# EL(LA) MAPPING: AN INTEGRATED ACCOUNT OF LEARNING CONTEXT, FEEDBACK AND AGREEMENT MORPHOLOGY IN THE PROCESSING OF $O_{cl}$VS SENTENCES IN ADVANCED L2 SPANISH

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

By

Silvia Susana Marijuan, M.S.

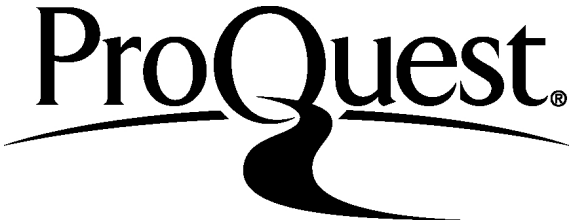
Washington, D.C.
June 18, 2015

ProQuest Number: 3717964

All rights reserved

INFORMATION TO ALL USERS
The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript
and there are missing pages, these will be noted. Also, if material had to be removed,
a note will indicate the deletion.



ProQuest 3717964

Published by ProQuest LLC (2015). Copyright of the Dissertation is held by the Author.

All rights reserved.
This work is protected against unauthorized copying under Title 17, United States Code
Microform Edition © ProQuest LLC.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346

Copyright 2015 by Silvia Susana Marijuan
All rights reserved

EL(LA) MAPPING: AN INTEGRATED ACCOUNT OF LEARNING CONTEXT, FEEDBACK AND AGREEMENT MORPHOLOGY IN THE PROCESSING OF $O_{cl}$VS SENTENCES IN ADVANCED L2 SPANISH

Silvia Susana Marijuan, M.S.

Thesis Advisor: Cristina Sanz, Ph.D.

## ABSTRACT

Previous literature has shown that beginning and intermediate English-speaking learners persistently misinterpret O-cliticVS sentences in Spanish, preferring word order over morphology when assigning semantic functions to the NPs of non-canonical sentences. Following the Competition Model (Bates & MacWhinney, 1982, 1989; MacWhinney, 2012), this dissertation investigated from an on-line and off-line account whether English-speaking advanced learners of Spanish are also prone to such misinterpretations, and whether their reconfiguration of L1 processing strategies benefits from: (a) mismatches in number agreement morphology, (b) immersion experience, and (c) computer-delivered feedback.

Three self-paced readings were conducted. In Experiment 1 (*N*=38), matching/mismatching number agreement between clitic and verb was manipulated to measure the usefulness of contrastive agreement to overcome word order bias. Experiment 2 (*N*=20) investigated whether a 5-week Study Abroad program promoted learners' $O_{cl}$VS sentence development. Experiment 3 (*N*=90) investigated potential interactions between agreement conditions and the provision or absence of computer-delivered feedback.

Results showed that advanced learners continued to rely on word order when interpreting $O_{cl}$VS sentences, but different agreement conditions were not processed alike. In Experiment 1, accuracy significantly improved in the mismatching condition

in which the morphological cue was on the verb. Participants also tended to present longer reading times in the verb of $Cl_{sg}V_{pl}S_{pl}$ structures. In Experiment 2, learners significantly improved at Week 5 and paid more attention to the verb and post-verbal subject, which are highly informative. In Experiment 3, the [+ Feedback] group outperformed the [- Feedback] group, which only received practice decoding manipulated input. Also, the [+ Feedback] group exhibited a speedup across experimental trials in the baseline condition ($Cl_{sg}V_{sg}S_{sg}$) whereas the [- Feedback] group exhibited a significant slowdown in the mismatching $Cl_{sg}V_{pl}S_{pl}$ condition.

This dissertation has implications for research on intra-subject L2 processing variation and emerging bilingualism. It also contributes to the debate on whether short-term studies overseas influence L2 grammar skills. Finally, the extent to which exposure to manipulated input alone helps learners reconfigure their L2 processing strategies is discussed along with feedback's potential role in enhancing this process.

ACKNOWLEDGEMENTS

Writing a dissertation is a unique learning experience, like riding a roller coaster, with ups and downs, ups and downs, until in the end, with your hair a little bit undone and your heart still beating hard, you look at yourself in the mirror and say, probably whispering, that you have made it and are ready to go on the next ride. But on every ride there is a co-pilot who helps you maneuver. I have to thank so many friends, family, and professors for having been my co-pilots, for having helped me navigate my dreams and my incipient accomplishments.

To my family, for always being there, through the good and the bad. To my husband, Scott Ferree, for being my iron man, my bridge to this culture, my guide, and my main supporter. You have helped me avoid going in circles, driven by perfectionism. Scott, I could write hundreds of pages that would not do justice to thank you for all you have done for me all these years.

To my parents, who have taught me to persevere, to work as hard as I can, and then work even harder, always keeping my goals in mind. They have also taught me the important lesson of "smiling at bad weather." I would not be half the optimist and fighter I am today if it were not for them.

To my mentor, Cristina Sanz, for all we have shared throughout these five years. When I think of Cristina, I see smiles, the sea, and a stimulating conversation that keeps me thinking afterwards. Cristina, like everyone, I admire you: for your personal strength, your creativity, and your unique sense of humor. I wish to be like you one day. You helped me find my voice. I have learned it from the Sanz-style. Thank you for the time you have dedicated to me, for helping me mature intellectually, for opening so many doors for me, and windows as well, and for

guiding me with the right questions. No matter how much time passes, I owe you, Cristina, for becoming the person I have become today.

To Lourdes Ortega for having taught me to look at the big picture and to maintain a critical mindset towards biases. Lourdes, you have shown me the importance of organizing my thoughts, and of synthesizing and prioritizing my projects. You have also helped me to rethink and revaluate my ideas about the process of writing, which, in turn, has helped me reconnect with my own writing. You have helped all my selves emerge and engage in a textual mesh! Thank you, Lourdes, for all you have done for me. There is a before and after knowing you!

To Nuria Sagarra, for all your recommendations and guidance in the preparation of the material for this dissertation. Nuria, you have always encouraged me to improve and defend my ideas clearly. Thank you for sharing your expertise with me, and your points of view on our common passion: sentence processing.

I owe my deepest gratitude to other scholars who have also shaped my thinking. I especially thank John Norris for having introduced me to TBLT and assessment, and for having encouraged me to think (even more) critically. John, you have taught me about the importance of purpose and true intentions, and the consequences of our decisions when we make plans. Questioning arbitrariness is another lesson I have learned from you. I would also like to recognize your influence on my teaching style, and my approach to tasks.

To Ron Leow, for accompanying me in my first steps into SLA and for guiding me in language program coordination. Ron, you inspired me with a passion for knowing more about processing and processes, and taught me that no job is hard when it is done with pleasure.

I would also like to thank Alfonso Morales-Front for have worked with me when I started my first year at Georgetown, and for always be willing to give me a good piece of advice.

I can never forget Héctor Campos for all the conversations we had about linguistics. Héctor, thank you for being always happy to exchange ideas.

To my friend, Emily Francomano, for having always believed in me and for having given me all your support. Thank you, Emily, for sharing your world with me. I will always remember the good moments we spent together discussing *Grisel and Mirabella.*

To Gwen Kirkpatrick, for sharing all your experience with me, for having interesting conversations with me, and for being authentically interested in graduate students' lives. Gwen, it has been a pleasure to work with you these last two years.

Thanks to Michael Ferreira for your great sense of humor and all your help (in spite of your preference for a different national team in soccer).

To all the amazing "hoyas" (joyas!) whom I have had the honor of meeting in these five years, your friendship will accompany me as long as I am able to remember. To Germán Zárate-Sández, my favorite Argentinian, for being at my side during one of the most difficult moments of my life. Thank you, Germán for being the perfect friend! To Julio Torres for being my first mentor, for having taught me all I needed to know to survive, and for all the good moments and the laughs we have shared. To the three stars: Natalia Curto, Sandra Martin and Janire Zalbidea, for all your support during this last year. I have so much enjoyed being around you, and will miss you so much! To Yoel Castillo Botello, for your beautiful smiles and all your help. Yoel you are a diamond! To Pablo for being there every time I needed to talk to a friend. To Charlie Nagle for all the good times in Barcelona, drinking *cafè tallat*. I

would like to mention all the people who have given me their support in one way or another: Ellen Johnson, for being such a good friend and answering all the questions that I always ask her, to Sergio Adrada Rafael, Gabriel Villaroel, Jessie Cox, Ross Karlan, Sarah Grey, Ariel Zach, Colleen Moorman, Lu Maimone, Angela Donate Velasco, Marisa Filgueras, Nick Pandža, Lauren Park, Anastasia Mozgalina, Mariko Uno, Brandon Tullock, and Aitzi Delgado.

Outside Georgetown, I would like to thank my good friend from the University of Maryland, Sol Lago. Thank you Sol for having taught me so much about Self-Paced Reading, eye-tracking, and experimental design in general. True to your style, you have also been my harshest critique, but I have to recognize that you have shown me the way to improve my argumentation "Maryland-style"!

To Ewan Dunbar for all your help in deciphering the conundrums of coding. Ewan, your mathematician's mind takes anyone into a parallel world!

I would like to thank all the students who have helped me in my studies, and especially the students from the Barcelona program.

Finally, I would like to thank Sue and Rick Ferree, my parents-in-law, for giving me all their support since I came to the United States. It is so comforting to know that I can always count on you! Gracias.

