = 53.64, p < .001, partial η2 = .63, power = 1) but not for Time (F 1, 31 = .032, p > .858, partial η2 = .001, power = .053) nor was there a significant interaction of Time and Group (F 1, 31 = .538, p > .469, partial η2 = .017, power = .11). This indicates that the variation in scores between the groups did not vary across time.

To further investigate the main effect for Group, independent sample t tests were run on both the immediate and the delayed post-test. Results from the immediate post-test showed that the differences between the Aware and the Unaware group on these two tests were both significant: (t(8.96) = -4.42, p < .001) and (t(8.47) = -5.04, p < .001), for the immediate and the delayed post-test, respectively (Table 24).
Therefore, it can be concluded that aware participants did only slightly better between the immediate posttest and the delayed posttest; however this difference was not statistically significant. However, it is noteworthy that aware participants did not present the customary decrease between the immediate posttest and the delayed posttest, whereas the unaware participants did.

In summary, Animate vs. Inanimate participants performed differentially according to their category of awareness. Aware participants performed significantly better than unaware participants on both tasks – recognition and production – as well as in the case of old and new items for both tasks. In the case of the second part of the Research Question 1 exploring the effects over a one week period, it was found that aware participants maintained the effects over this time period. Unaware participants performed at chance level on the immediate post-test and on the delayed post-test.

**Count vs. Mass**

The resulting groups consisted of 18 unaware participants, and 11 aware participants, most of whom reached a low level of awareness.

**Multiple-choice (as recognition task)**
To graphically see the distribution of the data a boxplot was produced containing the comparison between the results of the immediate post-test and the delayed post-test. Although the group distribution is not normal for both groups, an examination of the boxplot shows that the median of the scores tends to be similar in both groups with the exception of old items.

Figure 10

*Boxplot of recognition task across time: Count*

![Boxplot of recognition task across time: Count](image)

The descriptive statistics of the multiple-choice and the results of the repeated measures ANOVA are found in Table 25 and Table 26, respectively.

Table 25
Descriptive statistics for recognition task across time: Count

<table>
<thead>
<tr>
<th>Variable</th>
<th>Immediate</th>
<th>Posttest</th>
<th>Delayed</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Unaware</td>
<td>Aware</td>
<td>Unaware</td>
<td>Aware</td>
</tr>
<tr>
<td>Old items</td>
<td>46.29</td>
<td>65.15</td>
<td>45.83</td>
<td>53.03</td>
</tr>
<tr>
<td></td>
<td>(11.51)</td>
<td>(14.34)</td>
<td>(12.21)</td>
<td>(12.51)</td>
</tr>
<tr>
<td>New items</td>
<td>52.31</td>
<td>50.0</td>
<td>44.89</td>
<td>43.18</td>
</tr>
<tr>
<td></td>
<td>(11.71)</td>
<td>(13.43)</td>
<td>(11.12)</td>
<td>(13.34)</td>
</tr>
</tbody>
</table>

M (SD) Unaware N = 18, Aware N = 11

Table 26

Repeated measures ANOVA results for the recognition task: Count

<table>
<thead>
<tr>
<th>Source of variability</th>
<th>df</th>
<th>Sums of squares</th>
<th>Mean Squares</th>
<th>F</th>
<th>P</th>
<th>η²</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>MC Old</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>1</td>
<td>540.61</td>
<td>540.61</td>
<td>2.92</td>
<td>.099</td>
<td>.098</td>
<td>.378</td>
</tr>
<tr>
<td>Group</td>
<td>1</td>
<td>2316.99</td>
<td>2316.99</td>
<td>18.27</td>
<td>.000</td>
<td>.404</td>
<td>.985</td>
</tr>
<tr>
<td>Time × Group</td>
<td>1</td>
<td>466.26</td>
<td>466.26</td>
<td>1.80</td>
<td>.189</td>
<td>.055</td>
<td>.333</td>
</tr>
<tr>
<td>MC New</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>1</td>
<td>690.84</td>
<td>690.84</td>
<td>5.001</td>
<td>.034</td>
<td>.156</td>
<td>.578</td>
</tr>
<tr>
<td>Group</td>
<td>1</td>
<td>55.73</td>
<td>55.73</td>
<td>.351</td>
<td>.559</td>
<td>.013</td>
<td>.088</td>
</tr>
<tr>
<td>Time × Group</td>
<td>1</td>
<td>1.185</td>
<td>1.185</td>
<td>.009</td>
<td>.927</td>
<td>.000</td>
<td>.051</td>
</tr>
</tbody>
</table>

Unaware N = 18, Aware N = 11

Old items
Descriptive statistics of old items –both on the immediate and the delayed post-tests– are provided in Table 25 (cf. above) and presented graphically in Figure 11.

Figure 11

*Plot of the recognition task across time by groups (old items): Count*

A main effect was found for Group (F 1, 27 = 18.27 \( p = .000 \), partial \( \eta^2 = .404 \), power = .985) but not for Time (F 1, 27 = 2.92, \( p = .099 \), partial \( \eta^2 = .098 \), power = .378) nor was there a significant Interaction of Time and Group (F 1, 27 = 2.51, \( p = .125 \), partial \( \eta^2 = .085 \), power = .333). The lack of main effect for Time indicates that the scores did not vary across time. To further investigate the main effect for Group, independent sample \( t \) tests were run between the aware and the unaware groups on both the immediate and the delayed post-tests. Results from the independent sample-test revealed a significant difference between the aware and the unaware group on the immediate post-test (\( t(17.79) = -3.69, p = .002 \)) but not on the delayed post-test (\( t(20.86) = -1.51, p = .144 \)) (Table 26).
New items

Descriptive statistics of new items – both on the immediate and the delayed post-tests – are provided in Table 31 (cf. above) and presented graphically in Figure 12.

Figure 12

Plot of the recognition task across time by groups (new items): Count

A main effect in this case was found for Time (F 1, 27 = 5.001, p = .034, partial $\eta^2 = .156$, power = .578) and not for Group (F 1, 27 = .351, p = .559, partial $\eta^2 = .013$, power = .088) nor was there a significant Interaction of Time and Group (F 1, 27 = .009, p = .927, partial $\eta^2 = .000$, power = .051). The main effect for Time indicates that the scores vary across time, while the lack of a main effect for Group indicates that both groups performed similarly. An examination of the descriptive statistics (and the graphic representation of those statistics in Figure 12) evidenced that both the aware and the unaware groups drop steeply
between the immediate post-test and the delayed post-test. To further examine the main effect for Time and investigate this difference, paired samples $t$ tests were run for the aware and the unaware groups between the immediate and the delayed posttest. The results evidenced that the score difference was not significant for the aware participants ($t(10)= 1.39, p = .192$) nor for the unaware participants ($t(10)= 1.86, p = .08$) (Table 26).

**Fill-in-the-blank (as controlled production task)**

The descriptive statistics of the fill-in-the-blank test and the results of the repeated measures ANOVA are found in Table 27 and Table 28, respectively.

**Table 27**  
*Descriptive statistics for production task across time: Count*

<table>
<thead>
<tr>
<th>Variable</th>
<th>Immediate Unaware</th>
<th>Posttest Aware</th>
<th>Delayed Unaware</th>
<th>Posttest Aware</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Multiple-choice</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>46.76 (9.43)</td>
<td>54.16 (14.9)</td>
<td>49.07 (11.21)</td>
<td>48.1 (12.68)</td>
</tr>
<tr>
<td></td>
<td>Old items</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>48.61 (12.21)</td>
<td>45.37 (9.99)</td>
<td>50.92 (12.08)</td>
<td>45.45</td>
</tr>
<tr>
<td></td>
<td>New items</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>53.03 (20.16)</td>
<td>53.79 (19.49)</td>
<td>51.38 (16.48)</td>
<td>49.24</td>
</tr>
</tbody>
</table>

$M (SD)$ Unaware $N = 18$, Aware $N = 11$

**Table 28**  
*Repeated measures ANOVA results for the recognition task: Count*

<table>
<thead>
<tr>
<th>Source of</th>
<th>$Df$</th>
<th>Sums of Mean</th>
<th>$F$</th>
<th>$P$</th>
<th>$\eta^2$</th>
<th>Power</th>
</tr>
</thead>
</table>

167
<table>
<thead>
<tr>
<th>variability</th>
<th>squares</th>
<th>Squares</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>FIB</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>1</td>
<td>47.88</td>
</tr>
<tr>
<td>Group</td>
<td>1</td>
<td>141.556</td>
</tr>
<tr>
<td>Time × Group</td>
<td>1</td>
<td>239.47</td>
</tr>
<tr>
<td><strong>FIB Old</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>1</td>
<td>94.48</td>
</tr>
<tr>
<td>Group</td>
<td>1</td>
<td>3.779</td>
</tr>
<tr>
<td>Time × Group</td>
<td>1</td>
<td>333.94</td>
</tr>
<tr>
<td><strong>FIB New</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>1</td>
<td>7.408</td>
</tr>
<tr>
<td>Group</td>
<td>1</td>
<td>134.251</td>
</tr>
<tr>
<td>Time × Group</td>
<td>1</td>
<td>380.971</td>
</tr>
</tbody>
</table>

**Old Items**

Descriptive statistics of old items –both on the immediate and the delayed post-tests– are provided in Table 27 and presented graphically in Figure 13.

Figure 13
*Plot of the production task across time by groups (old items): Count*
Main effects were not found for Time (F 1, 27 = 0.366, p = .55, partial η² = .013, power = .09), Group (F 1, 27 = .968, p = .334, partial η² = .035, power = .158) nor was there a significant interaction of Time and Group (F 1, 27 = 1.829, p = .187, partial η² = .063, power = .257) (Table 28).

New items

Descriptive statistics of old items –both on the immediate and the delayed post-tests– are provided in Table 27 and presented graphically in Figure 14.

Figure 14
*Plot of the production task across time by groups (new items): Count*
Main effects were not found for Time (F 1, 27 = .042, \( p = .84 \), partial \( \eta^2 = .02 \), power = .054) nor for Group (F 1, 27 = .536, \( p = .470 \), partial \( \eta^2 = .019 \), power = .109) nor was there a significant interaction of Time and Group (F 1, 27 = .141, \( p = .155 \), partial \( \eta^2 = .073 \), power = .292) (Table 28).

Count vs. Mass participants performed differently than Animate vs. Inanimate participants. Aware participants in this experimental group performed significantly better than unaware participants only in the case of the Recognition task on the old items but not on new items. Aware participants also significantly decreased their recognition score from the immediate post-test to delayed post-test in the case of the new items. In the controlled production task, aware and unaware participants performed similarly in the case of old and new items.

**Research Question 2**

**Animate vs. Inanimate: First part of the RQ2**

In order to answer the first part of research question 2 (Is it possible to learn form-meaning connections without awareness (at the encoding stage) above chance level?), the assessments scores were submitted to a series of one-sample \( t \) test with the test value set at 50%. Unaware participants reached significance only in the case of old items (\( M = 57.7, p = .02, d = .30 \)) but with a low effect size and only on the recognition task (multiple-choice) (Table 29). The significance in the case of the new items was negative (\( M = 41.67, p < .001, d = .32 \)). Aware
participants performed significantly above chance on all the variables with a large
effect size (Table 30).

Table 29
One-sample t test summary for Unaware group: Animacy

<table>
<thead>
<tr>
<th>Variable</th>
<th>95% CI</th>
<th>M (SD)</th>
<th>t-value</th>
<th>p-value</th>
<th>Df</th>
<th>Effect size d</th>
</tr>
</thead>
<tbody>
<tr>
<td>MC Old items</td>
<td>1.24, 14.14</td>
<td>57.7 (15.97)</td>
<td>2.46</td>
<td>0.021</td>
<td>25</td>
<td>.30</td>
</tr>
<tr>
<td>MC New items</td>
<td>-12.8, -3.87</td>
<td>41.67 (11.05)</td>
<td>-8.33</td>
<td>0.001</td>
<td>25</td>
<td>.32</td>
</tr>
<tr>
<td>FIB</td>
<td>-4.03, 5.63</td>
<td>50.80 (11.96)</td>
<td>.342</td>
<td>.736</td>
<td>25</td>
<td>.30</td>
</tr>
<tr>
<td>FIB Old Items</td>
<td>-.76, 10.37</td>
<td>54.8 (13.77)</td>
<td>1.78</td>
<td>.07</td>
<td>25</td>
<td>.18</td>
</tr>
<tr>
<td>FIB New Items</td>
<td>-9.66, 4.30</td>
<td>48.08 (16.55)</td>
<td>-.592</td>
<td>.559</td>
<td>25</td>
<td>.07</td>
</tr>
</tbody>
</table>

Unaware N = 26
Table 30

One-sample t test summary for the Aware group: Animacy

<table>
<thead>
<tr>
<th>Variable</th>
<th>95% CI</th>
<th>M (SD)</th>
<th>t-value</th>
<th>p-value</th>
<th>Df</th>
<th>Effect size d</th>
</tr>
</thead>
<tbody>
<tr>
<td>MC Old items</td>
<td>28.28, 47.9</td>
<td>88.09 (10.6)</td>
<td>9.5</td>
<td>0.000</td>
<td>6</td>
<td>1.46</td>
</tr>
<tr>
<td>MC New items</td>
<td>3.95, 50.80</td>
<td>77.38 (25.32)</td>
<td>2.86</td>
<td>0.03</td>
<td>6</td>
<td>1.05</td>
</tr>
<tr>
<td>FIB</td>
<td>7.4, 46.16</td>
<td>75.52 (19.72)</td>
<td>3.38</td>
<td>.008</td>
<td>6</td>
<td>.99</td>
</tr>
<tr>
<td>FIB Old Items</td>
<td>-.73, 48.35</td>
<td>72.91 (24.70)</td>
<td>2.38</td>
<td>.05</td>
<td>6</td>
<td>.88</td>
</tr>
<tr>
<td>FIB New Items</td>
<td>13.19, 46.32</td>
<td>79.76 (17.9)</td>
<td>4.4</td>
<td>.005</td>
<td>6</td>
<td>1.15</td>
</tr>
</tbody>
</table>

Aware N= 7

Second part of the RQ2: Multiple choice

To address the second part of research question 2, do unaware and aware participants show learning of the undisclosed features besides gender?, the scores of the recognition and the production task then were submitted independently for old and new items to a 2 × 2 repeated-measures ANOVAs with animacy (correct or incorrect) and gender (correct and incorrect) as within-subject factors. The responses (dependent variables) were coded as correct animacy and correct gender (CACG), correct animacy and incorrect gender (CAIG), incorrect animacy and correct gender (IACG), and incorrect animacy and incorrect gender (IAIG).

Means and standard deviations for the multiple-choice recognition task are presented in Table 31. Table 32 and Table 33 present the ANOVA results for type of items by aware and unaware participants, respectively.
Table 31
*Descriptive statistics for the Recognition Task by Aware and Unaware participants: Animacy*

| Responses | Aware Old items | | Aware New items | | Unaware Old items | | Unaware New items |
|-----------|-----------------|---|-----------------|---|-----------------|---|-----------------|---|
|           | M    | SD  | M    | SD  | M    | SD  | M    | SD  |
| CACG      | 10.57 | (1.27) | 9.28 | (3.03) | 6.84 | (1.8) | 4.96 | (1.4) |
| CAIG      | .286  | (.487) | .429 | (.534) | .23  | (.58) | .461 | (.71) |
| IACG      | 0.714 | (1.11) | 2.29 | (3.04) | 4.88 | (1.9) | 6.08 | (1.4) |
| IAIG      | .429  | (.786) | 0    | 0    | .039 | (.19) | .538 | (.706) |

Unaware N = 26, Aware N = 7

Table 32
*Repeated measures ANOVA results for the Unaware participants: Animacy*

|              | Old items | | New items | | | | | | |
|--------------|------------|---|------------|---|---|---|---|
|              | F     | P   | \(\eta^2\) | Pow | F     | P   | \(\eta^2\) | Pow |
| Animacy      | 8.78  | .007 | .26 | .813 | 3.98  | .057 | .137 | .48 |
| Gender       | 2341.88 | .000 | .989 | 1 | 648.77 | .000 | .963 | 1 |
| Animacy \times Gender | 5.61 | .026 | .183 | .624 | 2.73  | .111 | .098 | .355 |

Unaware N = 26

Table 33
*Repeated measures ANOVA results for the Aware participants: Animacy*

|              | Old items | | New items | | | | | | |
|--------------|------------|---|------------|---|---|---|---|
|              | F     | P   | \(\eta^2\) | Pow | F     | P   | \(\eta^2\) | Pow |
| Animacy      | | | | | | | | | |
Table 32 shows that for unaware participants, there were main effects for gender for both old \( (F_{1, 25} = 2341.88, p < .001, \text{partial } \eta^2 = .989, \text{power} = 1) \) and new \( (F_{1, 25} = 648.77, p = .000, \text{partial } \eta^2 = .963, \text{power} = 1) \) items. In the case of animacy, there were main effects only for old \( (F_{1, 25} = 8.78, p = .007, \text{partial } \eta^2 = .26, \text{power} = .813) \) but not for new items \( (F_{1, 25} = 3.98, p = .057, \text{partial } \eta^2 = .137, \text{power} = .48) \). No significant interactions of Gender and Animacy were found for both old \( (F_{1, 25} = 5.62, p = .026, \text{partial } \eta^2 = .183, \text{power} = .624) \) or new items \( (F_{1, 25} = 2.73, p = 0.111, \text{partial } \eta^2 = .098, \text{power} = .355) \).

In the case of aware participants, Table 33 shows that there were main effects for gender for both old items \( (F_{1, 6} = 216.15, p = .000, \text{partial } \eta^2 = .973, \text{power} = 1) \) and new items \( (F_{1, 6} = 760.5, p = .000, \text{partial } \eta^2 = .992, \text{power} = 1) \). In the case of animacy, main effects were also found for both old items \( (F_{1, 6} = 111.87, p = .001, \text{partial } \eta^2 = .949, \text{power} = 1) \) and new items \( (F_{1, 6} = 10.84, p = .02, \text{partial } \eta^2 = .644, \text{power} = .783) \). Significant interactions of animacy and gender were also found for old items \( (F_{1, 6} = 131.2, p = .001, \text{partial } \eta^2 = .956, \text{power} = 1) \) and for new items \( (F_{1, 6} = 8.18, p = .029, \text{partial } \eta^2 = .577, \text{power} = .666) \).
To test whether learning took place, the distinction between the means of responses CACG and IACG for both old and new items were compared in the case of the aware and unaware participants. Paired-sample $t$ tests show that in the case of aware participants a main effect was found for both old items ($t (6) = 11.89, p < .001$), and new items ($t (6) = 2.19, p = .021$). In the case of the unaware participants, a main effect was only found for old items, ($t (25) = 2.7, p = .012$), and not for new items, ($t (25) = -1.91, p = .067$). These results seem to indicate that for the recognition task only aware participants – and not unaware participants – were able to select the correct determiner based on animacy and gender for old and new items. Unaware participants achieved this only in the case of old items, which suggests that memory played a role.

Second part of the RQ 2: Fill-in-the-blank (as controlled production task)

To test whether aware and unaware participants were able to produce the correct determiner above chance in the production task a series of one-sample $t$ tests were conducted. Table 34 shows the means and standard deviations of aware and unaware participants. Table 35 and Table 36 present the results of the ANOVAs conducted for aware and unaware participants.