TABLE OF CONTENTS

CHAPTER ONE: Introduction . . . . . 1

Statement of Problem . . . . . 1

Competition Model: Basic notions and predictions . . . . . 4

Cue strength readjustments and processing costs in L2 development 6

Gaps in previous literature and the significance of the present studies 10

Potential contributions to SLA theory and pedagogy . . . . . 20

Organization of the present studies . . . . . 22

Definitions of terms and procedures . . . . . 23
Self paced reading procedure . . . . . 23
Sentence processing . . . . . 25
L2 proficiency . . . . . 25

CHAPTER TWO: Contrastive Agreement . . . . . 27
Agreement, word order, and intra-subject variation in L2
Spanish sentence processing . . . . . 27

Introduction . . . . . 27

Background . . . . . 30
The role of morphology in previous studies . . . . . 30

On-line processing measurements in the study of $O_{cl}$VS sentences 37

Theories, concepts, and approaches to agreement . . . . . 38
What does agreement mean? . . . . . 40
Agreement and word order with advanced learners . . . . . 41
Subject-verb agreement and word order in L1 research . . . . . 42
Subject-verb agreement in L2 Spanish . . . . . 45
Summary . . . . . 46

Spanish direct object as a special type of clitic . . . . . 47

Experiment 1 . . . . . 49
Rationale for the present study . . . . . 49
Goals of the present study . . . . . 50

Research questions (RQs) and Predictions (Ps) . . . . . 51
Methods . . . . . 52
Participants . . . . . 52
Design . . . . . 53
$O_{cl}$VS sentences . . . . . 53

SV sentences . . . . . 54
Materials . . . . . 57
Procedures . . . . . 59

Analyses . . . . . 61
Accuracy data . . . . . 61
Latency data . . . . . 61
Statistical procedures . . . . . 62

Results . . . . . 63
Accuracy in $O_{cl}$VS sentences . . . . . 63
Reaction Times in $O_{cl}$VS sentences . . . . . 66
Reaction Times in SV sentences . . . . . 68

Discussion . . . . . 70

Conclusions . . . . . 73

CHAPTER THREE: Learning Context . . . . . 75
Evidence of morphosyntactic development in immersion contexts: An online and offline account of $O_{cl}$VS sentence processing in L2 Spanish . . . . . 75

Introduction . . . . . 75

Overview of previous SA studies . . . . . 77

Immersion experience and morphosyntactic development in L2 Spanish 79

Summary . . . . . 82

Online and offline measurements in SA contexts . . . . . 84

Research questions (RQs) . . . . . 86
In relation to $O_{cl}$VS sentences . . . . . 86
In relation to SV sentences . . . . . 87

Predictions (Ps) . . . . . 87
In relation to $O_{cl}$VS sentences . . . . . 87
In relation to SV sentences . . . . . 88

Methods . . . . . 88
Description of participants and inclusion criteria . . . . . 88
Characteristic of the SA program . . . . . 90
Research design . . . . . 92
$O_{cl}$VS sentences . . . . . 95
SV sentences . . . . . 96
Fillers . . . . . 97
Gramaticality Judgment Tasks (GJTs) . . . . . 98
Materials . . . . . 98

Procedures . . . . . 100

Analyses . . . . . 103
Accuracy data . . . . . 103
Latency data . . . . . 104
Statistical procedures . . . . . 104

Results . . . . . 105
$O_{cl}$VS sentences . . . . . 105
Accuracy in $O_{cl}$VS sentences . . . . . 105
Latency in $O_{cl}$VS sentences . . . . . 113
Processing costs in $O_{cl}$VS sentences overall . . . . . 117
Processing costs in $O_{cl}$VS sentences in the post-test 118
SV sentences . . . . . 118
Latency in SV sentences . . . . . 118
Grammaticality Judgment Tasks . . . . . 121

Discussion . . . . . 126

Conclusions . . . . . 130

CHAPTER FOUR: Feedback . . . . . 132
Assessing instructional factors in advanced learners' $O_{cl}$VS sentence processing: Does feedback make a difference? . . . . . 132

Introduction . . . . . 132

Background . . . . . 138
The role of feedback from a Competition Model (CM) perspective . . . . . 138
The effect of computer-delivered feedback on morphosyntactic development . . . . . 142

Computer-delivered feedback characteristics and operationalizations 144

Previous studies on $O_{cl}$VS sentence processing in L2 Spanish using computer-delivered feedback . . . . . 146

The Present Study . . . . . 150
Research questions (RQs) and Predictions (Ps) . . . . . 150

Methods . . . . . 152
Description of participants . . . . . 152
Research design . . . . . 153
Materials . . . . . 155
$O_{cl}$VS sentences . . . . . 155
Fillers . . . . . 157
Procedures . . . . . 158

Analyses . . . . . 162

Accuracy data . . . . . 162
Latency data . . . . . 163
Statistical procedures . . . . . 163

Results . . . . . 164
Accuracy in $O_{cl}$VS sentences . . . . . 164
No Feedback group . . . . . 168
Feedback group . . . . . 169
Aggregated accuracy data from the No Feedback and Feedback groups . . . . . 170
Latency in $O_{cl}$VS sentences . . . . . 173
RT data analysis for the clitic region . . . . . 180
RT data analysis for the verb region . . . . . 181
RT data analysis for the spillover regions: determiner and noun . . . . . 182

Discussion . . . . . 183

Conclusions . . . . . 192

CHAPTER 5: Conclusions . .. . . . . 195

Similarities and differences in the experiments conducted . . . . . 195

Discussion . . . . . 200
Accuracy outcomes in $O_{cl}$VS sentences . . . . . 200
$O_{cl}$VS sentence latency . . . . . 205

Methodological, theoretical and pedagogical implications . . . . . 209

Limitations and future research . . . . . 210

Overall conclusions . . . . . 212

APPENDIX A: Examples of grammar section of Spanish Placement Test 216

APPENDIX B: Bio questionnaire . . . . . 217

APPENDIX C: Exit questionnaire . . . . . 220

APPENDIX D: Elicited Imitation Task stimuli and scoring protocol 222

APPENDIX E: Vocabulary test in Quizlet (example) . . . . . 225

APPENDIX F: Sample of words in Quizlet . . . . . 226

APPENDIX G: Vocabulary answer sheet . . . . . 227

Bibliography . . . . . 228

LIST OF FIGURES

Figure 1: Word-by-word self-paced reading procedure . . . . . 24

Figure 2: Example of word-by-word self-paced reading procedure 24

Figure 3: Description of experimental session . . . . . 59

Figure 4: Accuracy (Trial Order × Condition) . . . . . 65

Figure 5: Mean residual Reaction Times (ms) in each condition ($O_{cl}$VS sentences) . . . . . 67

Figure 6: Mean residual Reaction Times (ms) for SV Sentences 68

Figure 7: Description of the three experimental sessions (pre and post-test) . . . . . 101

Figure 8: $O_{cl}$VS accuracy over the course of the experiment in the pre-test and post-test . . . . . 110

Figure 9: Mean residual RTs overall for all conditions in the pre-test and post-test . . . . . 113

Figure 10: RT data for SV sentences . . . . . 119

Figure 11: Accuracy for GJTs in the pre-test and post-test . . . . . 122

Figure 12: A mechanism for learning-on-error to determine L1 and L2 cue-strengths . . . . . 139

Figure 13: Description of two experimental sessions . . . . . 159

Figure 14: $O_{cl}$VS accuracy over the course of the experiment in the No Feedback and Feedback groups . . . . . 166

Figure 15: Mean Residual RTs overall for all conditions in the No Feedback and Feedback groups . . . . . 174

LIST OF TABLES

Table 1: Studies on word order ($O_{cl}$VS structures) in L2 Spanish 33

Table 2: Properties of third person direct object clitics in Spanish 48

Table 3: Sample set of experimental items . . . . . 56

Table 4: Mean accuracy $O_{cl}$VS sentences × Condition (overall) 63

Table 5: Logistic regression for $O_{cl}$VS sentences (Condition × Trial Order) . . . . . 64

Table 6: Effect sizes for $O_{cl}$VS sentences (accuracy) . . . . . 66

Table 7: Residual Reaction Times (ms) × Region × Condition ($O_{cl}$VS sentences) . . . . . 67

Table 8: Effect sizes for ungrammatical sentences in Foote (2011) and the present study. . . . . 69

Table 9: Inclusion criteria for participation in SA experiment . . . . . 90

Table 10: Sample set of experimental items in the SA experiment 92

Table 11: Overall accuracy for $O_{cl}$VS sentences in the pre-test and post-test . . . . . 106

Table 12: Logistic regression for $O_{cl}$VS sentences . . . . . 108

Table 13: Logistic regression for $O_{cl}$VS sentences in the pre-test 110

Table 14: Logistic regression for $O_{cl}$VS sentences in the post-test 111

Table 15: Linear models for the clitic region, the verb region, and spillover region . . . . . 115

Table 16: Mean accuracy for Grammaticality Judgment Tasks (GJTs) 122

Table 17: Logistic regressions for pre-test and post-test of GJTs . . . 123

Table 18: Mean accuracy on GJTs for left-handed participants . . . . 125

Table 19: Sample set of $O_{cl}$VS sentences (Latin Square design) . . . . 154

Table 20: Overall accuracy for $O_{cl}$VS sentences with feedback and no-feedback . . . . . 165

Table 21: Logistic regression for $O_{cl}$VS sentences in the No Feedback group . . . . . 167

Table 22: Logistic regression for $O_{cl}$VS sentences in the Feedback group . . . . . 167

Table 23: Logistic regression for $O_{cl}$VS sentences in the No Feedback and Feedback groups (aggregated accuracy data) . . . . . 170

Table 24: Linear models for the clitic region, the verb region, and spillover region . . . . . 176

Table 25: Linear models for the Residual RT data in the No Feedback and Feedback Groups (aggregated residual RT data) . . . . . 178

Table 26: Main differences in research design in the three experiments conducted . . . . . 196

## CHAPTER ONE: Introduction

### Statement of the Problem

The study of Object clitic-Verb-Subject ($O_{cl}$VS) sentences in Spanish as a Second Language (L2) continues to be productive after over three decades of research. Disentangling how non-canonical word order sentences are processed in the L2 taps into critical areas in the field of Second Language Acquisition (SLA) such as L1 transfer, L2 sentence processing, and L2 grammar development. In addition, research on $O_{cl}$VS sentences (e.g., *La mira el estudiante*, her-clitic.fem.sing. look at-3rd p. sing.- the student, "The student looks at her") is a fruitful arena for exploring how linguistic, pedagogical and contextual factors impact non-canonical word order sentence processing and development in the L2.

Empirical evidence to date suggests that $O_{cl}$VS structures in Spanish are persistently misinterpreted by English-speaking adult learners, regardless of level of proficiency (e.g., Lee, 2000; Lee & Malovrh, 2009; LoCoco, 1987; Montrul, 2010; Malovrh & Lee, 2010; Sanz & Morgan-Short, 2004; VanPatten, 1984; VanPatten & Houston, 1998, VanPatten & Borst, 2012, for beginner and intermediate leaners; and Lee & Malovrh, 2009; Malovrh & Lee, 2010 for advanced learners). Misinterpretations associated with $O_{cl}$VS sentence processing are the result of English-speaking learners' assigning the role of agent to the preverbal object of the sentence, instead of the role of patient. Such thematic role misassignment interferes with efficient communication, since what learners usually comprehend when confronted with the task of deciding *who* does what to *whom* is the inverse of what the sentence actually conveys (e.g., in the above example, the preverbal object clitic *la*, "her" is interpreted as *ella*,"she").

The question of how the human processor deals with sentences in non-

canonical word order began to be addressed in the neighboring field of cognitive psychology. Bever (1970), whose findings greatly influenced the early work conducted on $O_{cl}VS$ sentences in L2 Spanish (LoCoco, 1982, 1987; VanPatten, 1984), was the first to note that English-speaking children tended to make use of a processing strategy that led them to equate any given *Noun-Verb-Noun* (NVN) sequence with an *actor-action-object* structure, even at the expense of semantic plausibility (Slobin, 1989). According to Bever (1970), this habit results from children's exposure to an overwhelming number of sentences that conform to this specific pattern in English. In other words, the misinterpretations English speakers exhibit when processing non-canonical sentences in L1acquisition, as well as in second language acquisition, originate in the high frequency the SVO pattern presents in English.

The role of frequency in language acquisition and processing pattern productivity has been well documented by scholars who have adopted connectionist approaches to language acquisition (for a full review of connectionist language models, see Elman, 2003, 2005). In their view, structural regularities of language emerge from a learner's lifetime analysis of the distributional characteristic of the language input. Therefore, language representations consist of a statistical ensemble of language experiences (Ellis, 2003, Elman, 2001).

In connectionist models, the concept of *network*, which is neurally inspired (Bates & Elman, 1993; Ellis, 1998, 2003; Elman, 2001; Medler, 1998), is central to understanding the emergence of linguistic regularities, as well as the way linguistic information is processed. During language use, learners gradually make a stronger network of connections between different linguistic features in specific linguistic contexts, until the presence of one linguistic feature activates other feature(s); that is, parallel activation takes place in the learner's mind (Elman, 2003; Li & Zhao, 2012).

Consequently, what a system "knows" is, to a large extent, captured by patterns of connections as well as the weights associated with each connection. The network learns the values for the weights on its own (it self-programs). This means that, through learning, the network can adjust the weights in small incremental steps in such a way that, over time, the network's predictive accuracy improves (Elman, 2003).

As Ellis (2003) argued, connectionists (Christiansen, Chater & Seidenberg, 1999; Levy, Bairaktaris, Bullinaria & Cairns, 1995; McClelland & Rumelhart, 1986; Plunkett, 1998), emergentists (Bates & Elman, 1993, Elman, Bates, Johnson, Karmiloff-Smith, Parisi, & Plunket, 1996; MacWhinney, 1999) and functional linguists (Bates & MacWhinney, 1981; MacWhinney & Bates, 1989) all support a functional-developmental usage-based perspective on language. This dissertation supports the theoretical assumption which Ellis (2003) noted as a commonality in these scholars' work, that language use and language knowledge are inseparable; it therefore ascribes to the eminent "emergentist turn" that the SLA field has been witnessing in the last few years (Ortega, 2009, p. 102). Ascribing to this perspective entails being critical of traditional cognitivist notions (e.g., the dichotomy between competence vs. performance, Potter, 2000) and recognizing that L2 learning is grounded in experience and exposure to input.

Of the connectionist models that have focused on language acquisition, the Competition Model (CM) (e.g., Bates & MacWhinney, 1982, 1987, 1989; MacWhinney, 2012) is the theoretical framework that best fits the research questions of the present dissertation, which investigates morphosyntactic phenomena: form-to-function mappings in non-canonical sentences. A major tenet of the CM is that different languages employ different grammatical cues in a wide and varied way to

assign grammatical functions (Sokolov, 1989, p. 119). In fact, much CM research has focused on cue-role assignment, mainly because the model has the strength of being applicable to many languages, an aspect that "distinguishes it from the majority of other models of language processing which are based largely on work in English" (Kilborn & Ito, 1989). Such cross-linguistic utility affords an exploration of how a second language is acquired. In addition, the CM offers a dynamic view of how learners make predictions based on the way linguistic forms interact within a specific language (i.e., language mappings are predictive and non-deterministic). This allows the model to account for language transfer effects (e.g., MacWhinney, 1992). Most importantly, the CM contemplates strategic change, since, so as to increase its predictive power, the network architecture is thought to dynamically self-organize as a result of external constraints from the environment (e.g., McDonald, 1987, 1989).

The following section provides an account of the basic notions the model is built on, and states the predictions the model makes for the acquisition of non-canonical sentences in L2 Spanish.

**Competition Model: Basic notions and predictions**

The CM predicts a set of strategies involved in L1 and L2 sentence processing based on the following notions: (a) *cues*: the surface structures that are syntactic, morphological, phonological, and lexical in nature; Morett & MacWhinney, 2013) (e.g., word order, case morphology, agreement morphology); (b) *cue availability*: the number of times a cue is available over the times it is needed; (c) *cue validity*: the number of times a cue is correct over the total number of occurrences (i.e., the information value associated with a particular linguistic form within a particular language); and (d) *cue strength*: the strength of specific form-meaning mappings. Cue validity depends on cue availability and cue strength is based on cue validity.

Natural languages can vary either quantitatively, depending on whether a specific cue is present or absent in that language, or qualitatively, depending on whether a specific cue is more or less valid, and, as a consequence, stronger or weaker. Languages do not necessarily share the same type of cues, nor do they exhibit the same cue validity and strength configurations. In fact, such configurations or "hierarchies" can be radically different (Bates, MacWhinney, Caselli, Devescovi, Natale & Venza, 1984; MacDonald, 1989; Kail 1989).

According to the CM, *sentence comprehension* is viewed as a dynamic process of parallel cue activation, cue competition, and cue convergence. *Cue activation* refers to the activation of input cues in the sentence. *Cue competition* refers to different cues which point to competing functional interpretations of the input sentence: when cues are set into competition, the strongest cues in the hierarchy "win." Finally, *cue convergence* refers to the cluster of more than one cue pointing in the same direction (e.g., initial position, preverbal position, and nominative case "converge" to assign the first noun the role of agent of the sentence, MacWhinney, 2002).

When learners undergo the task of determining the agent of a transitive sentence, the cues available in the input sentence actively compete with one another or converge in different combinations, and as a result, a particular interpretation emerges as the best fit for that specific sentence. According to this approach, English-speaking learners' $O_{cl}$VS misinterpretations would be attributed to their overreliance on word order, a hierarchically stronger cue that overshadows morphological cues (e.g., agreement morphology and case morphology) (McDonald, 1987; Reyes & Hernandez, 2006).

English is a non pro-drop language lacking the morphological richness of Spanish: for example, English – unlike Spanish– does not exhibit person markers in

the past, and in the present, only the third person singular is morphologically marked (Sagarra, 2014); due to this fact, SVO is the most frequent word order configuration English exhibits, and word order is the most valid cue for agenthood (i.e., the probability of getting to the right interpretation of a sentence by assigning the role of agent to a sentence-initial element is overwhelmingly high in English). By contrast, in Spanish – a pro-drop language – the cue validity and strength of word order takes a second place position in its cue hierarchy (Kail & Charvillat, 1988), since word order in Spanish can vary extensively for structural reasons (e.g., object clitics combined with conjugated verbs need to be placed in a preverbal position) or for pragmatic purposes (Bates & MacWhinney, 1989). In the case of Spanish, the CM predicts that when morphological cues and word order do not agree, word order is ignored; the model predicts the opposite for English (Hernandez et al., 1994).

The weight that English-speaking L2 learners give to word order in their L1 (English) impacts the way sentences are processed in the L2 (Spanish); this includes sentences which do not follow a canonical SVO word order (i.e., L1 form-meaning mappings transfer to new structures in the L2). Therefore, elements in a sentence-initial position, such as object clitics in Spanish, are usually assigned an agent role in the sentence, even when the case morphological information encoded in the object clitic should prevent learners from interpreting a preverbal object clitic as the agent. However, because object clitics are rich in morphological variation, English-speaking learners who have little control over morphology have difficulties detecting them and processing them. They process word order as a cue more easily since processing word order is less taxing: the first noun is assigned the role of agent regardless of its ending.

**Cue strength readjustments and processing costs in L2 development**

How do L2 learners start to ‘distrust’ word order as the most reliable cue when

sentences in the L2 do not follow the SVO canonical order? According to the CM, learning a global construction in the L2, such as $O_{cl}$VS sentences in Spanish, is a gradual process (MacWhinney, 2008), which depends on how the L2 cue strength readjusts to L2 cue validity.

Whereas for the processing of canonical sentences it has been found that beginner learners who speak an L1 typologically similar to Spanish do not show "morphological transfer" when learning Spanish since the processing of morphological markers is cognitively taxing (Sagarra, 2014), the CM proposes that, at the early stages in learning, English-speaking learners, who do not rely on morphological markers in their L1, exhibit L1 transfer effects (i.e., preference for word order) when processing non-canonical sentences in a target language such as Spanish. This phenomenon, which has been well documented by a large body of CM research has been called "syntactic accent" (e.g., Bates & MacWhinney, 1981; Kilborn and Ito, 1989; McDonald, 1989).

The CM predicts that, as learners gain experience with the L2, they progressively "fine tune" the L2 cue strength ("a property of the learner") so that it better corresponds with L2 cue validity ("a property of the linguistic environment") (Hernandez et al., 1994, p. 422). In other words, while the process begins with L2 cue weight settings that are close to those of the L1, over time these cue weight settings start to change in the direction of the L2 cue hierarchy (MacWhinney, 1997, 2002, 2008). Based on this prediction, learning the correct interpretation for $O_{cl}$VS sentences in Spanish would entail moving away from overwhelming reliance on word order (or an 'SVO first-noun strategy,' according to Reyes & Hernandez, 2006) towards progressive reliance on morphological cues (e.g., agreement morphology), which are much more valid in Spanish.

Since readjusting the hierarchy of the L2 cue strength necessitates the progressive resetting of L1 cue values, as learners garner cumulative experience with the L2, morphological information is expected to receive more attention and gain strength. That is, it is expected that, over the course of L2 development, learners reach a state of differentiation between the L1 and L2 hierarchies (i.e., they use a separate set of strategies for each language), and that they exhibit optimal form-to-functions mappings as a result of increasing experience with the L2. Nevertheless, even though positive correlations have been reported between L2 proficiency and sensitivity to valid grammatical cues in the target language (Kilborn and Cooreman, 1987; McDonald, 1987), cumulative language experience (i.e., exposure or proficiency) is not necessarily accompanied by the disappearance of L1 transfer effects (Sanz, Park & Lado, 2015). In fact, it has been argued that *forward transfer* (i.e., the use of L1 strategies in L2 processing) seems to be the most common pattern amongst late L2 leaners (e.g., Kilborn & Ito, 1989; Kroll & Stewart, 1994; Nicol & Greth, 2003; McDonald, 1989; MacWhinney, 2005), to the point that its effects have been encountered in early bilinguals, and in some individuals who have been immersed in an L2 environment for more than 30 years (Bates & MacWhinney, 1981). Interestingly, *partially* differentiated patterns of *amalgamation* (i.e., a single set of strategies derived by merging L1 and L2 hierarchies) have also been found amongst early bilinguals (e.g., Hernandez, Bates & Avila, 1994).