Table 34

| Descriptive statistics for the Production Task by Aware and Unaware participants: Animacy |
|---------------------------------|-----------------|-----------------|
|                                 | Aware           | Unaware         |
|                                 |                 |                 |

### Table 35

**Repeated measures ANOVA results for the Unaware participants: Animacy**

<table>
<thead>
<tr>
<th></th>
<th>Old items</th>
<th></th>
<th></th>
<th>New items</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Old items</td>
<td>New items</td>
<td></td>
<td>Old items</td>
<td>New items</td>
<td></td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Animacy</td>
<td>6.36</td>
<td>.016</td>
<td>.210</td>
<td>.69</td>
<td>.515</td>
<td>.48</td>
</tr>
<tr>
<td>Gender</td>
<td>1101.94</td>
<td>.000</td>
<td>.978</td>
<td>1</td>
<td>601.66</td>
<td>.000</td>
</tr>
<tr>
<td>Animacy × Gender</td>
<td>3.72</td>
<td>.065</td>
<td>.13</td>
<td>.46</td>
<td>0.04</td>
<td>.843</td>
</tr>
</tbody>
</table>

Unaware N = 26

### Table 36

**Repeated measures ANOVA results for the Aware participants: Animacy**

<table>
<thead>
<tr>
<th></th>
<th>Old items</th>
<th></th>
<th></th>
<th>New items</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Old items</td>
<td>New items</td>
<td></td>
<td>Old items</td>
<td>New items</td>
<td></td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Animacy</td>
<td>5.66</td>
<td>(2.06)</td>
<td>6.54</td>
<td>(1.65)</td>
<td>5.66</td>
<td>(2.06)</td>
</tr>
<tr>
<td>CAIG</td>
<td>0</td>
<td>(0)</td>
<td>.286</td>
<td>(.488)</td>
<td>.342</td>
<td>(.845)</td>
</tr>
<tr>
<td>IACG</td>
<td>3.00</td>
<td>(3.1)</td>
<td>2.14</td>
<td>(2.16)</td>
<td>5.03</td>
<td>(1.7)</td>
</tr>
<tr>
<td>IAIG</td>
<td>.142</td>
<td>(.38)</td>
<td>0</td>
<td>(0)</td>
<td>.77</td>
<td>(.271)</td>
</tr>
</tbody>
</table>

Unaware N = 26, Aware N = 7
For unaware participants, there were main effects for Gender ($F_{1, 25} = 1101.94, p < .001$, partial $\eta^2 = .978$, power = 1) and Animacy ($F_{1, 25} = 6.36, p = .016$, partial $\eta^2 = .21$, power = .69), but not a significant interaction for Gender and Animacy ($F_{1, 25} = 3.72, p = .065$, partial $\eta^2 = .13$, power = .46) for old items. However, in the case of new items, there was a main effect only for Gender ($F_{1, 25} = 601.66, p < .001$, partial $\eta^2 = .96$, power = 1) and not for animacy ($F_{1, 25} = .515, p = .48$, partial $\eta^2 = .02$, power = .106) and no significant interaction of gender and animacy ($F_{1, 25} = .04, p = .843$, partial $\eta^2 = .002$, power = .054) (Table 35).

In the case of the aware participants, there were main effects for Gender ($F_{1, 6} = 1681, p < .001$, partial $\eta^2 = .1$, power = 1), Animacy ($F_{1, 6} = 5.64, p < \ .001$, partial $\eta^2 = .486$, power = .513) and a significant interaction for Gender and Animacy ($F_{1, 6} = 216.15, p = .043$, partial $\eta^2 = .521$, power = .571) for old items. For new items, there were main effects for Gender ($F_{1, 6} = 960, p < .001$, partial $\eta^2 = .99$, power = 1), Animacy ($F_{1, 6} = 19.33, p = .003$, partial $\eta^2 = .795$, power = .978), and a significant interaction for Gender and Animacy ($F_{1, 6} = 19.33, p = .005$, partial $\eta^2 = .753$, power = .952) (Table 36).
The means of the CACG and IACG items for the production task were submitted to paired-sample $t$ tests to assess whether the difference between those means was significant, that is, whether participants based their choices not only on gender but on animacy as well. In the case of the aware participants, performances on both the new ($t(6) = 4.64, p = .004$) and the old items ($t(6) = 2.47, p = .049$) were significant. The scores of the unaware participants were significant only in the case of the old items ($t(25) = 2.35, p = .027$), but not in the case of the new items ($t(25) = .291, p = .773$).

**Count vs. Mass**

**First part of the RQ2**

To test whether those groups performed above chance on the recognition and production tasks (First part of the RQ2) a series of one-sample $t$ tests were conducted with the test level set at .05. None of the scores of the unaware group reached significance (Table 37). The aware group reached significance only in the case of the old items ($t(10) = 3.5 \ p = .006$) (Table 38).

<table>
<thead>
<tr>
<th>Variable</th>
<th>95% CI</th>
<th>$M \ (SD)$</th>
<th>t-value</th>
<th>p-value</th>
<th>df</th>
<th>Effect size $d$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Table 37</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**One-sample t test summary for the Unaware group: Count**
Table 38  
**One-sample t test summary for the Aware group: Count**

<table>
<thead>
<tr>
<th>Variable</th>
<th>95% CI</th>
<th>M (SD)</th>
<th>t-value</th>
<th>p-value</th>
<th>df</th>
<th>Effect size $d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>MC Old items</td>
<td>-9.42, 2.02</td>
<td>46.29 (11.51)</td>
<td>-1.36</td>
<td>.19</td>
<td>17</td>
<td>.2</td>
</tr>
<tr>
<td>MC New items</td>
<td>-3.51, 8.14</td>
<td>52.31 (11.71)</td>
<td>.838</td>
<td>.41</td>
<td>17</td>
<td>.12</td>
</tr>
<tr>
<td>FIB</td>
<td>-7.93, 1.49</td>
<td>46.75 (9.43)</td>
<td>-.192</td>
<td>.163</td>
<td>17</td>
<td>.18</td>
</tr>
<tr>
<td>FIB Old Items</td>
<td>-7.46, 4.68</td>
<td>48.61 (12.21)</td>
<td>-.100</td>
<td>.636</td>
<td>17</td>
<td>.07</td>
</tr>
<tr>
<td>FIB New Items</td>
<td>-9.59, 3.39</td>
<td>45.37 (9.99)</td>
<td>-1.97</td>
<td>.06</td>
<td>17</td>
<td>.25</td>
</tr>
</tbody>
</table>

Unaware $N = 18$

Second part of the RQ2  
Since the scores of the aware group reached significance in the case of the recognition test, the scores of this test were submitted for old and new items to a 2 × 2 repeated-measures ANOVA’s with count (correct and incorrect) and gender (correct and incorrect) as within-subjects factors. The responses (dependent variables) were coded as correct count and correct gender (CCCG), correct count and incorrect gender (CCIG), incorrect count and correct gender (ICCG), and incorrect count and incorrect gender (ICIG). Mean and standard deviations for the multiple-choice recognition task are presented in Table 39. Table 40 and Table 41
present the ANOVA results for the type of items and aware and unaware participants.

Table 39

*Descriptive statistics for the Recognition Task by Unaware and Aware participants: Count*

<table>
<thead>
<tr>
<th></th>
<th>Unaware</th>
<th>Aware</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Old items</td>
<td>M</td>
</tr>
<tr>
<td>Responses</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CACG</td>
<td>5.58</td>
<td>(1.43)</td>
</tr>
<tr>
<td>CAIG</td>
<td>842</td>
<td>(1.42)</td>
</tr>
<tr>
<td>IACG</td>
<td>5.05</td>
<td>(1.43)</td>
</tr>
<tr>
<td>IAIG</td>
<td>.473</td>
<td>(.69)</td>
</tr>
</tbody>
</table>

Unaware N = 18, Aware N = 11

Table 40

*Repeated measures ANOVA results for the Unaware participants: Count*

<table>
<thead>
<tr>
<th></th>
<th>Old items</th>
<th>New items</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F</td>
<td>P</td>
</tr>
<tr>
<td>Count</td>
<td>1.38</td>
<td>.255</td>
</tr>
<tr>
<td>Gender</td>
<td>151.35</td>
<td>.000</td>
</tr>
<tr>
<td>Count × Gender</td>
<td>0.56</td>
<td>.816</td>
</tr>
</tbody>
</table>

Unaware N = 18

Table 41

*Repeated measures ANOVA results for the Aware participants: Count*

<table>
<thead>
<tr>
<th></th>
<th>Old items</th>
<th>New items</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F</td>
<td>P</td>
</tr>
<tr>
<td>Count</td>
<td>14.83</td>
<td>.004</td>
</tr>
<tr>
<td>Gender</td>
<td>320.21</td>
<td>.000</td>
</tr>
<tr>
<td>Count × Gender</td>
<td>18.99</td>
<td>.002</td>
</tr>
</tbody>
</table>

Aware N = 11
In the case of the unaware participants (Table 40), repeated measures ANOVAs show that there was a main effect only for Gender in the case of the old items (F 1, 17 = 151.35, p < .001, partial $\eta^2 = .894$, power = 1) and for the new items (F 1, 17 = 651.23, p < .001, partial $\eta^2 = .971$, power = 1), but not for Count in the case of old items (F 1, 17 = 1.38, p = 0.255, partial $\eta^2 = .71$) nor new items (F 1, 17 = 2.32, p = 0.145, partial $\eta^2 = .114$). No significant interactions of Count and Gender in the case of old (F 1, 17 = 0.56, p = 0.816, partial $\eta^2 = .148$) and new items (F 1, 17 = .01, p = .018, partial $\eta^2 = .271$) were found.

Results from the aware group (Table 41) showed that there were main effects for old items in the case of Count (F 1, 10 = 14.83, p = .004, partial $\eta^2 = .622$, power = .927), Gender (F 1, 10 = 320.21, p < .001, partial $\eta^2 = .973$, power = 1), and there was a significant interaction of Count and Gender (F 1, 10 = 18.99, p = .002, partial $\eta^2 = .678$, power = 978). In the case of new items, there was a main effect only for Gender (F 1, 10 = 596.45, p < .001, partial $\eta^2 = .985$, power = 1) but not for Count (F 1, 10 =.783, p = .08, partial $\eta^2 = .971$, power = .125), nor was there a significant interaction of Count and Gender (F 1, 10 =.141, p = 0.716, partial $\eta^2 = .015$, power = .06).

To test whether learning took place, the distinction between the responses CCCG and IACG for old and new items were compared only for the aware group. Paired-sample $t$ tests show that for the old items there was a main effect ($t$(10) = 4.16, $p = .002$), but not for new items ($t$(10) = .620, $p = .551$), which means that
the apparently more engaging approach to the task that this group seemed to have adopted according to the analysis of the think-aloud group did not result sufficient enough to show learning.

Unaware participants evidenced main effects only for Gender (the disclosed feature) but not for Count (the undisclosed feature). This means that the treatment worked. Aware participants, on the other hand, evidenced main effects for Gender and Count and a significant interaction for Count and Gender only in the case of old items. In the case of the new items, this group performed precisely the same than the Unaware group, e.g., a main effect only for Gender, not a main effect for Count nor a significant interaction for Gender and Count. Type of linguistic item may be the main reason why aware participants did not behave as Aware participants in the Animate vs. Inanimate experimental condition.

In conclusion, participants in the animate vs. inanimate experimental group behaved differently according to aware versus unaware participants. Aware participants performed above chance on the recognition and production task for both old and new items. Unaware participants performed above chance only in the case of the old items on the less demanding task (recognition assessment task). Aware participants also evidenced learning when the responses were analyzed according to the disclosed feature (gender) and the undisclosed feature (animacy) for the old and new items as well as in the case of the recognition and production assessment tasks. Unaware participants, on the other hand, failed to evidence
learning and their scores resulted significant only for the disclosed and the undisclosed feature only in the case of old items and on the recognition task. It is also revealed that aware participants performed significant higher than unaware participants on both assessment tasks as well as for old and new items. According to the analyses, unaware participants performed statistically better only on those variables that require memory and a low level of processing: old items and the recognition assessment tasks.

In the case of the count versus mass experimental group, aware participants performed above chance only in the case of the old items and in the case of the recognition task. Unaware participants failed to perform above chance on any of the variables: old items, new items and recognition and production assessment tasks. The repeated measures ANOVA run to test learning for the aware group resulted significant only in the case of the old items for gender (the disclosed feature) and count (the undisclosed feature).

Aware and unaware participants in both experimental conditions performed differently and this may be the result of type of feature to which they were exposed.

**Research Question 3**

**Animate versus Inanimate**

To answer the research question 3 (Does type of feature [+/-animacy] and [+/-count] have a differential effect on multiple-choice (recognition task) and fill-
in-the-blank (production task), the results of the immediate both recognition and the production tasks were analyzed by type of feature and aware and unaware participants.

The scores were submitted to a 2 × 2 repeated measures ANOVA, with group (aware vs. unaware) as the between-subject variable, and animacy (animate vs. inanimate) as the within-subject variable.

Descriptively, the results evidenced that for the aware participants, they consistently got higher scores in the case of animate than in the non-animate nouns; unaware participants, on the other hand, only obtained higher scores for animate nouns for the fill-in-the-blank, while they obtained higher scores in non-animate nouns for the multiple-choice (Table 42). However, main effects were found only for Group in the multiple-choice (F 1, 31 = 44.56, p < .001, partial η² = .957, power 1) as well as on the fill-in-the-blank (F 1, 31 = 14.76, p < .001, partial η² = .323, power 1) and not for Animacy in either the recognition task (F 1, 31 = .142, p = .709, partial η² = .005, power .065) and the production task (F 1, 31 = 3.287, p = .08, partial η² = .096, power .42) (Table 43).

Table 42
Descriptive statistics for animate and inanimate responses according to the recognition and production task: Animacy

<table>
<thead>
<tr>
<th>Variable</th>
<th>Animate Unaware</th>
<th>Animate Aware</th>
<th>Inanimate Unaware</th>
<th>Inanimate Aware</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple-choice</td>
<td>43.91 (19.79)</td>
<td>85.71 (19.07)</td>
<td>54.80 (19.60)</td>
<td>79.77 (16.57)</td>
</tr>
</tbody>
</table>
Table 43

Repeated measures ANOVA results for the Recognition task and the Production task: Animacy

<table>
<thead>
<tr>
<th>Source of variability</th>
<th>Df</th>
<th>Sums of squares</th>
<th>Mean Squares</th>
<th>F</th>
<th>P</th>
<th>η²</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>MC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Animacy</td>
<td>1</td>
<td>67.43</td>
<td>67.43</td>
<td>.142</td>
<td>.709</td>
<td>.005</td>
<td>.065</td>
</tr>
<tr>
<td>Group</td>
<td>1</td>
<td>12289.59</td>
<td>12289.59</td>
<td>44.56</td>
<td>.000</td>
<td>.957</td>
<td>1</td>
</tr>
<tr>
<td>Animacy × Group</td>
<td>1</td>
<td>782.92</td>
<td>782.92</td>
<td>1.65</td>
<td>.208</td>
<td>.051</td>
<td>.238</td>
</tr>
<tr>
<td>FIB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Animacy</td>
<td>1</td>
<td>883.84</td>
<td>883.84</td>
<td>3.287</td>
<td>.080</td>
<td>.096</td>
<td>.420</td>
</tr>
<tr>
<td>Group</td>
<td>1</td>
<td>6073.93</td>
<td>6073.93</td>
<td>14.76</td>
<td>.001</td>
<td>.323</td>
<td>.961</td>
</tr>
<tr>
<td>Animacy × Group</td>
<td>1</td>
<td>193.61</td>
<td>193.61</td>
<td>.72</td>
<td>.403</td>
<td>.023</td>
<td>.13</td>
</tr>
</tbody>
</table>

Unaware N = 26, Aware N = 7

To further analyze the main effect for group, an independent sample t test was run to compare the means of the aware and unaware groups on the recognition assessment task. The difference between the aware and unaware groups was statistically significant in relation to the animate (t(9.78)= -5.10, p < .001) and inanimate responses (t(11.00)= 3.39, p = .006).

To further analyze the main effect for group, an independent sample t test was run to compare the means of the aware and unaware groups on the production assessment task. The difference between the aware and unaware groups was statistically significant in relation to the animate (t(9.17)= -2.38, p = .04) and the
inanimate responses ($t(7.6)= -2.84, p = .023$). Paired sample $t$ tests evidenced that a main effect was found for type of feature in the case of the unaware group ($t(25)= 2.65, p = .014$), but not for the aware group ($t(6)= 1.18, p = .28$). This means that the unaware group provided more animate ($M = 59.29, SD = 18.30$) than inanimate ($M = 46.15, SD = 16.70$) responses correctly and this difference was statistically significant. The aware group, on the other hand, also descriptively identified more animate ($M = 78.57, SD = 19.16$) than inanimate ($M = 73.81, SD = 24.26$) responses but this difference was no statistically significant.

### Count vs. Mass

The results for this experimental group showed that the unaware group identified more count than non-count nouns on the multiple-choice recognition task, but equally identified count and non-count nouns on the fill-in-the-blank test. In the case of the aware group, they identified more count nouns on the multiple-choice recognition task but it was the opposite in the case of the fill-in-the-blank task (Table 44).

Table 44

*Descriptive statistics for recognition task across time: Count*
<table>
<thead>
<tr>
<th>Variable</th>
<th>Count Unaware</th>
<th>Count Aware</th>
<th>Mass Unaware</th>
<th>Mass Aware</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple-choice</td>
<td>61.84 (19.70)</td>
<td>65.90 (17.66)</td>
<td>36.40 (18.26)</td>
<td>47.72 (16.70)</td>
</tr>
<tr>
<td>Fill-in-the-blank</td>
<td>49.56 (21.96)</td>
<td>47.72 (22.07)</td>
<td>48.25 (25.39)</td>
<td>53.03 (23.05)</td>
</tr>
</tbody>
</table>

* M (SD) Unaware N = 18, Aware N = 11

---

Table 45

Repeated measures ANOVA results for the Recognition and Production task:

**Count**

<table>
<thead>
<tr>
<th>Source of variability</th>
<th>df</th>
<th>Sums of squares</th>
<th>Mean Squares</th>
<th>F</th>
<th>P</th>
<th>η²</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>MC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Count</td>
<td>1</td>
<td>6627.88</td>
<td>6627.88</td>
<td>14.742</td>
<td>.001</td>
<td>.345</td>
<td>.959</td>
</tr>
<tr>
<td>Group</td>
<td>1</td>
<td>825.12</td>
<td>825.12</td>
<td>3.66</td>
<td>.066</td>
<td>.116</td>
<td>.455</td>
</tr>
<tr>
<td>Count × Group</td>
<td>1</td>
<td>183.44</td>
<td>183.44</td>
<td>.408</td>
<td>.528</td>
<td>.014</td>
<td>.095</td>
</tr>
<tr>
<td>FIB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Count</td>
<td>1</td>
<td>55.38</td>
<td>55.38</td>
<td>.069</td>
<td>.794</td>
<td>.002</td>
<td>.057</td>
</tr>
<tr>
<td>Group</td>
<td>1</td>
<td>55.37</td>
<td>55.37</td>
<td>.069</td>
<td>.748</td>
<td>.004</td>
<td>.061</td>
</tr>
<tr>
<td>Count × Group</td>
<td>1</td>
<td>152.60</td>
<td>152.60</td>
<td>.069</td>
<td>.794</td>
<td>.002</td>
<td>.071</td>
</tr>
</tbody>
</table>
A main effect was found only for type of feature count in the case of the multiple-choice test (F 1, 31 = 14.74, p < .001, partial $\eta^2$ = .345, power .959), but not for Group (F 1, 31 = 3.66, p = .066, partial $\eta^2$ = .116, power .455), nor was the interaction of Count and Group significant (F 1, 31 = .408, p = .528, partial $\eta^2$ = .014, power .095) (Table 45).

To further analyze the main effects for Count, a paired-sample $t$ test was run to compare the count and mass responses of the aware and the unaware participants. The results revealed that the aware group correctly provided more count responses ($M = 65.9, SD = 17.66$) than mass responses ($M = 47.72, SD = 16.7$) and this difference was statistically significant ($t(10) = 2.78, p = .019$). In the case of the unaware group, they also correctly identified more count ($M = 61.84, SD = 19.7$) than mass responses ($M = 36.4, SD = 19.16$), and this difference was also statistically significant ($t(17) = 3.28, p = .004$). Independent sample $t$ tests were run to compare group responses according to group. The results revealed that for count responses, main effects were not found for the aware and the unaware group the ($t(22.98) = -.582, p = .566$), neither for the mass responses ($t(22.61) = -1.73, p = .098$).

For the production task, main effects were not found for Group (F 1, 31 = .069, p = .748, partial $\eta^2$ = .004, power .061), or Count (F 1, 31 = .069, p = .794, partial $\eta^2$ = .002, power .057), nor was the interaction of Count and Group
significant (F 1, 31 = .069, p = .794, partial $\eta^2 = .002$, power .071) (Table 51). A paired sample $t$ test was run to compare the count and mass responses of the aware and unaware participants. The results revealed that the aware group responded correctly to more mass ($M = 53.03, SD = 23.05$) than count responses ($M = 47.72, SD = 22.07$), but this difference was not statistically significant ($t(10) = -.447, p = .664$). The unaware group provided correctly an equal number of count ($M = 49.56, SD = 21.96$) and mass ($M = 48.25, SD = 25.39$) responses, and therefore, this difference was not statistically significant ($t(17) = .142, p = .889$). Independent sample $t$ test revealed that both groups performed the same in regard to count ($t(20.92) = .22, p = .828$) and mass responses ($t(22.74) = -.528, p = .603$), and therefore, the difference was not statistically significant.