CM scholars (Hernandez et al., 1994; Hernandez, Sierra, & Bates, 2000) have argued that bilinguals' processing mechanisms exhibit *qualitative differences* in comparison to those of monolinguals. For example, Hernandez et al., (1994), who investigated the interaction of word order and number agreement in an aural picture agent-choice task amongst a group of early English-Spanish bilinguals, found that, in

sentences in which word order and agreement information conflicted, (e.g., *El perro las vacas están correteando*; 'The dog the cows are chasing' vs. *El perro la vaca está correteando*; 'The dog the cow is chasing'), bilinguals were better at deciding who the agent of the sentence was. However, when agreement showed no contrast (e.g., *El perro la vaca está correteando*; 'The dog the cow is chasing'), word order played a strong role in bilinguals' sentence interpretations.

What Hernandez et al., (1994) have shown is that bilinguals may exhibit a processing pattern of "in-betweeness," and that such amalgamation of L1 and L2 strategy use can emerge when bilinguals process non canonical sentences conveying morphological qualitative differences: bilinguals seem to prefer morphology as a cue for agenthood when contrastive morphological markers are present, and seem to prefer word order when contrastive morphological markers are absent in the sentence. The authors also found that the control groups, English and Spanish monolinguals, behaved as predicted by the CM: the former consistently relied on word order and the latter consistently relied on agreement cues. These findings for monolinguals were replicated by Hernandez et al., (2000), who found that when Spanish monolinguals processed sentences with clitics (e.g., *El perro lo(s) está(n) corretiando la(s) vaca(s)*; 'The dog $_{\text{cl dog}}$ is /are chasing the cow(s)'), case turned out to be even more useful than contrastive morphological information as a cue for agenthood.

How then can we explain the preference of bilinguals for morphology as a cue for agenthood when contrastive morphological markers are present in non-canonical sentences? As MacWhinney (1997) pointed out, what makes a cue such as agreement useful is not only its availability in the input but also its contrastive effects. For example, in the sentence *El perro la vaca está correteando*; 'The dog the cow is chasing,' verbal morphology does not contrast in number for either NP, as both NPs

are singular. Here, the agreement cue is available but not contrastive, and therefore it is less informative and more ambiguous (i.e., it leads to more than one possible interpretation: 'The dog is chasing the cow' or 'The cow is chasing the dog'). By contrast, in the sentence *El perro las vacas están correteando,* 'The dog the cows are chasing,' the verb only agrees with the second noun; the agreement cue is not only available, but is also contrastive.

Finally, the CM makes specific predictions for sentence processing costs (i.e., reaction times, RTs) depending on the characteristics of the input: sentences conveying competition amongst cues should result in slower RTs than sentences with several converging cues. As MacWhinney (2012) has argued, cue cost factors arise primarily during the processing of agreement markers, since these markers cannot be used to assign thematic roles directly. Therefore, in the case of NVN sentences exhibiting a preverbal object (e.g., *El perro están correteando las vacas,* 'The dog - obj. are chasing the cows- subj.'), contrastive availability may have an impact on cue costs. In this example, the object in preverbal position (a full NP), *el perro* (the dog), cannot be assigned the role of subject because the plural verb *están* (are) requires a plural subject and so the processor searches for a plural noun, which comes afterwards. Similar processing patterns are expected for $O_{cl}$VS sentences when the clitic is interpreted as the agent (e.g., *Lo están correteando las vacas*. 'Him - obj. is chasing the cows - subj.'). If learners are in fact able to integrate agreement information online, this additional waiting and matching should require higher processing costs than when agreement is not contrastive (MacWhinney, 2012).

**Gaps in previous literature and the significance of the present studies**

Previous SLA and CM studies investigating non-canonical processing behaviors in both English-speaking learners and early English-Spanish bilinguals

raises important research questions which require further investigation so as to broaden our understanding of how $O_{cl}$VS sentences are processed in L2 Spanish.

As predicted by the CM, and as succinctly captured by (Kroll, Dussias, Boquiski, & Valdes Kroff , 2012, p. 240), proficiency may be "the most critical variable in accounting for grammatical performance in the L2." Examining the processing behaviors of learners at higher levels of proficiency can inform the field of the qualitative characteristics of emerging bilinguals' processing mechanisms. In the case of $O_{cl}$VS sentences, even though research in this area continues to be fruitful due to the frequency and communicative value these structures have for Spanish language users, the available evidence regarding how English-speaking advanced learners process $O_{cl}$VS sentences in Spanish is scarce and circumscribed to the aural modality (Lee & Malovrh, 2009; Malovrh & Lee, 2010). Since modality has also been argued to play a role in how sentences are processed by L2 learners (Sagarra & Abbhul, 2013) (e.g., L2 research has shown that L2 readers may be less cognitively constrained than L2 listeners due to a time factor, Leow, 1993, 1995), it is important to evaluate the potential differences that may emerge in learners' $O_{cl}$VS sentence processing due to the effects of modality. Therefore, this dissertation sought to replicate whether advanced English-speaking learners of Spanish show high misinterpretation rates, similar to learners at other levels of proficiency –as suggested by Lee & Malovrh, 2009, and Malovrh & Lee, 2010 for the aural modality– when assigning semantic functions to NPS in $O_{cl}$VS sentences in writing.

When advanced learners process $O_{cl}$VS sentences in reading, do L1 processing patterns still play a role? Can learners clearly differentiate when to use L1 and L2 cue strength hierarchies at this level, as predicted by the CM? Moreover, is there a relationship between the quality of the morphological information conveyed in the

input sentences and the triggering of an L1 or an L2 processing strategies? That is, do specific agreement combinations help advanced learners override L1 processing biases? A critical issue this dissertation examines, which has not been systematically addressed in previous literature (e.g., Lee, 2000; Lee & Malovrh, 2009; Montrul, 2010; Malovrh & Lee, 2010; VanPatten, 1984; VanPatten & Houston, 1998, VanPatten & Borst, 2012) is whether difficulties with $O_{cl}$VS sentence interpretation (i.e., L1 transfer effects) are observed in all type of $O_{cl}$VS sentences, regardless of the quality of the morphological information those sentences convey, or whether, by contrast, as observed by Hernandez et al., (1994) in early bilinguals, advanced learners of Spanish exhibit a processing pattern of 'in-betweeness' (i.e., cross-linguistic variance) when interpreting $O_{cl}$VS sentences conveying contrastive morphological markers. The effort to address how L1 and the L2 cue hierarchies may coexist in the emerging bilingual's mind is in line with recent efforts to evaluate how bilinguals "accommodate the presence of two grammatical systems that may sometimes converge and sometimes conflict" (Kroll et al., 2012, p. 241). If the findings reveal that advanced learners do not behave alike when processing $O_{cl}$VS sentences conveying contrastive agreement as compared to sentences which do not convey contrastive agreement, this would offer new evidence that emerging bilinguals can also exhibit cross-linguistic variance *within* their L2 language.

This dissertation also investigated the role of two other external factors that are at the heart of SLA – computer-delivered feedback (i.e., error correction) and immersion experience – in order to better address the environmental influence on grammatical performance, that is, the preference for morphology as the most valid cue for interpreting $O_{cl}$VS sentences in Spanish.

Does the provision of feedback on accuracy responses optimize L2 processing

strategies? Does feedback concerning agent accuracy make relevant morphological information stand out more easily? According to McDonald's (1987, 1989) cue strength model, a change towards a more target-like L2 cue strength, which allows the development of L2 learners' comprehension capability, will only occur when learners realize that their sentences interpretations are incorrect. When that happens, the strength of the cues suggesting a correct interpretation (e.g., agreement morphology) will progressively increase. By contrast, if the interpretation is correct, no change in cue strength will take place. In order to test this prediction, it is necessary to address the effects of error correction on cue strength alignment in relation to the nature of the input itself (Sanz & Morgan-Short, 2004), and also in relation to cue-costs (Sasaki, 1998).

Is the sole exposure to $O_{cl}$VS sentences conveying contrastive morphological markers a sufficient condition for L2 cue hierarchy reconfiguration? Heilenman and McDonald (1993) suggest that exposure to non-canonical (conflict) sentences should facilitate L2 cue strength optimization. But, if feedback has an impact on how quickly non-canonical sentences are processed (that is, on processing costs, Sasaki, 1998), it may be the case that feedback on accuracy responses could make L2 cue strength alignment occur more rapidly. In other words, the study of feedback in relation to the nature of the input not only contributes to the question as to whether feedback is effective for the processing of complex structures, such as $O_{cl}$VS sentences in Spanish, by advanced learners (e.g., Han, 2002, 2008; Mackey, Gass, & McDonough, 2000; Long, Inagaki, & Ortega, 1998, Li, 2010), but it also contributes to the debate as to whether structured-input is sufficient to push learners to process the target structures and alter their processing strategies (e.g., Cadierno, 1995; VanPatten & Cadierno, 1993; VanPatten & Oikkenon, 1996; VanPatten & Sanz, 1995, Sanz &

Morgan-Short, 2004).

Since frequency-based exposure to L2 input seems to be crucial to parsing (Kroll et al., 2012) (i.e., parsing decisions are susceptible to change as result of daily exposure to the L2), this dissertation additionally investigated whether immersion experience in Study Abroad (SA) contexts contributed to the degree with which the L2 cue hierarchy was used. The potential impact of immersion experience on morphosyntactic development was analyzed taking into account the quality and characteristics of the Study Abroad program in which participants took part. If the parser's configuration is related to intense language experience, emerging bilinguals' parsing preferences should change as a function of the frequency with which the target structure appears in the environment (Kroll et al., 2012), and as a function of the extent to which the L1 is inhibited (Linck, Kroll, & Sunderman, 2009). As such, the characteristics of the SA program will determine whether the ideal conditions for meaningful and continuous contact with the L2 can be created. This dissertation focused on the study of a 5-week intensive SA program, whose specific characteristics (e.g., Language Pledge, interaction with the local community, and fieldwork activities) sought to create the ideal conditions for learners to live in the target language and suppress their use of their L1.

Even though, a rich system of social support that provides high quality input in an implicit learning environment and promotes engagement in the rich ongoing use of the language should maximize learning, including learning of morphosyntax (MacWhinney, 2002), there is much debate in SLA as to whether immersion experience in Study Abroad (SA) contexts impacts morphosyntactic development (e.g., DeKeyser, 2014; Grey, Cox, Serafini, & Sanz, 2015; Jackson, 2013). While past studies have suggested that the immersion experience does not necessarily lead to

gains in grammatical accuracy (e.g., Howard, 2001, Isabelli–García, 2010, DeKeyser 2010, Magnan, 1986), more studies need to be conducted in order to fill the gaps that still remain in this literature. It should be noted that most studies have drawn their conclusion based on oral data and comparisons between SA groups of L2 learners – mainly at the low or intermediate levels of proficiency – and control groups staying at home (SH). This comparison, although valid, has the potential of introducing confound variables into the research design. Moreover, these SA vs. SH comparisons support a theoretical perspective that considers learning contexts in terms of dichotomies (See Ortega & Byrnes, 2008, Ortega, 2013). By contrast, the aim of this dissertation was not to make dichotomous comparisons between processing behaviors (native vs. non-native), or between learning contexts; rather the goal of the research conducted here was to take a closer look at what learners *can do* as result of immersion experience. Are learners able to optimize form-to-function mappings when processing non-canonical sentences in general (i.e., even when the agreement markers conveyed in $O_{cl}$VS sentences are not contrastive)? If results show that learners tend to prefer morphology as the most valid cue for agenthood assignment, then this finding would support the idea that experience with the target language affects how the L2 is used and processed.

Another central aspect to consider is that, as LaBrozzi (2009) has pointed out, the tools and tests that have usually been used to measure grammatical gains in SA contexts are not always sensitive enough to measure L2 learners' processing changes (p. 13). For this reason, LaBrozzi (2009) employed an online procedure (eye-tracking) to study whether immersion experience impacted real-time L2 processing. In his doctoral dissertation (published later in LaBrozzi, 2012), the author was interested in investigating how a SA group of intermediate English-speaking learners of Spanish

processed morphological cues (tense markers) and lexical cues (adverbs) in temporal reference, in comparison to a group without immersion experience.

LaBrozzi's (2009, 2012) findings partially supported the claim that immersion experience affected learner's processing preferences since results showed that, even though SA learners began to rely more on morphological cues to assign temporal reference in Spanish, they still continued to rely on lexical cues. The lack of a strong effect of immersion experience on grammatical development may, however, be explained by the fact that data were collected from learners at approximately three months after coming back from their immersion experience and due to the level of proficiency the researcher tested (intermediate learners).

Since this previous study cannot discard a potentially larger effect on learners' processing patterns *immediately after* immersion experience, it is necessary to investigate the immediate impact which immersion experience may have on the processing of morphological cues, especially at higher levels of proficiency, since, at present, there is a dearth of studies investigating grammatical skill development at the advanced proficiency level in SA research (Grey et al., 2015). The SA study in the present dissertation used an online methodology (i.e., self-paced reading) in an innovative manner: accuracy and latency data were collected in a pre-post test design *in situ*, that is, in the same environment the group of advanced learners under investigation were immersed.

Since this dissertation sought to advance our understanding of $O_{cl}VS$ sentence processing in emerging bilinguals, latency data (RTs) were included in the three experiments conducted. In other words, a combined analysis of accuracy and latency data was maintained throughout. As Jiang (2012) argued, any mental event takes time, and "time is cognition" (Lachman, Lachman, & Butterfield, 1979, p. 130); therefore,

RT data uniquely provide a sensitive means of revealing what happens in the learners' minds during language processing. Accuracy data alone, the "outcome of processing" (Jiang, 2012, p. 10) (i.e., what is observable as result of learners' mental processes and strategy use), are not sufficient to fully capture how learners process the target structures. However, previous findings on learners' $O_{cl}VS$ sentence processing in Spanish have mainly been based on accuracy data (e.g., McCarthy, 2008), and specifically on overall accuracy data (accuracy data can also be analyzed over the course of the experiment). The previous SLA studies that have, in fact, used online procedures, such as self-paced reading (SPR) (e.g., Foote, 2011), Event Related Potentials (ERPs) (Bañón, Florentino, & Gabriele, 2014; Dowens, Vergara, Barber, & Carreiras, 2010) and eye-tracking (ET) (Grüter, Lew-Williams, & Fernald, 2012) have focused on different agreement phenomena (e.g., violations to subject-verb agreement, noun-adjective gender agreement, determiner-noun number) involving *canonical* sentences.

With respect to the on-line methodology employed in the three experiments conducted here, the notion that "some tasks are more on-line than others" (Jiang, 2012, p. 5) was taken into account, meaning that the different online methodologies currently available will vary in the degree of information they provide to the researcher, as they also will impact the type of inferences that can be made. For example, whereas a picture-matching task can provide researchers with overall latency responses to the stimuli (i.e., total reading times), eye-tracking (ET) offers a wide range of dependent measures associated with eye-movement behavior, thereby tapping into different type of processing information. Within the so called "late measures" in ET (Leow, Grey, Marijuan, & Moorman, 2014), 'regressions' have been argued to inform about immediate anomaly detections, given that readers often

respond to processing difficulties by regressing to earlier portions of the sentence (e.g., Liversedge, Paterson, & Pickering, 1998; Rayner, 1998) (See Dussias, 2010 for an overview of eye-tracking).

This dissertation employed a non-cumulative word-by-word self-paced reading (SPR) task as its on-line methodology since: (a) SPR is still the most fundamental experimental measure employed by psycholinguists interested in processing at or above the level of the sentence (Jegerski, 2014); (b) the basic premise behind self-paced reading is that the eyes can be a window on cognition, as proposed by the 'eye-mind' assumption (Just & Carpenter, 1980), which states that the amount of time spent reading a word reflects the amount of time needed to process the word.(Jegerski, 2014); (c) presenting one word at a time has the virtue of showing that "a processing measure is associated with each word in the text "(Rayner, 1998, p. 391): since it is difficult to skip words (regions), and because regressions to previous regions in the text cannot take place, the RT data collected indefectibly corresponds to each of the regions of interest; (d) because participants cannot look back at the text, the cognitive demands of the non-cumulative word-by-word SPR task are high (Rayner, 1998); therefore, learners need to efficiently use their cognitive resources and engage with the task in order to successfully complete it; (e) SPR is a suitable methodology to study morphosyntactic phenomena (e.g., when stimuli become inconsistent with participants' preferred interpretations, Jegerski, 2014). In fact, research investigating sentence parsing has obtained similar results when SPR and eye-movement data were compared (Ferreira & Clifton, 1986; Garnsey, Pearlmutter, Myers, & Lotocky, 1997; Kennedy & Murray, 1984; Trueswell, Tanenhaus, & Kello, 1993). The reason self-paced reading and eye movement data can be quite similar when syntactic phenomena are studied is because, in these instances, "processing has

been disrupted" (i.e., slowdowns emerge from the data in both methodologies), and "readers must use some type of problem-solving strategy to compute the correct meaning"(Rayner, 1998, p. 392).

Another aspect taken into account when addressing $O_{cl}$VS sentence processing from an on-line and off-line account, was to consider what off-line data were to be measured. Whereas many SLA scholars are interested in using both online and offline tasks in order to uncover online processing patterns during *comprehension* (i.e., form-to-function mappings) (Sagarra, personal communication), other researchers analyze on-line data in combination with accuracy data that do not necessarily focus on learner interpretations' of the entire sentence (form-to-functions mappings), for example, using yes-no questions pointing to words previously mentioned in the sentence (e.g., LaBrozzi, 2012). It is not uncommon to focus mainly on the online data outcomes when the sentences under investigation convey morphosyntactic violations (e.g., subject-verb agreement violation in canonical sentences, e.g., Foote, 2011); in these cases, the main goal of the researcher is usually to address whether grammatical knowledge can be used (or 'integrated') in real-time – in which case, sensitivity to grammatical incongruencies (i.e., changes in processing behavior) emerge in the data as higher reaction times (RTs).

The focus of this dissertation was to address whether changes in $O_{cl}$VS sentence accuracy (i.e., form-to-function mappings) correlated with specific processing patterns (e.g., higher RTs in contrastive agreement markers, attention to the verb region, and overall speedups). In this way, previous findings can be replicated, and the internal processes that learners employ in order to comprehend L2 data can be analyzed.

To summarize, this dissertation sought to offer new empirical evidence, based on accuracy data and latency data, on how English-speaking advance learners of Spanish process $O_{cl}$VS sentences in the written modality under different agreement and learning conditions. That is, this dissertation helps to advance our understanding of the type of processing profile emerging bilinguals exhibit when they process non-canonical sentences in Spanish. In addition, it informs the field about the role of the nature of the input and the environment in the adjustment of L2 cue strengths. The external factors under investigation (agreement morphology, computer- delivered feedback and immersion context) were studied in three experiments so as to offer a broad picture as to how these factors interrelate. Not only does this dissertation look at the role of proficiency and modality, but it also implements methodological innovations, such as the adoption of online measurements *in situ* when examining SA learners abroad. Moreover, accuracy data is analyzed overall as well as over the course of the experiments so as to better interpret learners' patterns of improvement. The findings reported here replicate previous 'offline studies,' based mainly on accuracy data (Lee & Malovrh 2009; Malovrh & Lee 2010), and they complement other studies' findings coming from different theoretical frameworks (i.e., formal approaches to SLA) which have investigated related phenomena but that are not of the same type as those considered here.