In conclusion, aware participants in the animate vs. inanimate experimental condition identified an equal number of animate and inanimate items on both assessment tasks. However, the unaware participants identified an equal number of animate and inanimate items only in the case of the recognition task, while they identified correctly more animate than inanimate items on the more demanding task, the production task.

Aware and unaware participants in the count vs. mass experimental group correctly identified more count than mass items in the case of the recognition task. Whereas for the production task aware and unaware participants correctly identified an equal number of count and mass items.
Research Question 4

Animate vs. Inanimate

To answer research question 4 (Is type of noun ending (–o and –a) and (–e and –n) errors related to the features [+/- animacy] and [+/-count]?) a series of two-way group-independence chi-square tests were conducted, one for aware and unaware participants as a group, and another two that measured that relationship according to (un)awareness, one for aware participants and a different one for unaware participants. Results of the recognition task will be presented first followed by the production task.

Multiple-choice (as recognition task)

Data were submitted to a two-way group-independence chi-square to address the relationship between gender errors canonical ending (–o and –a) and arbitrary ending (–e and –n) and animacy (with two levels [+ animate] [- animate]). The contingency tables for these data are shown below in Table 46, 47, and 48.

The composite score of aware and unaware participants was statistically significant ($\chi^2 = 43.45, df= 1, p < .001$) with a small effect size of .25 (Table 46).

Table 46
Contingency table for animacy and type of feature (Aware and unaware participants): Animate

<table>
<thead>
<tr>
<th>Group</th>
<th>Animate</th>
<th>Inanimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canonical</td>
<td>227</td>
<td>163</td>
</tr>
<tr>
<td>Arbitrary</td>
<td>94</td>
<td>194</td>
</tr>
</tbody>
</table>

The results of the unaware participants ($\chi^2 = 35.55$, $df = 1$, $p < .001$) and aware participants ($\chi^2 = 9.123$, $df = 1$, $p = .003$) were statistical significant both with a small effect size, being .25 and .24, respectively (Tables 47 and 48, respectively). These results have to be interpreted with caution since they all have a small effect size.

Table 47
Contingency table for animacy and type of feature (Unaware participants): Animate

<table>
<thead>
<tr>
<th>Group</th>
<th>Animate</th>
<th>Inanimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canonical</td>
<td>180</td>
<td>128</td>
</tr>
<tr>
<td>Arbitrary</td>
<td>74</td>
<td>150</td>
</tr>
</tbody>
</table>

Unaware $N = 26$

Table 48
Contingency table for animacy and type of feature (Aware participants): Animate

<table>
<thead>
<tr>
<th>Group</th>
<th>Animate</th>
<th>Inanimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canonical</td>
<td>47</td>
<td>35</td>
</tr>
<tr>
<td>Arbitrary</td>
<td>21</td>
<td>44</td>
</tr>
</tbody>
</table>

Aware $N = 7$
Fill-in-the blank (as controlled production task)

The results of the two-way group-independence chi-squares were statistically significant for the composite score of aware and unaware participants ($\chi^2 = 45.07, df = 1, p < .001, w = .25$) (Table 49), both for the unaware ($\chi^2 = 37.91, df = 1, p < .001, w = .25$) (Table 50), and the aware participants ($\chi^2 = 7.6, df = 1, p = .006, w = .21$) (Table 51). In summary, the consistent results of the two-way group-independence chi-square tests across assessment tests (recognition and production) as well as level of awareness confirm that there is a significant association between type of feature (animate and inanimate) and type of noun ending (canonical) and (arbitrary) errors.

Table 49
Contingency table for animacy and gender (Aware and unaware participants): Animate

<table>
<thead>
<tr>
<th>Group</th>
<th>Animate</th>
<th>Inanimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canonical</td>
<td>321</td>
<td>224</td>
</tr>
<tr>
<td>Arbitrary</td>
<td>65</td>
<td>142</td>
</tr>
</tbody>
</table>

Table 50
Contingency table for animacy and gender (Unaware participants): Animate

<table>
<thead>
<tr>
<th>Group</th>
<th>Animate</th>
<th>Inanimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canonical</td>
<td>253</td>
<td>175</td>
</tr>
<tr>
<td>Arbitrary</td>
<td>51</td>
<td>114</td>
</tr>
</tbody>
</table>

Unaware $N = 26$
Table 51

Contingency table for animacy and gender (Aware participants): Animate

<table>
<thead>
<tr>
<th>Group</th>
<th>Animate</th>
<th>Inanimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canonical</td>
<td>68</td>
<td>49</td>
</tr>
<tr>
<td>Arbitrary</td>
<td>14</td>
<td>28</td>
</tr>
</tbody>
</table>

Aware N = 7

Count vs. Mass

Multiple-choice (recognition task)

The data was submitted to a two-way group-independence chi-square to assess the relationship between gender errors canonical ending (−o and −a) and arbitrary ending (−e and −n) and animacy (with two levels (+ animate) (− animate)). The contingency tables for the recognition and the production tasks are shown below in Tables 58 and 59.

Fill-in-the blank (as controlled production task)

The results of the two-way group-independence chi-square were not significant for the recognition test ($\chi^2 = 3.37$, df = 1, $p = .066$, $w = .07$) (Table 52), but they were statistically significant for the controlled production test ($\chi^2 = 48.89$, df = 1, $p < .001$, $w = .27$) (Table 53).

Table 52

Recognition task: Contingency table for count and gender (Aware and unaware participants): Count
For the animate vs. inanimate experimental group, the results evidenced that there is a relationship between canonical ending (–o and –a) and arbitrary ending (–e and –n) errors, which means that participants committed more errors in the arbitrary ending than in canonical ending. However, for the count and mass experimental group, the same statistical analysis was significant only in the case of the controlled production test and not on the recognition test.

**Research Question 5**

**Animate vs. Inanimate**

To determine if the type of gender [masculine] and [feminine] errors relate to the features [+ animate] and [-animate] a series of two-way group-independence chi-square tests were conducted. To probe deeper, three chi-squares
were run, one for the composite score of aware and unaware participants, and another two that measured that relationship according to (un)awareness: one for aware participants and a different one for unaware participants. Results will be presented in the following order: Recognition and Production.

Multiple-choice (recognition task)

Data were submitted to a two-way group-independence chi-square to assess the relationship between animacy (with two levels [+ animate] [-animate]) and gender (with two levels [+ feminine] [-feminine]). The contingency table for these data is shown below in Tables 54, 55, and 56.

The composite score of aware and unaware participants was not statistically significant ($\chi^2 = .266, df = 1, p = .606$) with a categorical small effect size of .01 (Table 60).

Table 54

Contingency table for animacy and gender (Aware and unaware participants):

<table>
<thead>
<tr>
<th>Animacy</th>
<th>Group</th>
<th>Animate</th>
<th>Inanimate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Masculine</td>
<td>190</td>
<td>184</td>
</tr>
<tr>
<td></td>
<td>Femenine</td>
<td>196</td>
<td>176</td>
</tr>
</tbody>
</table>

The results for the unaware group were not statistically significant ($\chi^2 = .399, df = 1, p = .528$) with a very small effect size of .02 (Table 55).
Table 55

Contingency table for animacy and gender (Unaware participants): Animacy

<table>
<thead>
<tr>
<th>Group</th>
<th>Animate</th>
<th>Inanimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Masculine</td>
<td>151</td>
<td>148</td>
</tr>
<tr>
<td>Feminine</td>
<td>154</td>
<td>136</td>
</tr>
</tbody>
</table>

Unaware $N = 26$

The results for the aware group were not statistically significant either ($\chi^2 = .596$, $df = 1$, $p = .440$) with a very small effect size of .06 (Table 56).

Table 56

Contingency table for animacy and gender (Aware participants): Animacy

<table>
<thead>
<tr>
<th>Group</th>
<th>Animate</th>
<th>Inanimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Masculine</td>
<td>30</td>
<td>36</td>
</tr>
<tr>
<td>Feminine</td>
<td>42</td>
<td>39</td>
</tr>
</tbody>
</table>

Aware $N = 7$

The consistency of these three results shows that there is not a relationship between gender (masculine or feminine) and the animacy feature ([+/- animate]) as measured on a recognition task. Although the feminine inanimate cell has a smaller frequency, this is not statistically significant.

Fill-in-the blank (as controlled production task)
The same two-way group-independence chi-square was used to assess the relationship between animacy (with two levels [+ animate] [-animate]) and gender (with two levels [+ masculine] [-masculine]). The contingency table for these data is shown below in Tables 57, 58, and 59.

Table 57
*Contingency table for animacy and gender (Aware and unaware participants): Animacy*

<table>
<thead>
<tr>
<th>Group</th>
<th>Animate</th>
<th>Inanimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Masculine</td>
<td>191</td>
<td>190</td>
</tr>
<tr>
<td>Feminine</td>
<td>193</td>
<td>179</td>
</tr>
</tbody>
</table>

The results for the combined group of aware and unaware participants were not statistically significant ($\chi^2 = .231, df = 1, p = .631$) with a small effect size of .018 (Table 58). The results for the aware ($\chi^2 = .213, df = 1, p = .644$) and the unaware ($\chi^2 = .024, df = 1, p = .876$) participants were not statistically significant either (Table 64 and Table 59).

Table 58
*Contingency table for animacy and gender (Unaware participants): Animacy*

<table>
<thead>
<tr>
<th>Group</th>
<th>Animate</th>
<th>Inanimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Masculine</td>
<td>150</td>
<td>149</td>
</tr>
<tr>
<td>Feminine</td>
<td>151</td>
<td>139</td>
</tr>
</tbody>
</table>

Unaware $N = 26$

Table 59
*Contingency table for animacy and gender (Aware participants): Animacy*
### Group Animate Inanimate

<table>
<thead>
<tr>
<th>Group</th>
<th>Animate</th>
<th>Inanimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Masculine</td>
<td>41</td>
<td>41</td>
</tr>
<tr>
<td>Feminine</td>
<td>42</td>
<td>40</td>
</tr>
</tbody>
</table>

Aware $N = 7$

The consistency of these three results shows that there is not a relationship between gender (masculine or feminine) and the animacy feature ([+/- animate]) as measured on a controlled production task.

**Count vs. Mass**

To determine if the type of gender [masculine] and [feminine] has a differential effect on the features [+ animate] and [-animate] a series of two-way group-independence chi-square tests were conducted. Results of the composite score (aware and unaware participants) will be presented in the following order: Recognition and Production.

The data was submitted to a two-way group-independence chi-square to assess the relationship between count (with two levels [+ count] and [-count]) and gender (with two levels [+ masculine] and [-masculine]). The contingency tables for the recognition and the production tasks are shown below in Tables 60 and 61.

**Multiple-choice (as recognition task)**
The results of the two-way group-independence chi-square for the recognition ($\chi^2 = .120, df= 1, p = .729$) and production ($\chi^2 = .071, df= 1, p = .790$) tasks were not significant for either task with a marginalized effect size of .013, and 0.1, respectively (Table 60 and Table 61).

Table 60:

**Recognition task: Contingency table for count and gender (Aware and unaware participants): Count**

<table>
<thead>
<tr>
<th>Group</th>
<th>Animate</th>
<th>Inanimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Masculine</td>
<td>174</td>
<td>163</td>
</tr>
<tr>
<td>Feminine</td>
<td>169</td>
<td>167</td>
</tr>
</tbody>
</table>

**Fill-in-the blank (as controlled production task)**

Table 61

**Production task: Contingency table for count and gender: Count**

<table>
<thead>
<tr>
<th>Group</th>
<th>Animate</th>
<th>Inanimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Masculine</td>
<td>176</td>
<td>175</td>
</tr>
<tr>
<td>Feminine</td>
<td>178</td>
<td>170</td>
</tr>
</tbody>
</table>

The consistency of these results evidences that there is not a relationship between gender (masculine or feminine) and the animacy feature ($[+/- animate]$) as measured on both a recognition and controlled production task.

The results of Research Question 5 are the only results that are consistent for both experimental groups, that is, animate vs. inanimate as well as count vs. mass. The results were also the same for aware and unaware participants. This indicates that there is no relationship in the errors between gender (masculine vs. feminine) and animacy feature ($[+/- animate]$) as measured on both a recognition and controlled production task.
feminine) and type of feature [+/- animate] and [+/- count] as measured on both recognition and controlled production tasks.

**Research question 6**

To address Research Question 6 (Does the type of ending canonical (–o or –a) and arbitrary (–e and –n) differ on (a) the recognition and (b) production task?), the results of the recognition and the controlled production tests were analyzed according to the noun endings to test whether those scores differed, and if so, whether this difference was statistically significant. Table 68 illustrates the comparisons that were analyzed. If noun ending rules gender selection, it is assumed that scores will differ according to the distinction (canonical gender marking vs. arbitrary gender marking). That is to say, the difference in scores between words that share the type of ending (canonical) is assumed to be not significant regardless of gender, and the same for the words that share type of ending (arbitrary). According to this assumption, for the former case, words that end in –o (masculine) and –a (feminine) will not differ statistically due to sharing type of ending (canonical) (Table 62).

**Table 62**

*Canonical and Arbitrary comparisons*

| Masculine –o (canonical) | vs. Feminine –a (canonical) |
Masculine –e or any other consonant (arbitrary) vs. Feminine –e or any other consonant (arbitrary)

Masculine –o (canonical) vs. Masculine –e or any other consonant (arbitrary)

Feminine –a (canonical) vs. Feminine –e or any other consonant (arbitrary)

Masc –o and Fem –a (canonical) vs. Feminine –e or any other consonant (arbitrary)

**Animate vs. Inanimate**

**Multiple-choice (as recognition task)**

Descriptively, the results evidence that participants committed more errors in arbitrary ending than in canonical ending (i.e., –o and –a), e.g., Masculine (canonical) \( M = 97.83 \) vs. Masculine (arbitrary) \( M = 90.3 \). It is also noteworthy to mention that participants committed more errors in Feminine (arbitrary) than in Masculine (arbitrary) e.g., \( M = 90.3 \) vs. \( M = 85.45 \), respectively (Table 63).

The results of the percentage of correct answers obtained on the recognition task by type of ending (canonical) and (arbitrary) were submitted to a paired-sample \( t \) test (Table 64). These results show that the comparisons between type of ending (canonical) versus (arbitrary) were significant in all cases; on the
contrary, the comparisons between groups that shared the same ending (e.g.,
canonical vs. canonical or (arbitrary vs. arbitrary) were not statistically different
in any case.

Table 63
Descriptive statistics Recognition task Canonical and Arbitrary comparisons:
Animacy

<table>
<thead>
<tr>
<th>Pair</th>
<th>N</th>
<th>M (SD)</th>
<th>M (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>M Canonical vs. M Arbitrary</td>
<td>33</td>
<td>97.83 (6.31)</td>
<td>90.3 (13.34)</td>
</tr>
<tr>
<td>F Canonical vs. F Arbitrary</td>
<td>33</td>
<td>99.13 (3.46)</td>
<td>85.45 (18.22)</td>
</tr>
<tr>
<td>M Canonical vs. F Canonical</td>
<td>33</td>
<td>97.83 (6.31)</td>
<td>97.13 (3.41)</td>
</tr>
<tr>
<td>M Arbitrary vs. F Arbitrary</td>
<td>33</td>
<td>90.3 (13.34)</td>
<td>85.45 (18.21)</td>
</tr>
<tr>
<td>M &amp; F Canonical vs. M &amp; F Arbitrary</td>
<td>66</td>
<td>98.48 (19.03)</td>
<td>87.88 (5.09)</td>
</tr>
</tbody>
</table>

Table 64
Paired-sample t test results Recognition task Canonical and Arbitrary
comparisons: Animacy

<table>
<thead>
<tr>
<th>Variable</th>
<th>95% CI</th>
<th>t-value</th>
<th>p-value</th>
<th>Df</th>
</tr>
</thead>
<tbody>
<tr>
<td>M Canonical vs. M Arbitrary</td>
<td>2.42, 12.64</td>
<td>3.00</td>
<td>0.005</td>
<td>32</td>
</tr>
</tbody>
</table>
Fill-in-the blank (as controlled production task)

Results of the paired-samples $t$ test were slightly different than those of the recognition task (Table 65). On this task, the statistically significant comparisons between different type of noun endings (canonical) and (arbitrary) were those between feminine (canonical) and feminine (arbitrary), and between the masculine and feminine (canonical) versus the masculine and feminine (arbitrary); the other comparison that was statistically significant was the one between masculine (arbitrary) and feminine (arbitrary) (Table 66).

Table 65

**Descriptive statistics Production task Canonical and Arbitrary comparisons:**

<table>
<thead>
<tr>
<th>Animacy</th>
<th>Pair</th>
<th>$N$</th>
<th>$M$ (SD)</th>
<th>$M$ (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M Canonical vs. M Arbitrary</td>
<td>33</td>
<td>96.63 (10.93)</td>
<td>96.97 (9.73)</td>
</tr>
<tr>
<td></td>
<td>F Canonical vs. F Arbitrary</td>
<td>33</td>
<td>98.61 (4.55)</td>
<td>84.85 (18.69)</td>
</tr>
<tr>
<td></td>
<td>M Canonical vs. F Canonical</td>
<td>33</td>
<td>96.63 (10.93)</td>
<td>98.11 (4.55)</td>
</tr>
<tr>
<td></td>
<td>M Arbitrary vs F Arbitrary</td>
<td>33</td>
<td>96.97 (13.34)</td>
<td>84.85 (18.69)</td>
</tr>
<tr>
<td></td>
<td>M &amp; F Canonical vs.</td>
<td>66</td>
<td>97.37 (8.34)</td>
<td>90.90 (15.99)</td>
</tr>
<tr>
<td></td>
<td>M &amp; F Arbitrary</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 66

*Paired-sample t test results Production task Canonical and Arbitrary comparisons: Animacy*

<table>
<thead>
<tr>
<th>Variable</th>
<th>95% CI</th>
<th>t-value</th>
<th>p-value</th>
<th>Df</th>
</tr>
</thead>
<tbody>
<tr>
<td>M Canonical vs. M Arbitrary</td>
<td>-3.37, 2.69</td>
<td>2.7</td>
<td>0.823</td>
<td>32</td>
</tr>
<tr>
<td>F Canonical vs. F Arbitrary</td>
<td>6.61, 19.9</td>
<td>4.65</td>
<td>0.000</td>
<td>32</td>
</tr>
<tr>
<td>M Canonical vs. F Canonical</td>
<td>-5.19, 2.24</td>
<td>-.807</td>
<td>0.426</td>
<td>32</td>
</tr>
<tr>
<td>M Arbitrary vs F Arbitrary</td>
<td>4.73, 19.51</td>
<td>3.34</td>
<td>0.002</td>
<td>32</td>
</tr>
<tr>
<td>M &amp; F Canonical vs. M &amp; F Arbitrary</td>
<td>2.52, 10.39</td>
<td>3.28</td>
<td>0.002</td>
<td>65</td>
</tr>
</tbody>
</table>

To summarize, in the case of the recognition task the comparisons between canonical ending and arbitrary ending were all statistical significant; however, for the case of the most demanding task (the controlled production task), two of the three comparison between canonical and arbitrary ending were statistical significant.

**Count vs. Mass**

The analyses of the percentage of correct answers obtained on the recognition task by type of noun ending (canonical) and (arbitrary) were submitted to a paired-sample *t* test. The descriptive statistics are shown in Table 67. The results showed that for the recognition task, two out of three comparisons that tested different type of noun ending (canonical) vs. (arbitrary) were
significant; those comparisons are between feminine (canonical) and feminine (arbitrary), and between masculine and feminine (canonical) and masculine and feminine (arbitrary) (Table 68). For the production task the only pair that resulted significant was for the feminine (canonical) and feminine (arbitrary) (Table 70). The descriptive statistics are shown in Table 69.