**Potential contributions to SLA theory and pedagogy**

What is of interest for SLA theory and pedagogy is under which conditions emerging bilinguals exhibit more target-like processing profiles; that is, which linguistic, pedagogical and contextual factors influence the optimization of L2 cue hierarchy and its development. Whereas most SLA research ascribing to a cognitive perspective on the study of $O_{cl}$VS sentences in Spanish has mainly focused on L2

learners at the early stages in learning (beginning and intermediate levels) – most probably due to an assumption that $O_{cl}$VS sentence misinterpretations are more problematic at these levels – the experiments conducted in this dissertation look mainly at the processing strategies of English-speaking advanced learners of Spanish. Additionally, while most SLA studies have tested $O_{cl}$VS sentence processing in the aural modality and have based their discussions mainly on off-line measurements (overall accuracy performance lacking time course information of sentence processing), the experiments conducted here address $O_{cl}$VS sentence accuracy in reading at the same time they offer a complementary online account of the processes taking place in the learners' mind. Lastly, whereas between-subjects designs have predominated in the study of $O_{cl}$VS sentence processing in Spanish, due to the interest of many studies in the impact of specific instructional interventions (e.g., pedagogical factors), it is also important to address whether there is intra-subject variation in the behavioral processing patterns of emerging bilinguals when they are confronted with different within-subjects agreement conditions. Therefore, building on findings from the CM framework, especially on those studies addressing cross-linguistic influence and bilingual processing behaviors, this dissertation represents a comprehensive effort to identify what external factors best account for differences in $O_{cl}$VS sentence interpretation employing both between-subjects and within-subjects designs. Identifying which factors significantly influence $O_{cl}$VS sentence interpretation is a fundamental question guiding research into adult L2 learners' non-canonical sentence processing. Results promise to contribute to SLA theory and pedagogy by investigating how successful advanced learners are at assigning morphological cues the weight they have in Spanish depending on the presence or absence of contrasting cues, the intensity of the input and the feedback provided; and how efficient they are

at integrating morphological information in non-canonical sentences during L2 sentence processing. Lastly, results contribute to the debate as to whether immersion experience has a positive impact on advanced learners' grammatical skills in the written modality, which has hardly been studied in this context because of the supremacy of oral development studies in Study Abroad research.

**Organization of the present studies**

This dissertation looks at three main external factors that may influence the readjustment of L2 cue weight in $O_{cl}$VS sentence processing: (i) the quality of the agreement information conveyed in the input sentences, that is, whether advanced learners are sensitive to the types of stimuli that are presented to them and whether they adjust their processing strategies accordingly (Chapter Two); (ii) the influence of the learning context, that is, whether immersion experience might have an impact on the stabilization of L2-based cue hierarchy over time (Chapter Three); and (iii) the role of feedback, that is, whether different training conditions might enhance the saliency of specific morphological cues and optimize L2 hierarchy use (Chapter Four). Each of these factors are presented as independent studies which relate to one another with the goal of providing a richer picture of the processing patterns emerging bilingual exhibit under different linguistic, contextual and training conditions. In addition, all the studies included in this dissertation share: (a) the same research methodology (non-cumulative Self Paced Reading tasks), which provides a considerable degree of information on how sentence processing unfolds over time; that is the RT data for each of the words read are submitted to statistical analysis (General Linear Models, GLM) so as to compare learners' responses in each of the experimental conditions. Such analysis allows for descriptions and explanations as to the role that the three factors under investigation may have on processing costs so that

a complimentary online account is provided along an offline account of accuracy data; (b) accuracy data is analyzed as overall means as well as across experimental items (i.e., longitudinally) in each of the experimental conditions; in this way, accuracy data is not only treated as an end-up product of sentence processing but as a dynamic trend of improvement, or the lack thereof, over the time course of the experiments; (c) all experiments in this dissertation test grammatical $O_{cl}VS$ sentences in the written modality conveying animate subjects and object clitics; (d) a within-subjects design (either a Latin square, a pre-post-test, or both) is included to examine participants' cross-linguistic variability in each of the experimental conditions under study.

Lastly, in the final chapter (Chapter Five), I discuss the main findings in relation to previous SLA and CM theoretical and empirical literature, draw conclusions based on these findings, consider implications for theory, assessment, and pedagogy, and conclude with limitations and suggest directions for future research.

## Definitions of procedures and terms

### Self-paced reading procedure

Self-paced-reading (SPR), the methodology employed in this dissertation, has been long used in adult native speakers (e.g., Schneider & Phillips, 2001) and child speakers (e.g., Sekerina, Fernandez & Clahsen, 2008) and it has been steadily growing in Second Language Acquisition (SLA), specifically in experimental studies investigating morphosyntactic sentence processing (e.g., Foote, 2011; Jiang, 2004, 2007; Sagarra & Herschensohn, 2010; 2011, 2012; Tokowicz & Warren, 2010, VanPatten, Keating & Leeser, 2012). As explained by Sagarra and Herschenson (2010), this procedure works as follows: Participants are shown rows of dashes such as the dashes below (Figure 1):

**Figure 1: Word-by-word self-paced reading procedure**

_____ _____ __ _____ _____ _____

Each dash represents a word in a sentence (i.e., the words are "masked") and are separated by spaces which represent the visible characters normally seen during natural reading. When participants press the space bar, the first word appears, and each time the space bar is pressed, a new word is revealed and the last word becomes dashes again, as seen in Figure 2:

**Figure 2: Example of word-by-word self-paced reading procedure**

The ____ ___ _____ _____ _____
(space bar press)
____ boys ____ _____ _____ _____
(space bar press)
____ ____ run _____ _____ _____

When a participant presses the key on the last word of the sentence, a comprehension question about the sentence is usually prompted, and participants mark their answer by pressing a button (e.g., "yes", "no"). The time spent reading each word is measured in milliseconds (ms.) as the time (i.e., Reaction Time, RTs) spent between two successive space bar key presses. Therefore, the dependent measure in SPR is the latency of button-pressing and processing (Jiang, 2007). It is assumed that as processing becomes difficult, participants slow down their rate of button-pressing, similarly to how in normal reading the eyes tend to rest longer before moving past a difficult-to-process word or region (Clifton & Staub, 2011). Therefore, the criteria readers use when they make a conscious decision to press a button to receive the next input word can be informative when analyzing the regions of processing difficulties

Depending on the research questions and the stimuli tested in the experiment, researchers decide a priori which of the words read represent the regions of interest. For example, if the research question addresses whether learners are sensitive to

agreement violations between adjectives and the nouns in Spanish as in Sagarra and Herschenson's (2010) study, (adjectives typically follow nouns in Spanish) (e.g., *el prototipo*-masc. sing. *famosa*-fem. sing, "the famous prototype"), the main region of interest is the adjective.

Researchers are also usually interested in analyzing the RTs of the regions immediately following the region of interest (RI), the so-called "spillover regions", since in these regions (usually one or two words following the RI) it is possible to detect late effects. These slowdowns in reading (i.e., higher RTs) are important because the onset of L2 processing can be delayed (Cf., Sagarra & Herschenson, 2010; Weber-Fox & Neville, 1996).

**Sentence processing**

Sentence processing is operationalized as form and meaning mappings on-line: that is, the mental processes involved while reading sentences in *real time*. Based on the CM, forms and meaning mappings are cue-driven, and when learners process L2 material they tend to rely on specific cues (e.g., word order) depending on which cue has a stronger weight in the learners' L1 (L1 cue bias effect). During L2 development, learners need to progressively adjust progressively the weight of a specific cue, such as word order, to make the correct form-function mapping during sentence processing. Many factors may influence cue re-adjustment, for example, input characteristics, such as their frequency, salience, and semantic value (what kind of cues are available in the sentence, e.g., contrastive agreement cues) and also the level of proficiency of L2 learners.

**L2 proficiency**

The proficiency level of L2 learners of Spanish in these studies is operationalized as institutional enrollment in advanced courses (6th semester); the

courses are offered through the Department of Spanish and Portuguese at Georgetown University. Spanish classes meet three days a week for fifty minutes each class period, for approximately fourteen weeks a semester. Students are enrolled in these courses either based on results of a summer or fall placement exam that has been in place for seventeen years (the majority of participants), or from performance on Advanced Placement (AP) exams, or from previous course enrollment. Additionally, in an effort to employ more robust proficiency assessment standards and to confirm that learners had homogeneous proficiency (Tremblay, 2011), participants were asked to complete the grammar section of an internal departmental Spanish placement exam.

## CHAPTER TWO: Contrastive Agreement

### Agreement, word order, and intra-subject variation in L2 Spanish sentence processing

### Introduction

SLA scholars have established that, in $O_{cl}$VS sentences, English-speaking adult learners – especially, but not only, in the early stages of acquisition – often misinterpret the initial object clitic as the agent instead of the patient of the verb (e.g., Lee 2000; Lee and Malovrh 2009, LoCoco 1987; Montrul 2010; Malovrh and Lee 2010; Sanz and Morgan-Short 2004; VanPatten 1984; VanPatten and Houston 1998; VanPatten and Borst 2012). For that reason, authors have looked at different external factors that may alter the processing strategies L2 learners use when interpreting $O_{cl}$VS sentences in Spanish. For instance, previous studies that investigated the potential benefits pedagogical techniques such as input enhancement (Lee 1987) and processing instruction (PI) (e.g., Sanz 1997; Sanz and Morgan-Short 2004; VanPatten and Cadierno 1993) may play in helping learners overcome processing biases, have been a productive and influential trend in word order research in L2 Spanish. This literature, however, has left lacunae worth investigating, including the development of processing strategies beyond the initial levels of acquisition, and the role of clitic/verb contrasting morphology in disambiguating semantic function assignment.

Several differences distinguish the present study from previous research. First, since this study adopts a functionalist approach (the Competition Model, Bates and MacWhinney 1981, 1987, 1989; MacWhinney 2012), rather than a competence approach to language processing, its main focus is not to make dichotomous

comparisons between processing behaviors of native and non-native speakers. Rather, the goal of this research is to take a closer look at what advanced learners of Spanish, who we will refer to as emerging bilinguals, *can* do when processing $O_{cl}$VS sentences exhibiting different agreement manipulations; that is, this study accounts for intra-subject processing variation.

Second, previous studies which have investigated $O_{cl}$VS sentences in the written modality have either excluded advanced learners from their sample or have not focused on questions related to the impact that clitic/verb contrasting morphology may have on $O_{cl}$VS sentence processing (e.g., Montrul, 2010; Montrul et al., 2008).

Finally, none of the studies in the previous literature have provided processing information concerning sentence-reading time: instead, reaction time data have been addressed only in terms of the latency of responses to comprehension questions. For instance, McCarthy (2008) did include the examination of $O_{cl}$VS sentences in advanced L2 learners, but the author looked mainly at production, measured offline, and the clitics conveyed in the $O_{cl}$VS sentences analyzed in that study were limited to sentences with inanimate referents. Other studies which have used online procedures such as Self-Paced Reading (SPR) tasks (e.g., Foote, 2011), Event Related Potentials (ERPs) (Bañón, Fiorentino, and Gabriele, 2014; Dowens et al., 2010) and eye-tracking (ET) (Grüter, Lew-Williams, and Fernald, 2012) addressed different agreement phenomena (e.g., violations to subject-verb agreement, noun-adjective gender agreement, determiner-noun number) in sentences with canonical word order only.

The main interest of this study is to investigate how English-speaking advanced learners of Spanish process $O_{cl}$VS sentences under different agreement conditions between the clitic and the verb. However, in order to answer this question,

this study first looks at previous studies that examined whether advanced learners are able to make use of agreement information online, another question this study addresses. On this issue, previous work by Foote (2011) has attempted to determine whether English-speaking advanced learners, whose L1 exhibits agreement morphology, were able to integrate number knowledge in Spanish during online sentence processing. To test this, Foote used a word-by-word Self-Paced Reading (SPR) task with canonical (SV) sentences that contained subject-verb number violations. SPR word-by-word methodology measures how much time learners spend reading at their own pace each of the words (regions) in a sentence; this allows researchers to identify which regions present processing difficulties. The assumption is that the longer it takes learners to move from one word to the next, the more costly the processing of cues in that region is. In her study, Foote found that when advanced learners encountered SV ungrammatical sentences, they exhibited slowdowns (i.e., higher processing costs) in the verb region. Such disruptions in the learners' reading comprehension were attributed to be evidence for advanced learners' sensitivity to agreement mismatches and, therefore, their ability to quickly integrate number morphology and compute agreement information online.

Does the sensitivity identified by Foote in SV structures translate into enhanced abilities to process clitic and verbal morphology in non-canonical transitive sentences for learners at this proficiency level? Previous work suggests that $O_{cl}$VS sentences are challenging for L2 speakers even at advanced levels of proficiency. For instance, Lee and Malovrh (2009) and Malovrh and Lee (2010) investigated the role of clitic morphology in clitic referential resolution in $O_{cl}$VS sentences in the aural modality; their studies provided empirical evidence suggesting that English-speaking

learners who have reached high levels of proficiency in Spanish still present difficulties interpreting these sentences.

In sum, previous research suggests that L2 learners find sentences with non-canonical word order challenging, even at advanced levels of proficiency. By contrast, L2 advanced learners show sensitivity to agreement information in canonical sentences. This chapter addresses the question of whether advanced learners of L2 Spanish were able to use agreement information to overcome word order bias when assigning agency to non-canonical sentences. To confirm that agreement information can be processed, a college population was selected that had an L1/L2 combination similar to Foote (2011); Lee and Malovrh (2009) and Malovrh and Lee (2010). The potential interaction between word order and agreement cues was examined by employing a within-subjects design that allowed us to compare accuracy rates and processing costs of $O_{cl}$VS sentences conveying different agreement conditions: two conditions presented contrastive agreement cues between the clitic and the verb, and the other two did not. By comparing learners' online and offline performance in both contrastive and non-contrastive agreement conditions, we were able to address whether emerging bilinguals make use of contrastive agreement morphology between the clitic and the verb to overcome word order bias.

## Background

### The role of morphology in $O_{cl}$VS sentence interpretation in previous studies

Table 1 (see page 33) offers an overview of the previous work and perspectives adopted in the study of $O_{cl}$VS sentences in SLA. A handful of IP studies investigating reference resolution have looked into the effects of clitic morphology manipulation. Lee (1987) found that plural inanimate object clitics in the third person

were interpreted as the subject significantly more often than singular inanimate object clitics (*lo/la;* "it"). Lee attributed such higher error rate for plural clitics to Slobin's (1979) *Operating Principle A*, observed in children, which stated: "Pay attention to the end of words" (p. 108). His interpretation of Slobin's principle for adult L2 learners is that, at the lower levels, the learners' syntactic processor is only able to process "one bit of morphological encoded information" at a time (p. 227). In this view, since plural markers are morphologically more complex than singular markers because they include an additional morpheme (*-s*), such morphological complexity would "block" a felicitous referential resolution between the clitic and its antecedent, resulting in higher misinterpretations. Based on this idea, clitic plural forms would be less salient (i.e., less 'detectable') than clitics in the singular forms. Malovrh (2006), who provided learners with a contextual sentence containing a familiar antecedent (one of the characters of the Simpsons) reported similar findings to Lee's (1987) but for animate direct object clitics: $O_{cl}$VS sentences with singular objects (*lo/la*, "him/her") received higher accuracy rates in reading.

Lee and Malovrh (2009) and Malovrh and Lee (2010) tested how third and fifth semester students, as well as advanced learners (students enrolled in a sixth semester Spanish course or in an upper Spanish course, e.g., Introduction to Spanish Linguistics), processed $O_{cl}$VS sentences. One of the independent variables they studied was clitic number morphology. The authors found clitic number morphology has no effect on accuracy when *aural* $O_{cl}$VS stimuli were presented to leaners. The authors claimed that one of the reason their results did not corroborate Lee's (1987) and Malovrh's (2006) outcomes was that proficiency may have been an influential factor, and that the effect of clitic number morphology might only be seen "at the very initial stages of language" (p. 114). Whereas the available research on the role of clitic

morphology on clitic resolution has found that number is an important factor in the how learners makes computations between the clitic and its antecedent, and that this computation seems to be mediated by proficiency, one question that arises is whether number manipulations on the clitic and on the verb (match/mismatching conditions) play a role in $O_{cl}$VS sentence interpretation. If learners assign the clitic the role of agent, do they expect clitics to 'agree' with the verb?

Of interest for this dissertation, Lee and Malovrh (2009) and Malovrh and Lee (2010) claimed that the object clitic system only begins to restructure at the advanced levels, suggesting that difficulties in $O_{cl}$VS sentences persist up to the higher levels or, in terms of the CM, that advanced learners still use L1 cue strength hierarchy to process $O_{cl}$VS sentences. In their two studies, advanced level students processed aural $O_{cl}$VS strings in the third person at either a 60% or 78% level of accuracy depending on whether they were enrolled in a sixth semester or in an upper level course (Malovrh & Lee, 2010, p. 242). By contrast, the clitic system at the lower levels (third and fifth semesters) was much more unstable: accuracy was at around or below 50% accuracy. The researchers called for replication of these findings so as to further address the developmental trajectory of $O_{cl}$VS sentences.

**TABLE 1: Studies on word order ($O_{cl}$VS structures) in L2 Spanish**

| Study | Level | Examples | Framework | Modality | Variables | Assessment & Results | Limitations |
|---|---|---|---|---|---|---|---|
| Houston (1997) | Fourth-semester *N*=28 | *A Raquel la contrata Don Pedro*. " Don Pedro hires Raquel" | Input Processing (IP) / Competition Model (CM) | Aural interpretation task. | +/- background knowledge, known and unknown characters, (between subjects design) | A verb in English was provided with two lines on its side where learners wrote the names of the subj. & obj. A significant effect for background knowledge was found. | Participants in the "Background knowledge" group knew the plot of the story (who did what to whom) |
| VanPatten and Houston (1998) | Third semester *N*=46 | *Ricardo está enojado porque lo insultó Susana en la reunión*. "Richard is angry because Susan insulted him in the meeting" | Input Processing (IP)/ Competition Model (CM) | Aural sentence interpretation task | +/- preceding sentence-internal context (between subject design) | As in Houston (1997), an effect was found for context. But only eight participants made use of context consistently. An effect for verb type was also found (e.g., *attack*, *insult* and *reject*). | The contextual sentence might have given extra information. In the example, it is more likely that Susan is the person who did the insulting since Richard is the one who got angry. |
| Lee (1987) | Second quarter | *Paco va a la biblioteca con la lección de geografía* [contextual sentence*] y <u>la</u> lee* . "Peter goes to the library with the geography lesson and reads it" | Input Processing (IP) | Reading sentence interpretation task | Number & gender between the antecedent and clitic- (1) sing subj. & sing obj. (2) pl. subj. & pl. obj.; (3) subj. & obj. = gender; (4) subj. & obj. ≠gender. Objects underlined. | Question "What does *lo /la /los /las* refer to?" Misinterpretation rate: 80 %. Plural object pronouns (*los/las* "them") were interpreted as subject significantly more often than the singular form (*lo/la* "it") (objects are inanimate). No significant difference was found for gender | Objects are always inanimate and in some of the plural examples they are followed by a verb like vender (to sell) which requires two arguments (e.g., *Ana y María compran plantas y <u>las</u> venden a otras personas*) |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| VanPatten & Cadierno (1993a) | Second year *N*=80 | *(Al chico) lo saluda la chica..*"The girl greets the boy" | Input Processing (IP) (Processing Instruction) | *Practice PI:* reading/ aural sentence interpretation; reading text; *Assessment:* aural interpretation & written production | Type of instruction: PI vs. traditional. Control group (no instruction) | The traditional instruction group improved in production and the PI group improved in both (interpretation and production) | The assessment interpretation task was similar to activities learners in the PI received. Sentence reading interpretation was not included in the assessment. Online data not collected. |
| Sanz (1997) | Second year *N*=44 & Control group | El chico ve a la chica y entonces..... "The boy sees the girl and then..." | Input Processing (IP) Procesing Instruction | Oral & written Production for all tasks. | Effect of Instruction on Mode (oral vs. written) & Type of assesment task (sentence level vs. question-answer vs. video clip narration) Pre and post test design. | Sentence completion task with picture support. Structured Interview task, video clip narration. Subj. & obj. gender were different. Objects were animate/ inanimate. Higher accuracy in the written mode in both groups. The experimental group did better in the oral sentence completion task. | The topic of the structured interview task changed during the interview and it may have been more demanding than the sentence completion task. |
| Lee (2000) | First-semester *N*=40 | *La ayuda bastante.* "He helps her a lot." (fragment from a passage) | Input processing (IP) | Reading comprehension | Orienting pre-reading conditions (form: input enhancement; meaning & control)(between -subject design). Gender (*lo/la*) | Multiple-choice (e.g., "Who calls whom?") & sentence recognition task (for intake). Gender (masc. and fem.). In reading, an effect was found for orientation to form and for gender (*lo* higher than *la* =75% vs. 62%) | Attention was drawn to form and meaning in the form condition due to the directions given. Clitics refer to subjects/ objects in preceding sentences. Referents are sometimes far away from clitics. |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Sanz and Morgan-Short (2004) | First or second year *N*=69 | *Lo lame el perro.* "The dog licks it= the cat". | Input Processing (IP) Processing Instruction | *Practice:* reading /aural sentence interpretation; reading text *Assesment:* aural interpretation & written production | +/- Explicit Feedback +/-Explanation (between-subject) | Based on VanPatten & Cadierno (1999a): aural sentences interpretation task & written sentence completion task. Also written video-telling task (pre-post test design). No effect was found for feedback nor for explanation. Importance of task-essentialness | A control group was not included. Reading sentence interpretation was not included as an assessment task. Online data were not collected. |
| Malovrh (2006) *N*=58 | Beginner *N*=58 | *Bart no saca buenas notas, pero Lisa sí, y a veces lo critica Lisa*. "Bart doesn't get good grades, but Lisa does, and at times she criticizes him." (the Simpsons characters) | Input Processing (IP) | Reading sentence interpretation task | +/- topic familiarity (within subject) +/- duplicated phrase "a+ noun" (between - subjects design). Clitic & subj. *gender* (different /same) Clitic & subj. *number* (pl. or sing.) | + Topic familiarity as in Houston (1997)/ –topic familiarity & picture selection task. Higher accuracy for sing. clitic in the + topic familiarity condition. Morphology was a better predictor for improvement (logistic regression). No effect for gender nor for sentences containing duplication phrases. | Different assessment tasks for the same variable. Double amount of distracters for [- familiar]. No counterbalancing between sets. Plural object referents consisted of two names (e.g., *Patti y Selma*). |
| Lee & Malovrh (2009) & Malovrh & Lee (2010) | 3rd, 5th semester, & students in two upper courses *N*=52 (1st study); *N*=35 (2nd study) | *María habla frecuentemente con su papá. Lo admira María mucho.* "María speaks with her father frequently. Maria admires him a lot." | Input Processing (IP) | Aural sentence interpretation Task 1st, 2nd and 3rd accusative and dative object | Contextual cues (object-oriented, subject-oriented and neutral) and linguistic factors (case, person, number, gender and homophony) | As in Lee's (2000)(e.g., "who did the inviting?"), only the most advanced processed $O_{cl}VS$ strings in the 3rd person at 78%. Accuracy was between 47% and 60% for the other groups. No effect found for number but was for gender in the 3rd semester group. Lo/los (masc.) was interpreted more accurately. | Uneven number of person, case, number (singular & plural) and gender (masc./fem. & unmarked) tokens. Some sentences conveying masc. clitics were biased to one interpretation. |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| VanPatten and Borst (2012) | Third-semester *N*=42 | *Lo ve la profesora*. "The professor sees him". | Input Processing (IP) Processing Instruction | Aural sentence interpretation | Explanation Instruction vs. Structured Input alone (Processing Instruction) | Picture matching & grammatical sensitivity test (aptitude). Trials-to-criterion results showed that EI did not play a significant role in PI with word order and clitic object pronouns in Spanish. Grammatical sensitivity was not a significant factor. | Fixed order in the distribution of $O_{cl}VS$ sentences & SVO sentences could have caused order effects. Participants had exposures to 10 of the last items of the treatment. |
| Liceras (1985) | Third & Fourth year *N*=60 Native control= 5 | *Compra el periódico y lo comienza a leer*. "(He) buys the newspaper and starts to read it" | Generative (UG) | Oral & written production | L1 typology (French & English) | Tell a story & write a dialogue based on images (pre-post test design). Native-like placement of object pronouns for both groups. Clitics seem to have different values in non-native grammars: they are used as affixes and words. | Language experience was not taken into account. |
| Montrul (2010) [Similar to Montrul et al., (2006)] | L2 learners & Heritage speakers; Low/ Intermediate. *N*=48 & control group | *Lo llaman por teléfono sus padres*. "His parents call him by phone" | Generative (UG) | Oral production & reading interpretation task | Age of acquisition /language experience effects (early bilinguals =heritage vs. late bilinguals= L2 learners) | Storytelling in past tense; 5-point written acceptability task with clitics in different positions & timed visual picture-sentence matching task. Oral tasks & GJTs showed both groups had knowledge of clitic grammaticality constraint. HS showed more native-like comprehension and use than L2 learners. | RT responses in the visual picture-matching task were considered as a between-subject variable only. No counterbalancing for the different $O_{cl}VS$ structures. Target pair of pictures appeared four times. |