Table 67

*Descriptive statistics for the Recognition task: Count*

<table>
<thead>
<tr>
<th>Pair</th>
<th>N</th>
<th>M (SD)</th>
<th>M (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>M Canonical vs. M Arbitrary</td>
<td>29</td>
<td>93.33 (11.52)</td>
<td>86.67 (25.67)</td>
</tr>
<tr>
<td>F Canonical vs. F Arbitrary</td>
<td>29</td>
<td>96.25 (5.82)</td>
<td>90.83 (12.25)</td>
</tr>
<tr>
<td>M Canonical vs. F Canonical</td>
<td>29</td>
<td>93.33 (11.52)</td>
<td>96.25 (5.82)</td>
</tr>
<tr>
<td>M Arbitrary vs F Arbitrary</td>
<td>29</td>
<td>86.67 (25.67)</td>
<td>85.45 (18.21)</td>
</tr>
<tr>
<td>M &amp; F Canonical vs. M &amp; F Arbitrary</td>
<td>29</td>
<td>94.79 (9.17)</td>
<td>88.75 (20.05)</td>
</tr>
</tbody>
</table>

Table 68

*Paired-sample t test results for the Recognition task: Count*

<table>
<thead>
<tr>
<th>Variable</th>
<th>95% CI</th>
<th>t-value</th>
<th>p-value</th>
<th>Df</th>
</tr>
</thead>
<tbody>
<tr>
<td>M Canonical vs. M Arbitrary</td>
<td>-2.62, 15.95</td>
<td>1.47</td>
<td>0.153</td>
<td>28</td>
</tr>
<tr>
<td>F Canonical vs. F Arbitrary</td>
<td>.562, 10.27</td>
<td>2.82</td>
<td>0.03</td>
<td>28</td>
</tr>
<tr>
<td>M Canonical vs. F Canonical</td>
<td>7.37, 19.98</td>
<td>4.42</td>
<td>0.263</td>
<td>28</td>
</tr>
</tbody>
</table>
### Table 69
Descriptive statistics for the Production task: Count

<table>
<thead>
<tr>
<th>Pair</th>
<th>N</th>
<th>M (SD)</th>
<th>M (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>M Canonical vs. M Arbitrary</td>
<td>29</td>
<td>96.33 (6.6)</td>
<td>98.33 (9.13)</td>
</tr>
<tr>
<td>F Canonical vs. F Arbitrary</td>
<td>29</td>
<td>98.09 (6.2)</td>
<td>92.00 (13.49)</td>
</tr>
<tr>
<td>M Canonical vs. F Canonical</td>
<td>29</td>
<td>96.33 (11.52)</td>
<td>98.09 (6.2)</td>
</tr>
<tr>
<td>M Arbitrary vs F Arbitrary</td>
<td>29</td>
<td>98.33 (9.13)</td>
<td>92.00 (13.49)</td>
</tr>
<tr>
<td>M &amp; F Canonical vs. M &amp; F Arbitrary</td>
<td>58</td>
<td>97.21 (6.46)</td>
<td>95.17 (11.86)</td>
</tr>
</tbody>
</table>

### Table 70
Paired-sample t test results for the Production task: Count

<table>
<thead>
<tr>
<th>Variable</th>
<th>95% CI</th>
<th>t-value</th>
<th>p-value</th>
<th>Df</th>
</tr>
</thead>
<tbody>
<tr>
<td>M Canonical vs. M Arbitrary</td>
<td>-6.43, 2.42</td>
<td>-0.92</td>
<td>0.363</td>
<td>29</td>
</tr>
<tr>
<td>F Canonical vs. F Arbitrary</td>
<td>0.649, 11.54</td>
<td>2.29</td>
<td>0.03</td>
<td>29</td>
</tr>
<tr>
<td>M Canonical vs. F Canonical</td>
<td>-5.45, 1.92</td>
<td>0.977</td>
<td>0.337</td>
<td>29</td>
</tr>
<tr>
<td>M Arbitrary vs F Arbitrary</td>
<td>-0.58, 12.72</td>
<td>2.03</td>
<td>0.052</td>
<td>29</td>
</tr>
</tbody>
</table>
To conclude, although not all the contrasts between different type of noun endings (Canonical) and (Arbitrary) resulted statistically significant, it is noted that only in one case the comparison between the same type of feature resulted statistically significant (cf. animate vs. inanimate fill-in-the-blank production task comparison between masculine (Arbitrary) and feminine (Arbitrary)).

Summary

To summarize, the statistical analyses with regards to the animate vs. inanimate experimental feature revealed that aware participants performed higher than the unaware group and those scores were statistically different in the case of both old and new items. Aware participants also performed better on the post-delayed tests than unaware participants although it was not statistically significant and they did not show the customary loss that is usually found in the delayed post-tests, as the unaware group did. Unaware participants statistically performed above chance only in the case of old items, while the aware participants performed above chance on both old and new items. Aware participants selected their responses based not only on gender –the disclosed element of the determiner system–, but on animacy too. Unaware participants, on the other hand, based their choices only on gender.
With regards to gender, the analyses indicated that statistical differences were found, mostly, for the comparisons between canonical ending vs. arbitrary gender, and not when nouns when compared according to [+ feminine] versus [-feminine].
CHAPTER 5: DISCUSSION AND CONCLUSION

Discussion

Overview

This dissertation sought to address the issue of whether there was a distinction between the performances of aware and unaware participants on implicit form-meaning mappings during an incidental reading task. It also explored from the Universal Grammar perspective whether type of feature [+/- animate], [+/- count], [+/- feminine] influenced the performances on the recognition and production assessment tasks. A central inquiry of the dissertation was the role of interpretable and uninterpretable features.

This chapter follows the same order of the research questions presented in the previous chapter. The first section compares the performances of the aware and unaware participants on the assessment tasks (recognition and production) and their behavior over a one-week period (i.e., Research question 1). Is it possible to learn form-meaning connections without awareness (at the encoding stage) above chance level? If so, do unaware and aware participants show learning of the undisclosed features besides gender? The following section addresses the possibility of implicit learning of form-meaning connections through a reading task. Aware and unaware participants’ responses were analyzed to examine if they were able to learn form-meaning connections above chance
The third section discusses aware and unaware participants’ responses according to the undisclosed features (animate versus inanimate and count versus mass) in order to investigate whether participants responded equally correctly to both type of features or if they presented a bias (i.e., Research question 3). The fourth section addresses aware and unaware participants’ responses according to both types of features (animate versus inanimate and count versus mass) and type of noun ending with regard to gender: canonical gender (–o and –a) and arbitrary gender (–e and –n) to address the behavior of those features as a cluster (i.e., Research question 4). The fifth section discusses aware and unaware participants’ responses according to gender (masculine and feminine) and type of feature (animate versus inanimate and count versus mass) to address the behavior of these features as a cluster (i.e., Research question 5). The sixth, and final section, addresses aware and unaware participants’ responses to the recognition and production tasks according to type of noun ending in regard to gender: canonical gender (–o and –a) and arbitrary gender (–e and –n) (i.e., Research question 6).

Each research question is presented prior to its subsequent discussion; the order of presentation is (1) animate versus inanimate followed by (2) count versus mass experimental groups.
Research question 1

Do aware advanced L2 learners perform significantly better than unaware learners, as measured on a recognition and a production assessment task?

In general, aware participants performed significantly better than unaware participants in this experiment; yet, these results differed according to the experimental conditions.

For the Animate versus Inanimate experimental group, results from Tables 21 and 22 for the recognition task, and Tables 23 and 24 for the production task evidenced that aware participants performed statistically better than unaware participants on all the tasks, as well as in the case of old and novel items with a large effect size in all the cases. However, for the count versus mass experimental group, aware participants performed significantly better than unaware participants only for the recognition task in the case of old items (Table 25 and Table 26). On the productive task, although aware participants scored better than unaware participants, this score was not statistically significant for the old or novel items (Table 27 and Table 28).

These conflicting results may be explained with regard to the varying levels of awareness reported by the aware groups in the two experimental conditions (animate versus inanimate and count versus mass). While participants in the Animate versus Inanimate group were able to report awareness at the level
of understanding, none of the participants in the count versus mass experimental group reported such a high level of awareness.

Five out of seven participants who became aware in the animate versus inanimate group were able to verbalize the animacy rule that underlay the determiner system, mentioning it during the on-line verbal protocol, and, in three cases, even verbalizing it on the off-line post exposure questionnaire. This higher level of awareness is reflected in the participants’ scores that support similar findings reported in the literature (Leow, 1997; Rosa & Leow, 2004; Rosa & O’Neill, 1999). On the contrary, none of the participants in the count versus mass experimental group reported being able to verbalize the rule. Only one of them mentioned a partially incorrect rule in the think-aloud protocol, failing to do so on the post-exposure questionnaire. What these findings appear to suggest is some types of linguistic items may not only need to be noticed but also processed deeper before becoming candidates for internalization in the system.

Results from this study appear to lend support to those theoretical and empirical claims that assume that awareness in second language processing – and learning – plays an important role (Hama & Leow, 2011; Leow, 1997, 2000; 2001; Robinson, 1995b; Schmidt, 1990 and elsewhere). The novelty of this dissertation was to explore the effects of awareness or lack thereof within two analogous tasks that differed with regard to type of feature. In this regard, the conclusion is that type of feature may play a role in whether or not the feature is
deeper processed. Task demands and type of feature should be variables that need to be included in experiments that address (un)awareness in order to further understand this phenomenon.

**Delayed post-test: Animate versus Inanimate**

With regards to the delayed post-test, results of the animate versus inanimate group indicated that there were main effects for Group on both assessment tasks (recognition and production) but not for Time nor was there a significant interaction of Time and Group (Table 22 and Table 24). The main effect for Group means that while aware and unaware groups performed statistically differently between them this score did not vary in time.

In relation to the aware participants, results revealed that this group maintained their scores above chance, and as a result, learning was retained for this period, with a small increase that was not statistically significant for the recognition (e.g., old items $M = 88.09$ versus $M = 94.04$), nor for the production (e.g., new items $M = 79.76$ versus $M = 84.52$), Table 21 and Table 23, respectively.

Unaware participants, on the other hand, performed at the chance level from the immediate post-test to the delayed post-test on the recognition task for old items ($M = 57.76$ versus $M = 50.64$) and below chance level for new items ($M = 41.67$ versus $M = 46.8$). For the production task, unaware participants
performed the same, above chance from the immediate post-test to the delayed post-test for old items ($M = 54.8$ versus $M = 53.84$) and below chance level for new items ($M = 48.07$ versus $M = 45.19$) (Table 21 and Table 23, respectively).

**Delayed post-test: Count versus Mass.**

The results of the delayed post-tests for the count versus mass presented a completely different finding. For the recognition task, there was a main effect for Group only for the old items, but not for new items. A main effect was not found for Time nor was there a significant interaction of Time and Group for either the old or novel items. The main effect for Group for the old items indicated that both groups performed differently only on the recognition task but not for the novel items. The low level of awareness may be the main reason why the beneficial effects of being aware did not translate to a better performance on the novel items as well as in the case of the effect across time (Table 26 and Table 28).

The animate versus inanimate experimental condition reported levels of awareness similar to the level of understanding (i.e., verbalization of the rule), while aware participants for the count versus mass distinction reported levels of awareness similar to noticing (i.e., mentioning or commenting upon the undisclosed feature).

While there is some evidence for an association between awareness and further processing of L2 data (Leow, 1997, 2000, 2001; Rosa & Leow, 2004,
Rosa & O’Neill, 1999), there is still some contradictory empirical evidence in regards to what extent noticing and understanding correlate with intake and learning.

For example, Leow (2000) found no dissociation between awareness and learning in a study designed to empirically test Schimd’t’s and Tomlin and Villa’s theoretical assumptions with regard to the role of awareness in L2 processing and learning. Participants coded as aware significantly took in and produced more instances of the targeted morphological form (third person irregular preterit in Spanish) using a problem solving task. However, studies addressing not only aware versus unaware participants, but also levels of awareness, have shown that different levels of awareness correlate differently with intake and learning. Rosa and O’Neill (1999) have found that both Understanding and Noticing (similar to the level of awareness detected for the aware participants for the count and mass experimental condition) led to pretest to posttest improvements. Robinson (1995b, 1996, 1997) proposed that only awareness at the level of Understanding led to higher levels of learning. Leow (2000) found that participants that demonstrated higher levels of awareness performed significantly better than learners who did not. Rosa and Leow (2004) found that participants at the level of Noticing (similar to the level of awareness detected for the aware participants for the count and mass experimental condition in this dissertation) increased their score significantly more when compared to the No Report group.
Faretta-Stutenberg and Morgan-Short (2011), within the same strand of meaning-form mappings under implicit conditions, directly addressed the issue of levels of awareness. They found that, descriptively, the Understanding group performed better than the Noticing or the No Report; nevertheless there were no significant differences between the three aware groups. More interesting is the way they further examined the learners’ performance in the second testing phase as all participants were asked to look for rules and all provided a guess of that rule. To approximate the classification according to Williams’ (2005) original study the No report group was eliminated, which resulted in two groups: Aware, which was the same as Understanding, and Unaware, which was equal to Noticing (italics mine). Although no statistical analyses were conducted due to the small numbers of participants in the aware group (i.e., awareness at the level of understanding), the results are descriptively similar to the count versus mass group. “Unaware” participants at the Noticing level performed better on old items, however at the chance level for the generalization items in Faretta-Stutenberg and Morgan-Short’s study, they performed just like aware participants in the count versus mass experimental group in this dissertation.

Is awareness at the level of Noticing not sufficient for furthering processing input or to generate novel items for certain type of linguistic features? There is no doubt that some of the contradictory results in studies addressing form-meaning mapping under implicit conditions (e.g., Williams, 2004, 2005;
Leung & Williams, 2011 versus Leow, 2000; Hama & Leow, 2011, Faretta-Stutenberg & Morgan-Short, 2011) are the consequence of (1) different instruments to measure awareness (think-aloud protocols versus post-exposure questionnaires), which results in (2) measuring awareness or lack thereof at a different level: low and evanescent instances of awareness versus higher levels of awareness, and (3) type of feature. These three factors might have impacted the level of awareness that they have detected.

In conclusion, in this study, aware participants in the animate versus inanimate experimental group maintained their gains between the immediate post-test and the delayed post-test; however, aware participants in the count versus mass condition declined between the immediate post-test and the delayed post-test for the recognition task (old items). These contradictory results may be the consequence of different levels of awareness detected for both groups of aware participants. Higher levels of awareness have been correlated with more complex intake processes or learning (Leow, 1997, 2001; Robinson, 1995b, 1996, 1997; Rosa & O’Neill, 1999). The complex processing needed for certain type of features require a higher level of awareness as it seems to be the case between the contradictory results in the two type of features under investigation in this research: animate versus non-animate and count versus mass. This will be reviewed thoroughly in the Research Question 2.
Research Question 2

Is it possible to learn form-meaning connections without awareness (at the encoding stage) above chance level? If so, do aware and unaware participants evidence learning of the undisclosed features besides gender?

Animate versus Inanimate

Results from the one-sample t-test to test whether participants were able to perform above the 50% chance level indicated that for the animate versus inanimate group, unaware participants performed significantly above chance only in the case of the old items on the recognition task (Table 29). The aware participants, on the other hand, performed significantly above chance for both assessment tasks and with regard to both old and novel items (Table 30).

Count versus Mass

In the case of the count versus mass group, unaware participants did not perform above chance in any of the assessment tasks (e.g., recognition or production) for either the novel or the old items. Aware participants performed above chance only on the recognition task for old items, but not for novel items. These aware participants did not perform above chance for the novel or old items on the production assessment task.
Since the one sample $t$-test examined whether participants performed only above the chance level regardless of the type of answers they provided, a different and finer statistical analysis was conducted (Hama & Leow, 2011). A $2 \times 2$ repeated measures ANOVA was conducted for old and new items separately in order to analyze correct and incorrect responses according to the disclosed and the undisclosed features. The crucial analysis is to compare the means of the correct and incorrect responses of the undisclosed feature (e.g., [+/- animacy] for the animate versus inanimate experimental group as well as [+/- count] for the count versus mass experimental group) to test whether this difference is statistically significant. This finer and more rigorous statistical analysis more accurately captures participants’ performance with regard to the undisclosed feature.

Unaware participants correctly selected their responses based on gender and animacy only in the case of the old items but not for the new items on the recognition task according to this more sophisticated analysis (Table 37). This means that while the training session was successful in teaching participants to recognize the trained items, they were unable to generalize this ability to novel items, which would be an indication of learning. On the contrary, aware participants correctly selected their responses on the recognition and production assessment tasks for the new and old items based on gender and animacy, the disclosed and the undisclosed feature, respectively (Table 38).
In the case of the count versus mass unaware participants based their responses only on gender but not on count (the undisclosed feature) on the recognition and production tasks. Aware participants based their selection on gender and count on the recognition task for old items, but not for new items on the recognition and the production task. These findings indicate that, unlike the feature animacy, participants’ performances with respect to the features count versus mass appears to be differential, most likely due to the opaqueness of the contrast or, as discussed above, to differential levels of awareness reported.

Methodological issues
Awareness or lack thereof was measured in this study using a hybrid design to triangulate data by gathering data at both the time of encoding (pre-training, treatment and testing phases) and after the experimental exposure (post-exposure questionnaire). Online data-elicitation techniques offer two advantages (1) control participants’ performance in the experimental groups; that is to say, they verify participants’ completion of the treatment and the assessment tasks as they were expected according to the condition under which they participated; and, more importantly for studies addressing (un)awareness (2) they detect any instance of awareness concurrently, that is, as participants perform during the whole treatment.
Relying on post-exposure questionnaires to measure awareness and, above all, on the verbalization of the rule as the crucial factor to draw the line between aware and unaware participants seems to be insufficient in detecting those levels of awareness below that level. Studies that have found some disassociation between awareness and learning (Williams, 2004, 2005; Leung & Williams, 2011a, 2011b) have relied on verbalization of the rule after completing the tasks to classify participants as aware, and even assumed that if “phenomenal awareness” might have been present, it was distinct from reflective awareness, which is reportable (Williams, 2005). According to Williams (2005), unaware participants did not “use these fleeting impressions to guide their responses in the test tasks (...) Whatever phenomenal awareness they might have had appeared to have been immediately forgotten and did not form the basis for rule learning or intentional behavior.” (p. 293-294). However, studies that have employed a finer measurement of awareness at the concurrent stage and at the post-exposure stage have been able to classify as aware those participants that represent lower levels of awareness (awareness at the level of noticing), and, consistently, have not been able to find any dissociation between awareness and learning (Hama & Leow, 2010; Faretta-Stutenberg & Morgan-Short, 2011).

Hama and Leow (2010) found that participants reporting awareness of the animacy rule while completing the assessment tasks failed to do so in the offline questionnaire. In this study, for example, one participant failed to report any
mention of the animacy rule in the post-exposure questionnaire. The analysis of the think-aloud protocol clearly revealed this participant’s early adoption of a data-driven strategy at the beginning, activating previous knowledge that triggered the animacy rule with a low level of confidence.

(1) Casa is a feminine word so it can be ru or ju. I go for ru casa

Educación is la educación. I’m sorry you are not supposed to use la. So it has to be ru or ju.

Sobrina is la sobrina. So ju sobrina or ru sobrina. I’m going to go with ju sobrina just because sounds better.

Enfermero? Ji enfermero or ri enfermero. Ri enfermero.

Río. Ji río or ri río. I’m going to go with ri río.

Vecina is feminine. So I’m going to go with ru vecina (…)

Masculine is also masculine so I’m going to go with ji.

Novio is masculine so I’m going to go with ji.

Vínculo. I remember this one in the readings so ri vinculo.

I kind of remember that family members were with j but I’m not positive.

Ji padre.

Ji nieto.

Ju familia. No, ru familia.

Pantalón. So I will go with ji pantalón. No, ri pantalón.

As illustrated in the think-aloud protocol, this participant hesitated about his answers, used his previous knowledge to answer correctly some of the answers using a data-driven strategy, and mentioned a partial rule almost at the end of the assessment tasks; however, this same participant offered the following
rationale for grouping the articles *ji, ju, ri, ru*: “They're paired into masculine and feminine article groups.”

Semi-aware participants, as called by Lambert (2002), are the aware participants detected in the Count versus Mass group. As it can be seen in Table 26 and Table 28, the scores of aware participants (with a low level of awareness) reached significance only in the case of old items, but not for novel items. However, a detailed analysis of the think-aloud protocols revealed that those participants could not be classified as unaware. These aware participants presented a data-driven process during the assessment tasks completion that clearly evidenced some conscious attempt at explicit knowledge formation. It is not surprising then that there is found a completely different behavior when two groups of participants are compared: one with the highest level of awareness possible (i.e., awareness at the level of understanding) with another group of participants that include, mostly, participants with a lower level of awareness (awareness at the level of noticing) unaware as well as aware participants.

As discussed above, some studies having registered some sort of learning have relied solely on post-exposure questionnaires, which capture only a higher level of awareness, for grouping aware and unaware participants and measured learning with a recognition assessment task that requires a low level of processing (Leung & Williams, 2011a, 2011b; Williams, 2004, 2005). This section discusses exclusively the issues regarding learning.
Studies that have claimed that participants were able to perform above chance, based those assertions primarily on statistical analyses that addressed whether or not those unaware participants were able to perform above chance using a one-sample t-test (e.g., Williams, 2004, 2005). On the contrary, Hama and Leow (2011) and this study came to similar conclusions using a 2 × 2 repeated measures, a more rigorous statistical design, to capture participants’ responses according to all the possible variables (e.g., disclosed and undisclosed items).

Since the amount of learning that takes place in implicit studies, although statistical significant, is not very large (Leung & Williams, 2011a, 2011b, Williams, 2004, 2005), the methodological design of implicit studies should be carefully revised. Increasing the number of options, especially in multiple choice tasks that only demand the selection of a correct answer out of two options, in which 50 percent reflects mere chance, strengthens the findings of the study. Studies having provided evidence of implicit learning have come to these conclusions based on a receptive task that requires a low processing of the target items as well as the selection of a correct response out of two possible options even though the target items that needed to be learned were four critical distinctions (Williams, 2004; 2005). This study and Hama and Leow (2010), which increased the number of possible options to four in order to have a distractor that covered all four possible options, found no dissociation between awareness and learning.
Based on the findings in this study and also on Faretta-Stutenberg and Morgan-Short (2011) and Hama and Leow, (2010), it seems that participants merely “learn” to recognize items (e.g., priming effect), which implies an early stage along the learning process. As a result, it is quite debatable to assume that, based on this early stage, any internalization has taken place. It is possible to at a certain extent support the view that sees implicit learning, at least in the early stages, as a process of statistical patterns involving memorization rather than abstract rule formation (Shanks, 2005)

For Boyer, Destrebecqz and Cleeremans (2005) within cognitive psychology, implicit sequence learning is essentially statistical in nature and only involves simple associative predictions, which explains the failure to transfer the effects of any implicit learning to novel material: “transfer decrement” was found, which suggests that rule abstraction did not occur (p. 396).

Empirical findings in SLA come from DeKeyser (1995) and Robinson (1996). DeKeyser concluded that categorical rules were learned better in an explicit condition. However, support for the hypothesis that implicit instruction works best with prototypical rules was less clear. Robinson (1996) found that explicit instruction seemed to be more effective for learning simple rules; yet, implicit instruction was not more effective for learning complex rules.

In this experiment and considering type of feature as a variable, in the count versus mass experimental group only those participants who reported some
low level of awareness appeared to be able to recognize only old items, a behavior not evidenced by the unaware group.

How is it possible that some aware participants performed above chance whereas others did not? A detailed analysis of the think-aloud protocols revealed that participants employed different strategies for assessment task completion. Aware participants in both experimental conditions presented two dissimilar strategies to perform the assessment tasks. Participants in the aware group for the animate versus inanimate experimental condition seemed to tackle the task in a general state of awareness reflected on their perfect or almost perfect scores on the assessment tasks; aware participants in the count versus mass experimental group demonstrated different levels of awareness.

Some participants employed data-driven processes while others employed conceptually-driven processes (Leow, 1998; Robinson, 1995b, 1997). Data-driven processing is the buildup and rehearsal of chunks previously encountered in memory without analysis. According to Robinson (1995b) it “may lead to the development of simple patterns of association between co-occurring items (p. 301-302). Conceptually-driven processing, on the other hand, is the elaboration of the input by activating schemata, which leads to a more sophisticated and abstract knowledge (Gillam et al, 2000). The former process leads to lower levels of awareness, while the second usually triggers higher levels of awareness (e.g., the possibility of verbalizing the rule). These strategies have been found previously
while participants are completing the treatment task and not only in the assessment phase (Leow, 1998, 2000, 2001; Robinson, 1995b, 1997; Rosa & Leow, 2004; Rosa & O’Neill, 1999).

**Animate versus Inanimate**

Aware participants in the animate versus inanimate experimental group followed for the most part a conceptually-driven process characterized by hypothesis testing and rule formation. This can be observed in the following participant’s excerpt in this experimental condition. This participant was testing earlier at the beginning of the first assessment task a rule that he had constructed, and therefore was testing this hypothesis as the rationale for selecting his answers:

(2) Juventud is not a person.
Ciudadana is ju.
Ru educación.
La familia. Ru familia.
Miembro is ji miembro. It’s a person. (…)
Vínculo. I know it’s not a person. Ri vínculo.

An example of data-driven processing is the following, in which a different participant in the animate versus inanimate experiment analyzed the items relying on gender –the disclosed feature– and memory.

(3) I know those i’s are masculine.
So this in entirely guessing.
Ju madre. I remember that one.
Biblioteca is feminine. I don’t remember this one. (…)
Basically I’m guessing.
Ri amor. I’m guessing.
Ru corazón.
Ji niño. I remember this one.

However, it seems that some participants did not solely follow a data-driven or a conceptually-driven processing from the beginning to the end; for some participants, the former processing took them to a conceptually-driven processing if there were the necessary conditions. Some participants in the animate versus inanimate experimental group transitioned from data-driven processing to conceptually-driven processing in a specific moment that could be detected in the think-aloud protocols:

(4) Ciudadana. I don’t know if it’s ru or ju? Wait. Ju persona
   Educación is masculine or feminine? Ru educación (…)
   El profesor. Ji is a person so. Ji profesor.
   Ru casa.
   Ju hija.
(5) Ri fiesta.
   Ji padrino.
   Honestly, I’m going to have family with j and all the rest with r. (…)
   Ru vecina.
   Ji profesor.
Count versus Mass

On the other hand, it appears that the participants in the count versus mass only relied on data-driven processing and therefore based their decisions on gender and memory only. In other words, it appears that they simply analyzed the items as a chunk without being able to decipher the rule.

(6) I can’t remember why it’s Pu crianza. I think I remember that.
   Pu importancia.
   Fu cuidado. No, no fi cuidado. No, pi cuidado sounds familiar. (…)
   Pu solidaridad sounds better.
   Pi comportamiento. I’m pretty sure that one was there.
   Pu amiga? Fu amiga?
   Pu padre. Pi padre. I remember this one.

Although Williams (1999) identifies data-driven processing with implicit learning and therefore learning with the absence of awareness, a detailed analysis of the think-alouds protocols revealed that a low level of awareness is also present in data-driven processing. For Robinson (1995b), both processes – data-driven and conceptually-driven – require awareness although the use of one or the other is subordinated to task demands. This dissertation believes that while the use of one or the other may depend also on individual differences – as Leow (2000) proposes – type of linguistic item or feature may also play a crucial role, as a detailed analysis of these participants’ think-aloud protocols evidenced. For example, Leung and Williams (2012), who investigated implicit learning of two
different type of features, found “implicit learning” was possible for the [+/- animate] feature but not for [+/- near] feature under exactly the same conditions. Therefore, based on the results obtained in this study and Leung and Williams (2012), it seems that some features are less amenable to be learned under implicit conditions. Participants in the [+/- animate] experimental condition were able to reach higher levels of awareness and decipher the rule, whereas participants in the [+/- count] were not successful in doing so. These results seem to support Hulstijn and de Graaff’s theoretical claims as well as Robinson’s empirical findings.

In summary, this study found that only participants with a level of awareness similar to understanding (i.e., verbalization of the rule) were able to perform above chance on both assessment tasks (recognition and production) with respect to trained and generalization items. Despite theoretical (Schmidt, 1990, 1995) and empirical (Leow, 2000; Rosa & Leow, 2004) claims assuming that awareness at the level of noticing is dissimilar to no report of awareness (e.g., unaware) and to awareness at the level of understanding, the aware participants in the count versus mass experimental condition with a level of awareness similar to noticing performed above chance on trained items only for the recognition task, but not for novel items. Unaware participants in both experimental conditions (animate versus inanimate and count versus mass) performed at the chance level.

Following a classification used by Rosa and O’Neill (1999) and Hama and Leow (2011), this study found no evidence of learning for the unaware group or
for the aware participants at the level or noticing with respect to the count versus mass contrast. It is crucial to employ finer measurements of awareness before making claims suggesting that certain learning has taken place under “the implicit condition” due to the fact that in some cases this evidence is the result of participants not performing the tasks according to their implicit condition (Robinson, 1996); in other cases, low or partial levels of awareness are not detected, and as a result partial learning is reported. In this dissertation, aware participants in the animate versus inanimate experimental group (most of them at the level of understanding, according to the what the analysis of the think’alouds seem to evidence) are the only ones that performed above chance for old and novel items. Playing roles in type of processing (data-driven versus conceptually-driven) may be individual differences and type of linguistic items or features. Obviously, before making claims about how type of feature affects the form-meaning mappings it is necessary to investigate other type of features.

Research Question 3

Does type of features [+/- animacy] and [+/- count] have a differential effect on the level of awareness reported by advanced L2 learners during the reading task?

The experimental design employed in this study followed previous studies in which a semi artificial determiner system responds to a combination of two features (cf. Williams, 2005, 2005, far versus close and animate versus
The novelty of this dissertation consisted in (1) the medium for delivering the treatment condition and (2) the selection of the features. Previous studies have delivered the target items embedded in isolated sentences via the aural mode; in contrast, the target items were equally and carefully distributed throughout the texts in this experiment. Participants had to map form-meaning connections while reading a text for meaning; as a result, participants were exposed to the input via a meaning-focused task.

The selection of features in this experiment responded to a careful selection that intended to explore the combination of a cluster of two features: one present in the repertoire of available features in the L1 and L2 ([+/- animate] and [+/- count]) with another feature that resulted crucial in the L2 inventory (+/- feminine)).

Although it can be said that both English and Spanish have gender, at least in a broad sense, only Spanish has syntactic agreement. Gender in contemporary English is not an inflectional category and there are few traces of the old English gender system, which was similar to present-day German (Brinton, 2006). There is gender only for pronouns (e.g., he/him or she/her), their possessive forms (e.g., his or her, hers), and their reflexive forms as well (e.g., himself, herself); nevertheless, there is no gender for nouns and, as a result, there is no gender agreement for determiners or adjectives, and consequently, no formal gender feature in any of the relevant features.
Gender in Spanish is usually marked with an ending, \textit{-o} for masculine, \textit{-a} for feminine, and \textit{-e} and any other consonant for either masculine or feminine in a more or less arbitrary fashion; however, nouns do not receive any morphological realization if they belong to the [\pm animate] or [\pm count] category. Hence, the idea was also to explore the combination of a feature (e.g., [\pm animate] and [\pm count]) that does not receive any morphological realization with one that does (at least in the agreement between a determiner and a noun), e.g., [\pm gender] in Spanish: \textit{el niño} [the (masculine) boy] versus \textit{la niña} [the (feminine) girl]. In this part of the discussion section, the features [\pm animate] and [\pm count] will be addressed, whereas gender will be discussed in the following research questions.

Previous studies within this strand of research analyzed answers to detect evidence of learning based on the scores of the undisclosed feature. If the answers were sensitive to both features – the disclosed, and, more importantly, the undisclosed –, it was considered evidence that the participant learnt or acquired the system. However, those studies did not analyze the responses according to type of features; that is to say, the analyses did not explore possible interpretations of participants’ errors. First, the animate versus inanimate experimental group will be analyzed, followed by the count versus mass.
Animate versus Inanimate

On the recognition assessment task (e.g., the receptive task), aware and unaware participants correctly identified an equal number of animate and inanimate nouns; however, on the fill-in-the-blank task (e.g., the productive task), only aware participants responded correctly to an equal number of animate and inanimate words, which is consistent with having learnt the system. That is to say, having an equal number of animate and inanimate words in the texts, and, since aware participants deciphered the rule, they identified an equal number of animate and inanimate words. However, the unaware participants in the case of the more demanding task, answered significantly more correct animate than inanimate words (Animate nouns $M = 59.29$ versus Inanimate nouns $M = 46.15$).

It is interesting to speculate as to the reasons why this occurred. One plausible explanation is that the features gender and animacy are closely related even in the L1 of the participants (English). In English, the only way to interpret gender, according to Carroll (2001, p. 83) is “as a semantic distinction based on the reference of nouns in context and relevant only for pronoun selection.” As a result both features are closely related, as it was mentioned in Chapter 2 when it was commented that English has a natural gender distinction that is required for the case of the selection of some pronouns ($he$ vs. $she$).
Count vs. mass

For the count versus mass experimental group, the aware group demonstrated a low level of awareness and none of the participants was able to either comment on or verbalize the rule. This low level of awareness seemed to be beneficial only for old items but not for novel items. In relation to the items that reflected knowledge of the system, both groups demonstrated similar performance. The aware and unaware group correctly identified significantly more count than mass target items on the recognition task. For the production assessment task, aware and unaware participants correctly identified a similar number of count and mass target items.

Bearing in mind that all the four features in the experiment are equally salient, it is expected to assume that detectability did not play a role. The experiment consisted of suffixal derivation, e.g., ji, ju (animate) versus ri, ru (inanimate) and pi, pu (count) versus fi, fu (mass). There is also evidence from the think-aloud protocols as well as from the post-exposure questionnaires that participants systematically noticed the target items while completing the tasks. However, it is known that suffixal derivation is less noticed than a suppletive realization, and as a result participants produced suffixal derivation less (Lardiere, 2008, 2009b). According to her, for example this explains that Case is consistently produced by L2 learners despite the fact that Case in pronouns is considered uninterpretable, “perhaps due to their suppletive rather than suffixal
nature” (Lardiere, 2009b, p. 410). As a result, Lardiere (2008) even proposes that “suppletive forms are acquired early and more robustly than affixal forms, certainly at least for adult language acquisition” (p. 121).

The features [+/-animate] and [+/-count] are part of the repertoire of available features in Spanish and English, and participants did not need to add any additional feature to their current inventory of L1 features. In the study, both features were also uninterpretable since they were agreement features in the determiners. Nevertheless, we see that unaware and aware participants with a low level of awareness demonstrated a bias toward animate and count nouns while selecting their responses. Experiments have evidenced that people can recognize animals with little or no focal attention and with incredible speed (Li, VanRullen, Koch, & Perona, 2002 and Kirchner & Thorpe, 2006 cited in Leung & Williams 2011b), which demonstrates that there is an evolutionary benefit in detecting animate entities.

The context in which the mass and count nouns appeared in the texts could also be a factor that impeded the detection of mass and count nouns as exemplars of one or the other category. The fact that all the nouns were used in singular and in a [+ definite] context obscures the most recognizable approach to detect nouns as count or mass. In English – more than in Spanish – common nouns are separated according to distributional factors: Only count nouns can be preceded by an indefinite determiner, only count nouns can appear in plural and singular,
and there are different quantifiers for count and mass nouns (much/many and little/few) (Gathercole, 1997; Gathercole et al. 1997).

Definite determiners (the ones that were used in the treatment) are compatible with mass and count interpretations in both Spanish and English. Singular definite determiners can occur in both languages with mass and count nouns. Definite determiners in Spanish also need to agree in number and gender with the noun they precede (el, la, los, las). A difference between Spanish and English is that in Spanish, singular definite determiners – the ones used in the experiment – can adopt a count and a mass interpretations, that is to say, the definite determiner in Spanish permits count and mass interpretations (example taken from Bosque, 1999, p. 11):²

(4) A María no le gusta el café

María does not like coffee

The sentence in (4) can be interpreted as mass suggesting that María does not like coffee in general, but also as count signifying that María does not like the specific cup that she is drinking. The fact that participants were presented only to

² Gathercole (1997) even doubts whether or not Spanish actually has a linguistic a count/mass distinction. She proposes that Spanish – contrary to English – does not obligatory assign nouns one category or the other: “In English, the mass / count distinction obligatory applies to every noun and determines the distributional properties for that noun, and it determines the way in which the referent of the noun is viewed by the speaker / hearer.”(p. 833)
singular nouns and that they appeared with definite determiners (that can receive both readings as count and mass) did not allow participants to make any form-meaning connection that identified nouns as part of one category or the other.

There is also the fact that the count-mass distinction does not relate to gender. As it was mentioned above, gender and animate-inanimate are related in the L1, English (e.g., pronouns he and she) and the L2, Spanish (e.g. articles él ‘he’ and ella ‘she’ and nouns niño and niña ‘girl’). Neither English nor Spanish associate the features count or mass with gender.

In conclusion, this research question intended to explore participants’ performance on the assessment tasks – recognition and production – with regard to the type of features. For the animate versus inanimate, the experimentally aware participants, who reached a higher level of awareness, correctly identify a similar number of animate and inanimate nouns since they seem to have acquired the system (MC recognition test: Animate $M = 85.71$ versus Inanimate $M = 79.77$; FIB production test: Animate $M = 78.57$ versus Inanimate $M = 73.81$). On the other hand, unaware participants performed differently according to the assessment tasks. For the recognition task, unaware participants correctly identify similar numbers of animate and inanimate nouns (Animate $M = 43.91$ versus Inanimate $M = 54.8$); for the production task – an assessment task that needs deeper processing of the L2 data – unaware participants correctly identify
significantly more animate than inanimate target items (Animate $M = 59.29$ versus Inanimate $M = 46.15$) (Table 42).

For the mass versus count experimental group, both groups – aware and unaware – performed similarly while it is noted that awareness reported was only at the level of noticing. On the recognition task, both groups correctly identify significantly more count than mass items (e.g., aware participants: Count $M = 65.9$ versus mass $M = 47.72$; unaware participants: Count $M = 61.84$ versus Mass $M = 36.4$). For the production task, the two groups did not significantly differ with regard to correctly identifying count and mass items (e.g., aware participants: Count $M = 47.72$ versus mass $M = 53.03$; unaware participants: Count $M = 49.56$ versus Mass $M = 53.03$). It is important to mention that regardless of the correct label, participants did not prefer to identify more inanimate or mass nouns (Table 43).

This was explained based on three assumptions (1) type of linguistic item or feature, (2) saliency, and (3) prefixal affix. Lardiere (2009a, 2009b) proposes that feature acquisition responds more to issues of “detectability” rather than a distinction between interpretable or uninterpretable features. For her, detectability is the base of learnability as “the learner will associate a difference in a minimally contrasting form with some difference in meaning or grammatical function and construct some sort of representation for it,” (Lardiere, 2009a, p. 214), such as in the case of “student – students” (Lardiere’s example, p. 214) or $pi$ versus $pu$ or $pi$
versus fi, in this experiment. The analysis of the think-aloud protocols revealed that aware participants reportedly were aware of the distinction between masculine and feminine and the affixal ending used to mark masculine and feminine gender during the assessment tasks (e.g., –i for masculine and –u for feminine). They also revealed that participants noticed or detected that there was a consonant at the beginning of the determiner system that carried a distinction, the j- and the r-, for the first experimental condition, and p- and f- for the second experimental group. Aware participants in the animate versus inanimate experimental group were aware of this distinction and they even commented upon this and/or verbalized the rule; in other words, they were able to explain the system according not only to the disclosed feature (e.g., gender) but also to the undisclosed feature (e.g., +/- animacy). Aware participants of the count versus mass experimental group commented that there was a distinction between the features that each consonant marked, nevertheless, they were not able to grasp the rule and as a result they were not able to verbalize or comment upon the rule (e.g., gender plus +/- count). They noticed the undisclosed feature (e.g., +/- count) mentioning that there was a distinction between the consonants (e.g., p- or f-), but they were not able to decipher the rule either implicitly or explicitly. Jiménez (2003) stated that studying the effects of attention and implicit and explicit learning: “could tell us something very important about which kind of regularities our cognitive systems are prepared to capture immediately, and which other
contingencies can be grasped exclusively by relying on a series of strategic, resource-demanding, and conscious operations” (p. 6-7). There is no doubt that the bundle of features [+/- animate] and gender were present in more meaningful contexts than the bundle of features [+/- count] and gender were better detected and acquired.

Research questions 4 and 5 will be discussed and analyzed in conjunction with Research question 6 since all of them address feature acquisition in regards to type of noun ending: canonical ending (–o for masculine, and –a for feminine) and arbitrary ending (those nouns ending in vowel –e or any other consonant). First, a concise summary of the findings will be presented.

Research question 4
Is type of noun ending (–o and –a) and (–e and –n) errors related to the features [+/- animacy] and [+/-count]?

Animate versus Inanimate

For the animate versus inanimate experimental group, results from the chi-squares for the recognition assessment task showed that participants were less accurate on arbitrary ending (–e and –n) than on canonical ending (–o and –a) on the composite score of aware and unaware participants ($\chi^2 43.45, df = 1, p < .001$)
as well as in the case of aware participants ($\chi^2 9.123, df = 1, p < .003$) and unaware participants ($\chi^2 35.55, df = 1, p < .001$).

These results were also confirmed for the production assessment task, in which the composite scores of aware and unaware participants ($\chi^2 45.07, df = 1, p < .001$) as well as the individual scores of the aware ($\chi^2 37.91, df = 1, p < .001$) and the unaware participants ($\chi^2 7.6, df = 1, p = .006$), were also found to be significant.

**Count versus Mass**

For the count versus mass experimental group, the composite scores of the aware and unaware participants were analyzed and were significant only on the production task ($\chi^2 48.89, df = 1, p < .001$) but not on the recognition task ($\chi^2 3.37, df = 1, p = .066$), which means that participants were less accurate on arbitrary ending [-e and -n] than on canonical ending [-o and -a] only in the case of the production but not on the production task.

In conclusion, for those participants in the animate versus inanimate experimental group, despite the type of task demands (recognition versus production) as well as (un)awareness condition, participants made more errors in type of noun ending arbitrary (e.g., (-e and -n)) than on canonical (e.g., (-o and -a)). However, for the count versus mass experimental group, the same is observed for the production assessment task but not for the recognition assessment task. As
mentioned above, on the recognition task both groups of participants, aware and unaware, had responded to a statistically higher number of count versus mass items. This may be the explanation as to why there was an association between type of feature (count and mass) and type of noun ending (canonical) and (arbitrary), although further investigation is needed.

**Research question 5**

Is type of gender [masculine] and [feminine] errors related to the features [+/- count] and [+/- animacy]?

**Animate versus Inanimate**

For the animate versus inanimate experimental group, participants on the recognition assessment task were statistically equally accurate on masculine nouns as on feminine nouns regardless of the endings of those words. The composite scores of the aware and unaware participants for the recognition task were not statistically significant ($\chi^2 = 2.266, df = 1, p = .606$); similarly no significant results were found for individual scores of the unaware ($\chi^2 = 2.399, df = 1, p = .528$) and aware participants ($\chi^2 = 2.596, df = 1, p = .440$).

For the production task (Tables 44, 45, and 46) similarly insignificant results for the composite scores of the aware and unaware participants for the recognition task were also found for both groups.
For the count versus mass experimental group, Tables 47 and 48 evidenced the same results, for the composite score of the aware and unaware participants ($\chi^2 = 12, df = 1, p = .729$) for the recognition and production task ($\chi^2 = .071, df = 1, p = .79$).

**Research question 6**

Does the type of ending canonical (–o or –a) and arbitrary (–e and –n) differ on (a) the recognition and (b) production task?

Gender is the appropriate space to test differences in the availability of features in the L1 and the L2. Grammatical gender in Spanish presents a challenge to English L1 speakers. Although English marks gender in the pronouns system, all nouns in Spanish are specified for genders. This grammatical gender poses a challenge to L1 English speakers: Abstract gender is not instantiated in English, therefore transfer is not available from the L1 (Keating, 2009). Is the acquisition of gender L2 features possible after puberty when they are not present in the repertoire of available features in the L1?

Two opposing positions attempt to explain this issue. Failed Functional Features Hypothesis (FFFH) claims that uninterpretable features are subject to a critical period, which leads to representation deficits in the L2 interlanguage grammars (Franceschina, 2001; Hawkins, 2001; Tsimpli, 2003). The Full Transfer Full Access hypothesis (FTFA) claims that uninterpretable features can be
acquired, and that L2 speakers have the same underlying grammatical representations (White, 2003). However, Lardiere (2009a, 2009b) has challenged these proposals. Lardiere (2009a) affirms that feature acquisition –interpretable and uninterpretable– seems to respond more to detectability peculiarities, assuming that the “basis for detectability is the observation of any formal contrast, such as the difference between student ~ students, or xuesheng ~ xueshengmen” (p. 214). The observation of these contrasts will lead the learner to associate them with some difference in meaning or grammatical function, and hence, construct a certain type of representation for it.

She eludes to state her definition of detectability; still, in the case of feature acquisition, she rejects any intrinsic distinction between interpretable and uninterpretable features: “My own view is that any feature contrast that is detectable is, in principle, ultimately acquirable (although it might not be actually acquired in any given particular case for independent reasons) (Lardiere, 2009a, p. 214).

The results from this dissertation seem to support claims like Lardiere’s (2008, 2009a, 2009b), who proposes that inaccuracies respond more to problems with mapping features to morphology rather than to an indication that there are impaired features (either interpretable or uninterpretable) or that they have not been acquired at all. Although these participants do not control agreement marking with a native-like accuracy, and proficiency level seems to be a variable
that needs to be considered, participants in this experiment demonstrated gender with a higher level of accuracy despite the fact that their L1 lacks a [gender] feature for nouns and adjectives. Gender errors seem to be more of a production problem and not a representational problem or the lack of a formal feature, as the small number of gender errors in the gender assessment task seems to show.

Before continuing with the analysis of the errors, it is necessary to mention that the percentage of errors on the gender assessment task in which participants had to assign the correct determiner to each of the nouns that were used in the assessment tasks (multiple-choice and fill-in-the-blank) was on average 97.56% for the animate versus inanimate and 98.08% for the count versus mass. This means that (1) participants in both experimental conditions had the same baseline, and (2) participants knew the gender of the words they were tested on. This increases the reliability of the assessment tasks that were part of the experiment since it was proved that participants were familiar with the words and correctly assigned gender on those words with a high accuracy. With regard to gender assignment, these results are consistent with other studies that found that determiner + noun gender agreement errors are fewer, compared with noun and adjective agreement (Bruhn de Garavito and White, 2002; Fernández, 1999; Franceschina, 2001). For example, Franceschina (2001) reports an accuracy of gender assignment of 94.2% on articles compared with an accuracy of 77.36% on adjectives. This 94.2% on gender assignment on determiners is similar to the
accuracy of 97.56% for the animate and inanimate experimental group and 98.08% of the count versus mass experimental group that was reported in this experiment for the gender assignment test. These percentages varied after participants completed the experiment as it is reported below and possible interpretations are offered.

The results from Research question 4 and Research question 5, which addressed errors only on noun ending and gender for each experimental condition separately, indicate that participants relied on morphosyntactic cues, e.g., word endings, as cues to assign gender favoring the (–o and –a) over (–e and –n). Research question 5 addressed gender assignment based on the intrinsic gender of the noun regardless of the ending. That is to say, it is not the gender of the nouns itself (masculine versus feminine) that seems to explain the errors but noun ending, as Research question 6 evidenced. Errors in gender assignment are not statistically significant when masculine nouns versus feminine nouns are compared (Research question 5), but only when the noun endings are compared (–o and –a) versus (–e and –n) (Research question 6). It is noun ending, then, and not the gender of nouns, that can be seen as the factor that explains errors in gender assignment. Participants seem to have had a harder time assigning gender to nouns not ending in (–o and –a). These results would thus to confirm that participants relied on noun ending cues to assign gender.
Cain, Weber-Olsen and Smith (1987) investigated learners of L2 Spanish and Spanish-speaking children to compare whether the adult learners of L2 Spanish employed the same strategies for acquiring Spanish gender as the ones that Spanish-speaking children do. The participants were 20 children, ages 3-12, and 20 college-age adults L2 learners of Spanish. Results that are pertinent to this study indicated that morphophonological clues (nouns endings –o, –a) carried more weight than semantic clues (sex of the noun), but the importance of the semantic clues increased according to age for the L1 and proficiency level for the L2. In the same vein, Sagarra and Herschensohn (2010) found that intermediate students were less accurate in gender assignment with animate than with inanimate nouns. The former nouns are cognitive more demanding for gender assignment since they usually activate the masculine and the feminine form, e.g., 
*esposo* (‘husband’) and *esposa* (‘wife’).

The errors of the two experimental conditions were analyzed in order to explore more deeply the variability found between the low percentages of errors produced on the gender questionnaire completed at the end of the experiment and the more numerous gender errors participants produced on the assessment tasks during the experiment. For the animate versus inanimate experimental condition it was found that 9% of the errors for the multiple-choice and 32% of errors for the fill-in-the-blank consisted of words ending in –o or –a (20.5 % on average for both assessment tasks). However, for the count versus mass experimental
condition, this trend was reversed: the majority of the errors concentrated on words’ endings –o and –a: 55% for the multiple-choice and 60% for the fill-in-the-blank (on average 57.5 for both tasks). These results evidenced that for both groups adding a feature (i.e., [+/- animate] or [+/-mass]) modified the organization of a system that clearly responded to morphophonological clues (–o and –a versus –e and consonant), and as a result those canonical clues (–o and –a) were no longer as “visible.” Participants seem to be in a process of feature reassembly in which they test the values of the new features that translate in erratic behavior (cf. Table 70 and Table 71). Nevertheless, both experimental groups did not behave completely identically. The experimental group [+/- count] produced considerably more (–o and –a) errors than the [+/- animate] experimental group: 57.5% versus 20.5% of errors, respectively. This once again provides evidence to assert that both types of features are not equal. The bundle of features [+/- count] plus [+/- feminine] seems to result in an overload of information for the L2 learner, which may have led to a domino effect that misadjusts and obscures some cues (noun endings) that are the ones that override determiner selection.
Table 71

*Errors according to the noun endings Animate versus Inanimate experimental group*

<table>
<thead>
<tr>
<th>Assessment tasks</th>
<th>Gender questionnaire</th>
<th>Multiple-choice (Recognition Task)</th>
<th>Fill-in-the-blank (Production task)</th>
</tr>
</thead>
<tbody>
<tr>
<td>–o, –a</td>
<td>–o, –a</td>
<td>–o, –a</td>
<td></td>
</tr>
<tr>
<td>3%</td>
<td>9%</td>
<td>32%</td>
<td></td>
</tr>
<tr>
<td>–e, –n</td>
<td>–e, –n</td>
<td>–e, –n</td>
<td></td>
</tr>
<tr>
<td>97%</td>
<td>91%</td>
<td>68%</td>
<td></td>
</tr>
</tbody>
</table>

Key: n = any consonant

Table 72

*Errors according to the noun endings Count versus Mass experimental group*

<table>
<thead>
<tr>
<th>Assessment tasks</th>
<th>Gender questionnaire</th>
<th>Multiple-choice (Recognition Task)</th>
<th>Fill-in-the-blank (Production task)</th>
</tr>
</thead>
<tbody>
<tr>
<td>–o, –a</td>
<td>–o, –a</td>
<td>–o, –a</td>
<td></td>
</tr>
<tr>
<td>25%</td>
<td>55%</td>
<td>60%</td>
<td></td>
</tr>
<tr>
<td>–e, –n</td>
<td>–e, –n</td>
<td>–e, –n</td>
<td></td>
</tr>
<tr>
<td>75%</td>
<td>45%</td>
<td>40%</td>
<td></td>
</tr>
</tbody>
</table>

Key: n = any consonant

Participants in this study seemed to have relied on morphophonological clues; however, considering the low percentage of errors in nouns ending in either –e or any other consonant, and the variability – in some cases they made the correct gender agreement on the gender questionnaire while in others the same word was incorrect on the assessment tasks – , it is possible to conclude that participants seemed to be in a process of relying on all available clues – semantic, morphophonological and syntactic clues – to better assign gender. It may be
concluded that participants seemed to have acquired a gender [feature] and consistently assigned gender correctly; the existent errors seemed to be the consequence of mapping features to form (Lardiere, 2008, 2009a, 2009b).

**Conclusions, Limitations and Future Research**

**Overview**

This section summarizes the results of this dissertation presented in the previous sections and proposes some theoretical and pedagogical implications drawn from this study.

**Conclusions**

While models and hypotheses in second language acquisition based on cognitive psychology propose that attention is necessary for learning to take place (Robinson, 1995b; Schmidt, 1990, 1993, 1994, 2001; Tomlin & Villa, 1994), the role of awareness in further intake processing or learning is the object of current scrutiny in second language acquisition research. Schmidt’s noticing hypothesis (1990) supports the idea that focal attention is isomorphic with awareness, and that intake is not possible without some level of awareness present during attention paid. This hypothesis suggests the existence of two levels of awareness, (1) awareness at the level of noticing, and (2) awareness at the level of understanding. The former relates with intake, and the latter with system learning or restructuring. Tomlin and Villa (1994), on the other hand, categorize attention
as a process dissociated from awareness. The closest construct to noticing is their attentional function of detection that is within selective attention, although it does not necessarily involve awareness. They even recognize a difference between detection with or without awareness for those cases that do not implicate awareness or for processing of non-target items.

Hama and Leow (2010), Leow (1997, 1998b, 1998b, 2000), Rosa and O’Neill (1999), Rosa and Leow (2004) have incorporated definitions of noticing, attention and awareness. Results from these studies incorporating verbal reports to operationalize and measure awareness have found that attention is necessary for intake processing and that there is no evidence of dissociation between awareness and intake or learning. Williams (2004, 2005), and Leung and Williams (2011a, 2011b), on the other hand, have presented evidence of mapping form and meaning connections in the absence of awareness. These contradictory results validated a further investigation of this contentious issue.

On the other hand, second language learners’ interlanguage offers a high degree of variability in their productions. UG explanations assume that this erraticism can be explained based on the availability or lack thereof of different types of features, L2 interpretable features are available even if they are not part of the repertoire of features in L1, while L2 uninterpretable features are not available for further acquisitions unless they also exist in the L1 (Hawkins, 2003; Franceschina, 2001). Other proposals assume that the variability can be attributed
to difficulties learners experience in mapping abstract representations to the morphological surface form (Lardiere, 2007, 2008, 2009a, 2009b). Lardiere (2009a, 2009b) proposes that a more cogent explanation of feature acquisition is to understand learnability as a consequence of saliency and cues detection, and not as the result of any intrinsic dissimilarity between interpretable and uninterpretable features.

The main independent variables of this study were type of awareness (unaware versus aware), type of grammatical feature [+/− animacy], [+/− count], and gender [+/− feminine]. Within the context of a computer delivered task, participants read two texts for meaning in which the target items were equally distributed throughout the texts. This dissertation investigated the immediate and delayed effects of the provision of two different experimental conditions (animate versus inanimate and count versus mass), on the performances of new and old items presented in the experimental texts as measured on two assessment tasks: a multiple-choice (as a recognition task) and a fill-in-the-blank (as a production task). It also addressed the acquisition of the features [+/− animate], [+/− count] as well as [+/− feminine] under the UG perspective. Before any analysis of the responses on the assessment tasks were conducted, participants were grouped into aware and unaware based on the use of a hybrid design that operationalized and measured the construct awareness at both a concurrent (think aloud protocols) and nonconcurrent (postexposure questionnaire) stage in an effort to triangulate
information at different phases of the treatment and different stages of the learning process.

On the basis of the results reported in previous chapters, it is possible to enumerate the following conclusions:

1. Awareness seems to be beneficial for further processing and learning of the target items; however, level of awareness is a factor that needs to be considered according to what the think-alouds appeared to indicate. Higher levels of awareness, as measured by the level of awareness reported in the online think-aloud protocols and the post-exposure questionnaire, seemed to result beneficial for intake and learning of old and novel items; lower levels of awareness resulted beneficial on receptive tasks and for old items only. These conclusions needs to be taken cautiously since the main purpose of the dissertation did not address levels of awareness as a independent variable.

2. This dissertation found a disassociation between unawareness and further processing or learning of the target items. Unaware participants performed significantly better only on old or trained items, which means that memory played a role.

3. Regarding the effects over time, from the immediate posttest to the delayed posttest one-week later, aware participants under the animate versus inanimate experimental group maintained the gains although this
resulted beneficial only for those participants who reported a higher level of awareness (at the level of understanding).

4. The qualitative analysis of the think-aloud protocols revealed that participants adopted two different strategies while performing the treatment and the assessment phases of the experiment. Data-driven and conceptually-driven processes have been previously studied in the literature (Leow, 1998; Robinson, 1995b; Williams, 1999), and it is well known that they are the result of task-demands and individual differences; however, it is worthy to consider another variable, namely, type of linguistic item or feature, which might contribute to the usage of one or the other. It is noted that there were more numerous participants that adopted a conceptually-driven process in the animate versus inanimate condition while none of the participants in the count versus mass condition adopted such a process and solely relied on data-driven strategies.

5. Unaware participants responded correctly significantly more to animate than inanimate nouns in the fill-in-the-blank production assessment task, and significantly more to count than mass nouns on the recognition assessment task.

6. Regardless of levels of experimental condition, gender was considerably well supplied and errors on both assessment tasks were not numerous, a
result that was confirmed by the gender test in which the 48 words that were part of the assessment tasks were tested to check whether or not participants could assign the correct gender to the words being tested on the assessment tasks. The average score on that gender test was substantially high, 98%, which is similar to previous studies that have proved that gender errors are less common in the determiner and noun agreement than in noun and adjective agreement.

7. Regardless of awareness (aware or unaware) or type of feature ([+/- animate] or [+/- count]), participants seemed to rely on ending cues (i.e., -o for masculine and –a for feminine) for gender assignment. This explains why gender assignment errors were not significant for the distinction [+/- feminine], and they were only significant when type of noun ending was explored canonically (–o and –a) and arbitrarily (–e and –n).

8. Studies addressing gender acquisition have stated that there is some parallelism between L1 and L2 gender acquisition. L2 learners seem to follow the same stages that L1 children demonstrate while acquiring gender assignment. In both cases they rely more on morphophonological clues at earlier stages and then move to syntactic clues as they are evolving (Franceschina, 2005).
9. An analysis by gender of the errors produced on the assessment tasks found that the morphophonological clues that override gender assignment became less evident. The percentage of errors on nouns ending on –o and –a for masculine increased for both experimental conditions after completing the experiment; however, the [+/- count] experimental group presented significantly more gender errors –o and –a than the [+/- animate] experimental group. Participants in the [+/- count] experimental condition became more oblivious of the morphophonological clues than participants in the [+/-animate] experimental condition. It seems that the former bundle of features results in an overload of information.

**Pedagogical implications**

Although this dissertation developed a semi-artificial determiner system, the results obtained in this dissertation present some pedagogical implications. Since this dissertation found no dissociation between awareness and intake or learning, it is central for material developers and instructors to recognize the importance of the constructs of attention and awareness. Some researchers, especially those that support some sort of implicit learning, consider that implicit learning is beneficial for complex structures (Reber 1976, 1993). Robinson (1996), for example, tested this assumption and found that the implicit condition was second best out of four conditions (the other being incidental, enhanced and instructed) and instructed
(i.e., explicit) worst out of the four for easy rules. However, individual differences must be taken in mind since rule difficulty is an “individual issue” according to DeKeyser (2003): “What is a rule moderate difficulty for one student may be easy for a student with more language learning aptitude or language learning experience” (p. 331). According to DeKeyser (2003, p. 334) “The harder it is to learn something through simple association, because it is too abstract, too distant, too rare, too unreliable, or too hard to notice, the more important explicit learning processes become.” And this is precisely what was found; participants in the animate versus inanimate experimental condition were able to decipher the semi-artificial determiner rule with more consistency than those participants in the count versus mass experimental condition. Participants in the former experimental condition were able to transform the implicit exposure into explicit knowledge; participants in the latter experimental condition were not able to do so, and as a result, a more explicit form of instruction could have resulted more beneficial.

In either case, knowing the importance of attention and awareness in language intake and possible further processing and learning is necessary to design pedagogical tasks that draw learners’ attention and help to raise awareness on those types of linguistic features of the target language.
Limitations and Future Research

Limitations

The following dissertation presents a series of limitations that will be enumerated here.

The first limitation refers to the cell size. The original number of participants was 117 participants and attrition reduced the number to 62. Due to the different levels of (un)awareness, uneven cell distribution resulted. In the animate versus inanimate experimental group, the animate group contained $N = 33$ participants, who were distributed into $N = 26$ unaware participants and $N = 7$ aware participants. The count versus mass experimental condition ($N = 29$) resulted in $N = 18$ unaware participants and $N = 11$ aware participants. A more evenly distributed number of participants would have resulted in greater statistical power, especially when group comparisons needed to be carried out. A larger number of participants (and therefore cells more evenly distributed) increases the statistical power (Field, 2005).

A second limitation relates to the number of target items. 24 target items were carefully chosen intending to have those nouns representing prototypical instances of their categories: animate, inanimate, count, and mass as well as masculine and feminine. Nevertheless, although it was also intended to select the nouns according to their morphophonological clues (e.g., noun endings), it was not possible to have exactly the same instances of nouns ending in $-o$, $-a$, $-e$, or
any other consonant across all the categories. The mismatches in numbers were only two or three elements at a maximum, but in order to be able to carry out finer comparisons a larger number of elements is needed.

Finally, a third limitation is the delayed post-test. Even though this dissertation is the first study addressing the effects over time of the form-meaning mappings, a one-week period may be viewed as being relatively brief. Future studies should consider longer-term delayed posttests.

Future research

A fruitful future research refers to individual differences. Williams (2004, 2005) and Faretta-Stutenberg and Morgan-Short (2011) have addressed the possible impact of individual differences in implicit learning. Due to the homogenous nature of the population that participated in this study, it was not possible to address individual differences. Williams (2004, 2005) found a correlation between L1 knowledge of languages that encode gender and implicit learning of form and meaning connections. Since one of the main rationales of this study was the analysis of L1 versus L2 languages with different feature availability, it was clearly intended not to use participants with L1 that encoded gender. Another individual difference that has been explored is years of instruction. This population, having the same background, had on average the same length of instruction and therefore this variable could not be explored either.
A different individual difference that could prove fruitful to the study is working memory. There are ample studies that have posited that there is some evidence between working memory and SLA early stages such as input processing strategies (e.g., segmentation), noticing and working memory (Dörnyei & Skehan, 2003). Knowing the importance of individual differences in SLA processing and acquisition, empirical studies should include these variables in further research.

Another fruitful area for future research in relation to gender assignment in combination with the animacy feature is that proposed by Sagarra and Herschensohn (2011), who found that processing of animate nouns is cognitively more demanding than that of inanimate nouns for basic and intermediate students since they also presented lower accuracy rates in sentences with animate than with inanimate nouns. Although the purpose of this study was not reaction times and this study found that participants errors varied according to type of ending (canonical versus arbitrary) and not according to type of feature [+/-animate], in future research it can be possible to study this effect by having exactly equal number of animate and inanimate nouns that have exactly the same number of nouns ending in canonical (e.g. transparent) and arbitrary (e.g., opaque) endings. These researchers investigated the noun endings using the same distinction canonical (–o and –a) and arbitrary (–e and –n) with a different label transparent and arbitrary, respectively.
APPENDIX A

Instructions

Instructions for the animate vs. inanimate experimental condition
Texts 1 and 2

This is an experiment about how people formulate sentences in different languages. Different languages have different ways of saying things, which raises the question of whether this forces people to think slightly differently in order to speak and understand in different languages. Here I am investigating this issue in a situation where the sentences are almost entirely in Spanish, apart from the articles.

When you read the text below, please note that the Spanish articles *el* or *la* are not used but instead…
Masculine articles correspond to
JI and RI and
Feminine articles correspond to
JU and RU

Please keep this in mind as you read the texts for the cultural information reported in them. You will be asked to answer questions based on the content of the texts.

Instructions for the count vs. mass experimental condition

Texts 3 and 4

This is an experiment about how people formulate sentences in different languages. Different languages have different ways of saying things, which raises the question of whether this forces people to think slightly differently in order to speak and understand in different languages. Here I am investigating this issue in a situation where the sentences are almost entirely in Spanish, apart from the articles.

When you read the text below, please note that the Spanish articles *el* or *la* are not used but instead…
Masculine articles correspond to
FI and PI and
Feminine articles correspond to
FU and PU
Please keep this in mind as you read the texts for the cultural information reported in them. You will be asked to answer questions based on the content of the texts.
APPENDIX B
Experimental Texts
*The experimental texts for the animate vs. inanimate experimental condition*

Text 1
Ju familia: Definición
De acuerdo a ru Declaración Universal de Derechos de ji hombre y ju mujer, ru familia es ri elemento natural y fundamental de cualquier sociedad. Por ello, ri Estado debe garantizar ri respeto de ji miembro que pertenece a ella. Ésta se define a partir de dos tipos de unión:

a) *unión de afinidad*: surge entre ji esposo y ju esposa, mediante una unión matrimonial que suele habitar en ru casa familiar; y b) *unión de consanguinidad* (es decir, de sangre): ri vínculo entre ji padre y ju madre para ru crianza de ji hijo y/o ju hija o aquella unión que se establece entre ji niño y/o ju niña que descienden de un mismo padre.

Ru *institución familiar* es ru cara social de ju persona que vive en ru ciudad o ri campo. En Latinoamérica incluye no solo a padres e hijos sino a ju abuela, ji abuelo o ji padrino, esto independentemente de ru dinero o clase social a que se pertenezca. No hay consenso sobre cuál es ri factor que la define; sin embargo, en realidad ésta es muy diversa y tiende a cambiar con ri tiempo.

Familia extensa en Latinoamérica
No existe otra institución cultural más importante en América Latina. Factores sociales como ru solidaridad, ru educación y ri amor desembocan y se combinan entre sí, surgiendo tensión entre ji esposo y ju esposa.

Típicamente una familia latinoamericana que vive en ri campo o en ru ciudad es extensa, independientemente de ru clase social o de ri dinero que cuente; incluye en ri seno familiar no solamente a ji padre, ju madre y ji hijo y ju hija, sino también a ji abuelo y ju abuela y a ji tío y y ju tía, y hasta ji padrino o ju madrina, quienes pueden intervenir en ru crianza y ru educación. Por lo tanto, en
algunas ocasiones esto ofrece una oportunidad para que ju persona conviva en ru casa con tres generaciones. Ri tiempo que conviven ji niño, ju niña y ji mayor es ri factor que ayuda a mantener ri respeto y ri amor familiar y explica que ri vínculo entre ellos sea muy fuerte.

Ru familia extensa es de fundamental importancia social. En ella ji hombre y ju mujer encuentran apoyo moral, ya que casi siempre se cuenta con algún pariente, generalmente femenino: madre o abuela, que aconseja o reconforta. Dentro de de ella, ji miembro aprende valores como ru solidaridad, entre otros.

Normalmente, ji abuelo es importante y ayuda en ru crianza y ru educación de sus nietos, por ello son respetados y no son considerados una carga. Por otro lado, ji niño o ju niña no son ri centro de atención y aprenden desde muy jóvenes ri respeto a sus mayores. Cuando ji padre y ju madre salen a una fiesta, usualmente llevan a sus hijos consigo. Cuando eso no es apropiado, pueden dejarlos con ju abuela, ji pariente o, incluso, ju empleada en ru casa.

Otra característica de ru familia extensa es que se ayudan mucho entre ellos, por ejemplo, en caso de que se busque empleo, ascenso en su trabajo o recomendaciones para encontrar uno. Algunos ámbitos profesionales y políticos en América Latina, principalmente cuando se trata de ji hombre, pero también de ju mujer, todavía se conducen en su mayor parte según ri vínculo que ofrece ri personalismo. Es decir, que si ji esposo, ju esposa, ji hijo o ju hija, por ejemplo, se quedan sin trabajo, en caso de que se reciba una recomendación de ji tío o ji padrino, se mejora significativamente ru posibilidad de conseguir trabajo. Una recomendación de un pariente o amigo vale mucho más que aquella recomendación de ju persona desconocida. Por ello, de alguna manera, ru solidaridad y ri amor familiar aseguran un estatus económico similar y se mantiene ri dinero en unas pocas familias. No queremos sugerir que no hay mérito individual o que sólo este fenómeno llamado personalismo explique cómo
funciona en la sociedad latinoamericana. Lo que queremos señalar es que, a pesar de algunos cambios en el tiempo, en América Latina el personalismo es casi siempre el factor que explica porque un miembro familiar que vive en la ciudad consigue un empleo.

Text 2
Cambio familiar
En el tiempo actual, la realidad familiar latinoamericana, principalmente de aquella que vive en la ciudad, es un fenómeno en transición, todavía cambiante, resultado de una transición de valores tradicionales, propios de modelos familiares fijos, a valores nuevos que promueven el respeto, la equidad y el amor entre quienes cohabitan en la casa familiar. Entre estas ideas que promueven una nueva relación entre sus miembros están: Igualdad de la mujer y el hombre, una nueva relación con el padre y la hija y sobre todo un ideal de que la educación es el factor que permite una mejor vida.

Este cambio ha modificado instituciones que forman parte de la familia. De cualquier modo, el esposo y la esposa se empeñan en apoyarse para la crianza de los niños y en mejorar el vínculo con el miembro familiar. La madre y la abuela cumplen un papel central en todo ello. A continuación revisaremos algunos aspectos de su organización y su estado actual y cómo afectan a la persona que vive en familia.

Proceso de cambio: Introducción
En el tiempo pasado, el patriarcado era la organización familiar tradicional. La esposa e hijos se subordinaban a un miembro masculino: el esposo y el padre.

También era de gran importancia, una definición estricta de la papel sexual de la persona, principalmente de la madre o la hija. El factor más importante para que el hombre se ganara el respeto era ser un buen proveedor, es decir, el dinero
que traía a casa. Ju mujer tenía un papel muy destacado para ofrecer ri amor, ayudar en ru casa y en ru crianza y ru educación de ji niño y ju niña.

Otro aspecto importante de ru vida familiar es ri compadrazgo. Mediante él, ru solidaridad se extiende más allá de vínculos de sangre o de casamiento ya que ji padrino y ju madrina se vuelven parte integral de ru familia. Generalmente es espera que ju comadre y ji compadre den regalos y tengan atenciones especiales, por ejemplo, ayuden con ri dinero para celebrar un día especial. Ji padrino y ju madrina son unas personas especiales que brindan apoyo.

Ri madresolterismo

Ri vínculo y ru solidaridad familiar entre ji miembro y sus demás familiares siguen intactos y fuertes en muchas partes de América Latina. Sin embargo, para millones de personas, todo es diferente. Por ejemplo, en ru clase social baja, tanto en ru campo como en ru ciudad, en ocasiones ji padre abandona sus responsabilidades con ju esposa y sus hijas por cuestiones económicas, sobre todo cuando ju pareja no está casada legalmente. Ri madresolterismo –ese escenario en que ju madre, sin ji esposo, es responsable de sus hijos– es común en toda América Latina sobre todo en clases bajas. Ju mujer invierte tiempo trabajando para dar de comer a ji niño y/o ju niña, quienes raras veces pueden asistir regularmente a ru escuela. Por ello, ru familia en que habitan estos niños, carece de una figura paterna, y en cambio, se cuenta con sólo una madre, y, en ocasiones, ju abuela que apoya a su hija en ru crianza. Ri amor materno representado en esas dos mujeres ayuda a sacar adelante esas familias.

Ji joven urbano

Una marcada desigualdad socioeconómica es ri factor que mayormente afectó a ri vínculo familiar principalmente a partir de 1980. Actualmente se calcula que dos tercios en América Latina son pobres y que 40% tiene menos de 15 años de edad.

Con estos datos, se puede apreciar que una porción significativa se compone de jóvenes de familias sin ri dinero suficiente. Ji joven carece de ru
educación mínima para superar su situación económica y socialmente inferior. En 
ri campo, ju hija tiene que ayudar a sus padres con trabajar dentro y fuera de ru 
casa. En ru ciudad mediana o grande hay incluso jóvenes que viven sin sus padres 
y no tienen casa. Sin ju persona extra que brinde ru solidaridad necesaria, como ji 
padrino o ju madrina o ji abuelo o ju abuela, ri tiempo libre no es productivo y 
enseñanza de valores como ri respeto es casi nula.

The experimental texts for the count vs. mass experimental condition
Text 3
Pasado y presente: cambios individuales y familiares
Frente a tantas transformaciones, tanto fi hombre como fu mujer se encuentran 
ahora ante un dilema: escoger entre pi progreso de pi presente y fu imagen 
idealizada de pi pasado. Estas tensiones son vividas sobre todo por aquellos que 
viven en pu ciudad, y principalmente por aquellos que pertenecen a fu clase 
media. En muchos casos, este choque y cambio de valores produce tensiones y 
confusión de papeles entre pareja, lo cual se relaciona con un aumento en fi 
divorcio.

Muchas familias están tratando de acomodarse a fu época nueva y 
cambiante, entre estos dilemas se encuentran, cómo dar forma a pi amor de pareja, 
qué papeles deben desempeñar fi padre y fu madre y cuál es pu importancia actual 
de fi vínculo familiar. Un ejemplo de esto es fu familia profesional de clase 
media, para quien pu infancia es muy valiosa y, por ello, buscan pu protección y fi 
cuidado de sus hijos. Pu educación y fu expectativa de ambos sexos a seguir 
trabajando permiten contar con pi dinero suficiente para pu crianza de fi hijo y/o 
fi hija y garantizan pi respeto y pu solidaridad entre sus miembros. Estas familias, 
que viven con optimismo pi futuro, representan en realidad un porcentaje menor. 
En su mayoría estas familias han sido víctimas de pu pobreza, pi desempleo y de 
otras consecuencias económicas negativas que afectan gravemente pi 
comportamiento de sus miembros.
Otras formas de unión familiar: unión libre y divorcio
Además, a pesar de lo que se piensa, su aceptación en el presente de la unión libre –aquella unión en que el hombre y la mujer no están casados pero viven como si lo estuvieran– es cada vez mayor: hay datos con porcentajes que van desde un 4% en Chile y 6% en Brasil, hasta un 24% en México y un 47% en Guatemala. Para estas personas, el amor y el respeto entre ellos es lo que valida su relación más que otro tipo de aprobación externa.

En su época actual, el divorcio, ahora legalizado y que cuenta con la protección y el cuidado del Estado, puede dar lugar a matrimonios sucesivos en el futuro. El dinero, que ofrece la estabilidad económica, no es la causa principal para casarse ya que en su clase media y alta tanto el esposo como la esposa cuentan con ingresos. El vínculo familiar ha dejado de ser principal lugar para la educación y la crianza de valores morales, ni tampoco el padre ni la madre son quienes únicamente los transmiten. La importancia de lazo de sangre, es sustituido por relaciones de compañerismo y de amistad selectiva a nivel personal.
Tanto el hijo como la hija establecen desde edades tempranas relaciones muy cercanas con distintas personas y la solidaridad familiar se extiende fuera de la familia. Todo ello ha dado lugar a una diversa concepción familiar: en ella aparece un grado mayor de informalidad, lo que modifica su comportamiento y permite nuevas formas de integración familiar.

El vínculo familiar urbano y el cuidado de la infancia
Problemas económicos ocasionan que la clase media también sienta presión por el dinero. Hoy en día, por ejemplo, la madre sale a trabajar y ayuda económicamente. Esto le permite estar en una posición de igualdad y compartir con su esposo acuerdos familiares, por ejemplo, cuántos hijos se desean tener, así como tener un papel más activo en la crianza y el tiempo que se pasa con ellos. Esto disminuye su importancia de patriarcado y de su papel central de padre.
Dentro de la familia urbana en la época actual se presenta una dinámica en que el niño y la niña aprenden ciertas concepciones sobre género; éstas se reflejan en una división de tareas que cumple cada miembro. Mediante la educación recibida en infancia desde que son niños, se les enseña el comportamiento esperado en sociedad.

Conforme el niño y la niña toman consciencia de su sexo temprano en su infancia, hacen propias ciertas concepciones sobre el hombre y la mujer, por lo que desde niños comienzan a mostrar diferencias de género. Antes era más común que la niña ayudara en actividades domésticas, tal como, el lavado de ropa, preparación de comida, limpieza, etc.; mientras que al niño se le fomentara actividades de mayor independencia, y que incluso se le diera alguna responsabilidad como la protección de su hermana. Todo esto en el presente está cambiando y es importante mencionar que el hijo y la hija aprenden igualdad de género, reflejado en valores como el amor, el respeto y la solidaridad, para lo cual hay la esperanza de que todo esto se pueda observar en el futuro.

Text 4
Algunos cambios actuales

El vínculo familiar urbano de clase media se está haciendo más nuclear. Para evitar que en el futuro haya menos contacto entre sus miembros y para mantener el afecto y la solidaridad familiar, la mujer, más comúnmente que el hombre, se mantiene en contacto por teléfono, reuniéndose a tomar un café y arreglando algún problema familiar. Así la madre mantiene vivo su amor y su protección familiar y se adapta a su presente.

Familia urbana en la época actual puede recurrir a otras formas para cuidar de sus niños durante su infancia, dejándolos con algún familiar o incluso con alguna empleada de confianza, quien interviene así en su crianza y su educación de un niño o una niña de sus jefes. Sin embargo, estas opciones casi no
existen para la obrera, quien no tiene dinero para pagar una empleada, y
tampoco recibe ayuda de su esposo en el cuidado de sus hijos. De ahí, que estos
tipos de centros infantiles, existentes sobre todo en la ciudad, permiten a la clase
obrera contar con un lugar para que le cuide a su hijo o hija. Esto trae algunos
cambios; valores como la solidaridad o el respeto serán aprendidos no sólo a
partir de una visión familiar sino a partir de una más social.

Otra innovación actual que afecta el comportamiento familiar es el hogar
para adultos mayores donde se brinda cuidado a ancianos. Espacios más pequeños
para vivir ocasionan que el adulto mayor viva separado de sus hijos y que la
importancia de éste disminuya en el círculo familiar.

Cambios en la maternidad y en la paternidad
Cambios en el vínculo familiar causan una nueva relación entre el esposo y la
esposa de la clase media y sus. Con una autoridad que no es exclusivamente
paterna, con un menor apoyo de la familia extensa y principalmente con menos
hijos, la importancia de ellos es mucho mayor en el presente que antes. Cada
etapa desde la infancia hasta la adolescencia es muy importante y se busca
garantizar la protección y el cuidado de ellas. Esto tiene ventajas, desde luego,
pero a veces impone mucha presión en el hombre o la mujer para salir bien en el
estudio, para dar validez a su dinero invertido y el amor de sus padres. Por otro
lado, el padre y la madre no han soltado totalmente el mando tradicional y buscan
que la crianza y la educación de sus hijos sea de lo mejor; sin embargo, también
quieren influir en el comportamiento de sus hijos. Por una parte, quieren enseñar
valores más acordes con la época actual, como la independencia y el
individualismo; pero por otra parte, quieren vigilar el tiempo, la vida social y
sexual de su hijo y, principalmente, de su hija. Obviamente, esta contradicción
causa muchos conflictos entre el adolescente y el padre y la madre. En caso de que
haya confianza, algún familiar puede intervenir para ayudar a solucionar
conflictos en pi futuro. Sin embargo, estos problemas son también ejemplos de una tendencia hacia relaciones familiares más abiertas y democráticas entre sus miembros, en que se busca pi respeto entre todos.

Hacia un nuevo sentido familiar

A pesar de problemas que afectan a fu clase baja y media, principalmente, no puede decirse, sin embargo, que fu vida familiar esté en proceso de desaparición; al contrario, hay una revalorización de fu unión familiar como ese espacio para pu protección y fi cuidado de fi individuo. Estamos, pues, ante un panorama de cambios y de expectativas inquietantes, pero también en una época en que pu esperanza se mantiene.

En resumen, en pi presente pu importancia de fi vínculo familiar y pi compadrazgo siguen siendo fuertes en América Latina aunque se enfrentan a intensos cambios. Sin embargo, por importantes que sean, en fu época actual representan solamente una parte. Para lograr una visión actual de fu familia, hay que tomar en cuenta también cambios en pu educación en general, igual acceso a pi dinero de ambos esposos. También es necesario incluir cambios en pu crianza de fi hijo y fu hija que reflejan una importancia mayor de pu infancia, así como cambios en pu masculinidad que vive fi hombre, en su doble papel: fi padre y fi esposo y, sobre todo, pi feminismo que modificó pi comportamiento de fu mujer.

Lo aquí expuesto refleja desafíos para pi futuro, sin embargo, pi respeto, pi amor, pu solidaridad y fi apoyo entre sus miembros, a pesar de todo, sigue siendo igual.
APPENDIX C

Instructions to think-aloud
In this experiment, I am interested in what your thoughts are as you perform some tasks that I give you. In order to do this, I am going to ask you to THINK ALOUD as you do the different tasks. What I mean by “think-aloud” is that I want you to say aloud everything you would silently say to yourself. When you are instructed to begin, I would like you to begin thinking aloud, that is, verbalize your thoughts, and continue to do so the entire time as you are doing the tasks. Speak in ENGLISH.

These are examples of a think-aloud:
If I were working through a math problem, it might sound like this:
24 + (8 x 4) =
“OK. I have to multiply first…so eight times four… um, that’s 32…. Add 24 to that… OK. Two plus four is six… three plus two is five… so, uh, 56, yeah, 56.”

Now, please practice to think-aloud with these examples. They are math problems and verbal tasks. Say aloud whatever calls your attention. Use the headphones and the microphone. Please speak LOUDLY and CLEARLY enough as much as possible so that the recording will be audible to the researcher. Don’t worry if you don’t know the answer. The purpose of the exercise is to think-aloud:

1. 45 - (7 x 3) =
2. 24 + (2 x 6) =
3. 30 - (7 + 3) =
4. 15 - (10 - 2) =

Please speak LOUDLY and CLEARLY enough so that the recording will be audible to the researcher. Do not mumble.
APPENDIX D

Comprehension Questionnaires

*Comprehension questionnaires for the animate vs. inanimate experimental condition*

Text 1

Comprehension Questions

Please answer the following questions based on the reading. Choose the answer that best reflects the ideas presented in the text. Please choose only ONE answer for each question.

1. According to the article, a typical Latin-American family consists of the following members:
   a) a mother and a father without kids.
   b) a mother, a father, and several sons and daughters.
   c) a mother, a father, children, grandparents, aunts and uncles.

2. The typical Latin-American family offers an opportunity to…
   a) have all members of the family eat together at least once every day.
   b) live three different generations under the same roof.
   c) share their expenses.

3. The extended family is socially important because…
   a) it offers relief and moral support to its members.
   b) it helps to reduce poverty in society.
   c) it teaches children how to behave within society.

4. According to the article, which members of the family typically tell the other family members what to do in difficult times?
   a) The grandparents.
   b) The grandfather, the father or the husband.
   c) The grandmother, the mother or the wife.

5. The elderly members of the family are…
a) respected, yet this is rapidly changing.
b) considered an important part of the family and help to raise their grandchildren.
c) not respected anymore.

6. In the Latin-American family children…
a) are the center of the family.
b) are not the center of the family.
c) are as important as the grandparents.

7. If the parents get an invitation to a get together or a party, 
a) the father attends the event and the mother stays home with the children.
b) they usually never go to these events because they prefer to spend time with their children.
c) they usually take their children with them to the event.

8. Another important aspect of the extended family is…
a) to recommend its members when they need to find a job.
b) to help its members start a business.
c) to give the members money when they are in need.

9. According to the article, one difference in politics and business between Latin America and the United States is that …
a) businesses are smaller and politics is more corrupted.
b) the norms of the personalismo are more important in Latin America.
c) there are not big differences because of the existing trade agreements.

10. According to the article, personalismo is…
a) worthless since the Latin-American society has changed.
b) important in few Latin-American countries but not in all of them.
c) a common way to get a job.

Text 2
Comprehension Questions
Please answer the following questions according to the reading. Choose the answer that best reflects the ideas presented in the text. Please choose only ONE answer for each question.

1. The urban current Latin-American family…
   a) is the same traditional family since the values and ideals have not evolved.
   b) is in transition because of a conflict between old and new values.
   c) has evolved completely and does not share anything with the traditional family.

2. Patriarcado is an institution that…
   a) represents traditional family.
   b) should be extinct.
   c) women fight against.

3. According to the article, the role of women in the patriarcado is…
   a) to home school the children.
   b) subordinate to the man’s role, but important since housekeeping and child care are essential.
   c) completely subordinate to a man’s role and unimportant.

4. Under el patriarcado, man’s main responsibility is …
   a) to be in charge of the education of the children.
   b) to be a good provider.
   c) to help with the household chores.

5. Compadrazgo is an institution that is formed by…
   a) members of the extended family.
   b) the oldest members of the extended family.
   c) friends that become as close as any member of the family.

6. The godfather and the godmother are expected to…
   a) give presents to the children and offer support.
b) look after the children in cases when the parents cannot take care of them.
c) give a certain amount of money periodically to support children.

7. Which of the following situations exemplifies a typical instance of a man who leaves his family in Latin America?
   a) A man who leaves his family after filing for divorce.
   b) A man who leaves his family after having an affair with another woman.
   c) A man who leaves his family because of economic issues without being married.

8. The concept of *madresolterismo* is…
   a) uniformly common throughout Latin America and more common in the lower classes.
   b) more common in cities in which men leave their family to work as immigrants.
   c) more common in the upper classes because women want to fulfill their professional careers.

9. What does the article say about the situation of young people in Latin America?
   a) That more than 50% of the population is young.
   b) That a good percentage of the population is young and poor.
   c) That young people live with their parents.

10. According to the article, how does a portion of the younger poor population live in cities?
    a) On the streets.
    b) Sharing a room with many friends.
    c) With their parents until they get married.

---

*Comprehension questionnaires for the count vs. mass experimental condition*

Text 3

Comprehension Questions
1. According to the article, an increase in the divorce rate is the result of?
   a) Changes in values that create conflict between the spouses.
   b) The influence of the mass media on the Latino American family.
   c) The fact that more women are working today than in the past.

2. According to the article, professional middle class in Latin America?
   a) Have more children because both parents work.
   b) Value childhood more.
   c) Marry older.

3. According to the article, the social acceptance of those couples living together without being married?
   a) Has increased substantially in Latino America.
   b) Keeps steady in the last twenty years.
   c) Has only increased in the middle class.

4. Due to an increase in the divorce rate and because more people live together without being married,…
   a) the family is not the only place where children learn values.
   b) Women work outside the home.
   c) fewer families are having children.

5. What is the effect in the traditional family when more women work outside the home?
   a) The mother and the father share the responsibility of raising a family.
   b) The father abandons the family.
   c) Families have more children since they have more income.

6. The dynamics children learn inside the family…
   a) reflect the division of labor inside the family.
b) influence how they behave outside the home.
c) influence how well they behave at school.

7. Boys and girls become conscious about their role in the family…
a) as they become adolescents.
b) early in the childhood.
c) when they start a relationship.

8. It used to be more common for girls inside the Latino American family…
a) to take care of their children.
b) to help with household chores.
c) to respect their brother(s).

9. It used to be more common for boys inside the Latino American family…
a) to be responsible for the protection of their sisters.
b) to find a job in order to bring home some money.
c) not to even finish elementary school.

10. Despite that some traditional roles have been kept,
a) boys and girls learn some gender equality.
b) girls are told not to marry.
c) boys are expecting a submissive wife.

Text 4

Comprehension Questions
Please answer the following questions according to the reading. Choose the answer that best reflects the ideas presented in the text. Please choose only ONE answer for each question.

1. It is a fact that the urban Latin American family…
a) lives closer to the other members of the family.
b) is having more children.
c) is becoming more nuclear.
2. What do the female members of the family do in order to avoid the weakening of family ties?
   a) They talk to each other, and get together for coffee.
   b) They get together and organize family parties.
   c) They get together every other weekend.
3. What can middle class urban families do that working families cannot do in order to take care of their children?
   a) They pay can either pay for child care or leave them with an employee.
   b) They take care of them since the mother does not work.
   c) They can host get to gathers at home.
4. Child care centers in cities are intended to serve…
   a) working families.
   b) middle class families.
   c) both middle and working class families.
5. Why do some elderly people live without their families?
   a) Because of the urban lifestyle and small apartments.
   b) Because both spouses work and they can’t take care of the elderly person.
   c) Because couples are not having kids and they don’t need the grandmother to take care of the children.
6. The urban Latin American family…
   a) is not marrying any more.
   b) revolves around the children more than it did before.
   c) is divorcing more frequently.
7. There is more pressure on young people to fulfill their parents’ expectations because…
   a) they demand more from their parents in return.
   b) parents are expecting their children to support them economically when they are old.
c) parents want their children to fulfill their parents expectations and have good grades.

8. The article says that there is a contradiction between parents’ old and the new values. Which are the new values that the article mentions?
   a) Independence and individualism.
   b) Independence and competence.
   c) responsibility and individualism.

9. Which are the “old practices” that parents find difficult to abandon?
   a) Keeping an eye on their young sons and daughters’ religious lives.
   b) Telling their children what to study.
   c) Keeping an eye on their young sons and daughters’ social and sexual lives.

10. According to the article, young people in Latin America…
    a) are in a difficult period because their values are changing.
    b) are in disadvantaged because they have fewer opportunities than their parents.
    c) face the same problems as their parents’ generation.
APPENDIX E

Assessment tasks: Fill-in-the blank (production task) and Multiple-choice (recognition task)

The experimental assessment tasks for the animate vs. inanimate experimental condition

Texts 1-2: Fill-in-the-blank

Please fill in the blanks with the word that preceded a similar word in the texts that you have just read.

1. __________________ padre.
2. __________________ solidaridad.
3. __________________ canción.
4. __________________ paz.
5. __________________ madre.
6. __________________ respeto.
7. __________________ odio.
8. __________________ libro.
9. __________________ crianza.
10. __________________ biblioteca.
11. __________________ familia.
12. __________________ niña.
13. __________________ novio.
14. __________________ doctora.
15. __________________ gato.
16. __________________ esposo.
17. __________________ nieto.
18. __________________ pantalón.
19. __________________ factor.

282
20. __________________ vínculo.
21. __________________ prima.
22. __________________ miembro.
23. __________________ amiga.
24. __________________ abuela.

Texts 1-2: Multiple-choice
Please mark with an (X) those sentences that are correct.

( ) 1. __________ río.
   a) ri río  b) ji río  c) ji río  d) ju río

( ) 2. __________ fiesta.
   a) ri fiesta  b) ru fiesta  c) ji fiesta  d) ju fiesta

( ) 3. __________ hija.
   a) ri hija  b) ru hija  c) ji hija  d) ju hija

( ) 4. __________ niño.
   a) ri niño  b) ru niño  c) ji niño  d) ju niño

( ) 5. __________ amor.
   a) ri amor  b) ru amor  c) ji amor  d) ju amor

( ) 6. __________ ciudadana.
   a) ri ciudadana  b) ru ciudadana  c) ji ciudadana  d) ju ciudadana

( ) 7. __________ tiempo.
   a) ri tiempo  b) ru tiempo  c) ji tiempo  d) ju tiempo

( ) 8. __________ padrino.
   a) ri padrino  b) ru padrino  c) ji padrino  d) ju padrino

( ) 9. __________ ciudad.
   a) ri ciudad  b) ru ciudad  c) ji ciudad  d) ju ciudad

( ) 10. __________ escuela.
   a) ri escuela  b) ru escuela  c) ji escuela  d) ju escuela
11. ________ calle.
a) ri calle  b) ru calle  c) ji calle  d) ju calle

12. ________ oxígeno.
a) ri oxígeno  b) ru oxígeno  c) ji oxígeno  d) ju oxígeno

13. ________ dinero.
a) ri dinero  b) ru dinero  c) ji dinero  d) ju dinero

14. ________ educación.
a) ri educación  b) ru educación  c) ji educación  d) ju educación

15. ________ alumno.
a) ri alumno  b) ru alumno  c) ji alumno  d) ju alumno

16. ________ hombre.
a) ri hombre  b) ru hombre  c) ji hombre  d) ju hombre

17. ________ casa.
a) ri casa  b) ru casa  c) ji casa  d) ju casa

18. ________ profesor.
a) ri profesor  b) ru profesor  c) ji profesor  d) ju profesor

19. ________ sobrina.
a) ri sobrina  b) ru sobrina  c) ji sobrina  d) ju sobrina

20. ________ mujer.
a) ri mujer  b) ru mujer  c) ji mujer  d) ju mujer

21. ________ odio.
a) ri odio  b) ru odio  c) ji odio  d) ju odio

22. ________ persona.
a) ri persona  b) ru persona  c) ji persona  d) ju persona

23. ________ vecina.
a) ri vecina  b) ru vecina  c) ji vecina  d) ju vecina

24. ________ enfermero.
a) ri enfermero  b) ru enfermero  c) ji enfermero  d) ju enfermero
The experimental assessment tasks for the count vs. mass experimental condition

Texts 3-4: Fill-in-the-blank

Please mark with an (X) those sentences that are, in your opinion and based on the texts you have just read, correct.

1. ________________ padre.
2. ________________ solidaridad.
3. ________________ bondad.
4. ________________ máquina.
5. ________________ madre.
6. ________________ respeto.
7. ________________ odio.
8. ________________ jugo.
9. ________________ crianza.
10. ________________ leche.
11. ________________ familia.
12. ________________ importancia.
13. ________________ novio.
14. ________________ riqueza.
15. ________________ maestro.
16. ________________ esposo.
17. ________________ escuela.
18. ________________ costo.
19. ________________ presente.
20. ________________ carro.
21. ________________ comportamiento.
22. ________________ cuidado.
23. ________________ amiga.
24. ________________ mujer.
Texts 3-4: Multiple-choice

Please mark with an (X) those sentences that are, in your opinion and based on the texts you have just read, correct.

( ) 1. __________ río.
   a) fi río    b) fu río    c) pi río    d) pu río

( ) 2. __________ fiesta.
   a) fi fiesta  b) fu fiesta  c) pi fiesta  d) pu fiesta

( ) 3. __________ hija.
   a) fi hija   b) fu hija   c) pi hija   d) pu hija

( ) 4. __________ hijo.
   a) fi hijo   b) fu hijo   c) pi hijo   d) pu hijo

( ) 5. __________ amor.
   a) fi amor   b) fu amor   c) pi amor   d) pu amor

( ) 6. __________ vecina.
   a) fi vecina b) fu vecina c) pi vecina d) pu vecina

( ) 7. __________ futuro.
   a) fi futuro b) fu futuro c) pi futuro d) pu futuro

( ) 8. __________ vínculo.
   a) fi vínculo b) fu vínculo c) pi vínculo d) pu vínculo

( ) 9. __________ café.
   a) fi café   b) fu café   c) pi café   d) pu café

( ) 10. __________ armonía.
   a) fi armonía b) fu armonía c) pi armonía d) pu armonía

( ) 11. __________ leche.
   a) fi leche   b) fu leche   c) pi leche   d) pu leche

( ) 12. __________ oxígeno.
   a) fi oxígeno b) fu oxígeno c) pi oxígeno d) pu oxígeno

( ) 13. __________ dinero.
a) fi dinero  b) fu dinero  c) pi dinero  d) pu dinero

( ) 14. ___________ educación.
a) fi educación  b) fu educación  c) pi educación  d) pu educación

( ) 15. ___________ vaso.
a) fi vaso  b) fu vaso  c) pi vaso  d) pu vaso

( ) 16. ___________ hombre.
a) fi hombre  b) fu hombre  c) pi hombre  d) pu hombre

( ) 17. ___________ época.
a) fi época  b) fu época  c) pi época  d) pu época

( ) 18. ___________ falda.
a) fi falda  b) fu falda  c) pi falda  d) pu falda

( ) 19. ___________ blanco.
a) fi blanco  b) fu blanco  c) pi blanco  d) pu blanco

( ) 20. ___________ zapato.
a) fi zapato  b) fu zapato  c) pi zapato  d) pu zapato

( ) 21. ___________ infancia.
a) fi infancia  b) fu infancia  c) pi infancia  d) pu infancia

( ) 22. ___________ clase.
a) fi clase  b) fu clase  c) pi clase  d) pu clase

( ) 23. ___________ basura.
a) fi basura  b) fu basura  c) pi basura  d) pu basura

( ) 24. ___________ protección.
a) fi protección  b) fu protección  c) pi protección  d) pu protección
## Gender assignment test

<table>
<thead>
<tr>
<th></th>
<th>_______</th>
<th></th>
<th>_______</th>
<th></th>
<th>_______</th>
<th></th>
<th>_______</th>
<th></th>
<th>_______</th>
<th></th>
<th>_______</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Hijo</td>
<td></td>
<td>31</td>
<td></td>
<td>comprensión</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>mesa</td>
<td></td>
<td>32</td>
<td></td>
<td>rubí</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>felicidad</td>
<td></td>
<td>33</td>
<td></td>
<td>valor</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>costumbre</td>
<td></td>
<td>34</td>
<td></td>
<td>cometa</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>vejez</td>
<td></td>
<td>35</td>
<td></td>
<td>problema</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>clima</td>
<td></td>
<td>36</td>
<td></td>
<td>enigma</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>mano</td>
<td></td>
<td>37</td>
<td></td>
<td>cama</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>planeta</td>
<td></td>
<td>38</td>
<td></td>
<td>esperanza</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>toro</td>
<td></td>
<td>39</td>
<td></td>
<td>virtud</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>yegua</td>
<td></td>
<td>40</td>
<td></td>
<td>rama</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>día</td>
<td></td>
<td>41</td>
<td></td>
<td>refrán</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>letra</td>
<td></td>
<td>42</td>
<td></td>
<td>calor</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>pared</td>
<td></td>
<td>43</td>
<td></td>
<td>enjambre</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>amor</td>
<td></td>
<td>44</td>
<td></td>
<td>equipaje</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>libra</td>
<td></td>
<td>45</td>
<td></td>
<td>reloj</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>alcalde</td>
<td></td>
<td>46</td>
<td></td>
<td>cámara</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>rey</td>
<td></td>
<td>47</td>
<td></td>
<td>foto</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>profesora</td>
<td></td>
<td>48</td>
<td></td>
<td>señal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>ángel</td>
<td></td>
<td>49</td>
<td></td>
<td>paleta</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>color</td>
<td></td>
<td>50</td>
<td></td>
<td>animal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>gente</td>
<td></td>
<td>51</td>
<td></td>
<td>lumbre</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>amistad</td>
<td></td>
<td>52</td>
<td></td>
<td>silla</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>libro</td>
<td></td>
<td>53</td>
<td></td>
<td>refrigerador</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>especie</td>
<td></td>
<td>54</td>
<td></td>
<td>sed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>felicidad</td>
<td></td>
<td>55</td>
<td></td>
<td>canción</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>___</td>
<td>diamante</td>
<td>56</td>
<td>___</td>
<td>papel</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>___</td>
<td>cara</td>
<td>57</td>
<td>___</td>
<td>paz</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>___</td>
<td>lugar</td>
<td>58</td>
<td>___</td>
<td>actitud</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>___</td>
<td>depresión</td>
<td>59</td>
<td>___</td>
<td>rapidez</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>___</td>
<td>persona</td>
<td>60</td>
<td>___</td>
<td>serie</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX G

Background Questionnaire

Name: _____________________________________________
Date:_____________________
Current Spanish Course: _____________________   Professor: ____________________________

All information will remain absolutely confidential. It only will be seen and used by the researcher according to the purpose of the research. Please put it in the envelope when you finish.

1) Biographical and background information

a) Gender:    Male      Female

b) Nationality: _____________________

c) Native language: ___________________

d) What year in school are you?

Freshman   Sophomore   Junior   Senior   Graduate (Ma/ Ms) Graduate (Phd)

e) What is your current major (s) / minor(s)

____________________________________________

f) What other languages have you studied and form how many semesters?

Language: _______________ Number of semesters: ___________
Language: _______________ Number of semesters: ___________

g) Why are you taking Spanish?

1) Career   2) Fun    3) Requirement    4) Study abroad

5) Communicate with Spanish speaking people   6) Other:

____________________________________________

h) Amount of time studied Spanish.

____________________________________________

i) Have you been to the country of the target language? Yes/ No

If yes, amount of the time spent in the country.
BIBLIOGRAPHY


Lardiere, D. (2009b). Further thoughts on parameters and features in second language acquisition: A reply to peer comments on Lardiere's 'some thoughts on the contrastive analysis of features in


Tsimpi, I. M. & M. Mastropavlou. (2008). ‘Feature interpretability in L2 acquisition and SLI: Greek clitics and determiners.’ In Liceras, J.