**On-line processing measurements in the study of $O_{cl}$VS sentences**

As shown in Table 1, $O_{cl}$VS sentences in L2 Spanish have been studied in both the aural and written modalities in comprehension and production. Learning outcomes have been reported by means of different assessment tasks which have provided us mainly with end-product (off-line) accuracy data (e.g., multiple-choice, sentence completion tasks, matching tasks, story telling). Few studies have used specific assessment procedures to track learner responses during the course of the treatment (Montrul, 2010; Montrul et al., 2006; VanPatten & Borst, 2012).

Montrul (2010) and Montrul et al., (2006) used a timed-grammaticality judgment task along with an online visual picture-sentence matching task. By pressing A or B on the keyboard, participants had to decide as quickly as possible which of the two pictures was described by the sentence. The time spent making that decision (RT data) was measured in milliseconds. Using a different online measure, VanPatten and Borst (2012) used a trial-to-criterion procedure to detect whether the provision of explicit information (EI) about the target structure speeded up $O_{cl}$VS sentence processing. Their procedure consisted of measuring the number of sentences it took learners to begin processing input strings correctly (i.e., correctly choosing between two drawings based on the aural stimulus they received). The criterion consisted of correctly answering three target $O_{cl}$VS items and at least one distractor SVO item in a row (e.g., for a participant who began answering correctly at item 12, her trials-to-criterion would be 11, that is, the number of items that preceded her correct responses). All participants had to perform at above 60% correct after trials-to-criterion. A potential limitation of this study is that the sentences in

the treatment were distributed in a fixed order so that $O_{cl}$VS sentences were followed by an SOV, in successive cycles (e.g., 1. OVS, 2. OVS, 3. OVS, 4. SOV, 5.OVS, 6. OVS, 7. OVS, 8. SOV, and so on). This distribution ensured that the measure of trials-to-criterion could be met. The regularity of this recurrent pattern, however, together with the adjacency of the experimental items (there were not enough mediating fillers), could have caused order effects. Compounding these issues, VanPatten and Borst (2012) used the last ten items of the treatment (in inverse order) as a pretest, creating a double exposure to these items.

As argued in Chapter One, the study of $O_{cl}$VS sentences in Spanish needs to be addressed both online and offline in order to fill the gaps in previous literature. Self-paced reading word-by-word is the methodology used in this dissertation since it allows for counterbalancing and randomization, at the same time that it also allows for the comparison of RTs in each of the regions of interest (e.g., clitic, verb, post-verbal subject) in each of the different experimental conditions. The study in this chapter addresses whether differences in the morphological information conveyed in the $O_{cl}$VS sentences have an impact on how English-speaking bilinguals process these structures.

**Theories, concepts, and approaches to agreement**

Agreement is a fundamental aspect of natural languages and language learning per se: the abundance of conceptions and approaches dedicated to its study bears witness to the importance of better understanding its developmental and processing mechanisms. For several decades, agreement has been a productive area of research in diverse linguistic fields such as L1 acquisition (e.g., Clahsen & Penke, 1992; Verhagen & Blom 2014), second/heritage language acquisition (Montrul, 2010), psycholinguistics (e.g.,

Haskell and MacDonald, 2005; Wagers, Lau, Phillips, 2009) and neuroscience (e.g., Barber & Carreiras, 2005; Van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005). In second language acquisition and bilingualism, the study of agreement is closely related to broader issues debated in the field. Depending on whether the approach to the study on agreement is formal or cognitive, questions are triggered, either related to the possibility of acquiring morphological features not instantiated in the L1 (e.g., gender acquisition for English speaking learners) or to the learner's ability to integrate morphological prior knowledge into the underlying processes that take place during L2 sentence processing.

Whereas recent empirical studies conducted on agreement in richly inflected languages such as Spanish have placed their focus primarily on internal agreement relationships within the noun phrase, that is, gender/number agreement between the article and the noun, or the noun and the adjective (e.g., Alarcón, 2009; Dussias, Kroff, Tamargo, & Gerfen, 2013; Keating, 2009; Lew-Williams & Fernald, 2010; Montrul, Foote, & Perpiñan, 2008; Sagarra & Herschesohn, 2010, 2011), fewer studies have been conducted to explore other agreement relationships, such as number agreement between the subject and the verb (e.g., Foote, 2011; Grey, Cox, Serafini, & Sanz, 2015; VanPatten, Keating, & Leeser, 2012) or number agreement in non-canonical sentences. The aforementioned studies have looked into agreement computations between the clitic and its antecedent in $O_{cl}$VS sentences in Spanish, as well as the work conducted within the CM amongst English-Spanish bilinguals discussed in Chapter One, (Hernandez et al., 1994; Hernandez et al., 2000) are exceptions. In short, Hernandez et al., (1994) found that contrastiveness between morphological cues in non-canonical sentences (e.g., *El perro*

*las vacas están correteando*; 'The dog the cows are chasing') seemed to help bilinguals override an L1 processing bias (L1 cue hierarchy use). However, it remains unexplored whether contrastive availability in the input (MacWhinney, 1997) makes a cue such as agreement useful for emerging bilinguals when interpreting $O_{cl}VS$ sentences in Spanish.

**What does agreement mean?**

What is understood by the term 'agreement' may be perceived as a simple matching task, but in fact, agreement is a complex phenomenon (Corbett, 1988, 1991). Corbett (1988, 1991) argued that the extent of the phenomena covered by the term "agreement" (i.e., what type of constituent agrees with another constituent) is in dispute. In a wide and traditional sense, that is leaving aside any directionality, agreement refers to the *matching of properties* between elements in specific syntactic configuration; adjectives are said to agree with nouns; verbs are said to agree with subjects (Humphreys & Bock, 2005). There are more detailed and strict definitions of agreement, in structural terms; for example, Lehmann (1988) and Moravcsik (1988) specified that in order to say that the verb agrees with the subject – the element that 'controls' the agreement relationship, the *controller* – in a specific category (e.g., third person singular) there must be a *syntagmatic relationship* between them (they must form a construction). The categories involved (e.g., person/ number) need to exhibit a paradigm with subcategories ($1^{st}$, $2^{nd}$, $3^{rd}$, for person; singular/plural, for number) and the subcategory specification of the controller, the subject, (e.g., singular) is independent of the presence or nature of the verb. This definition serves to illustrate that agreement is referential in nature and therefore redundant: it helps *identify* or *reidentify* referents (Lehmann, 1988, p. 55). It also shows that in order to make agreement computations, paradigms (i.e., subcategories)

are necessary (e.g., the first person contrasts with the third person, the singular form contrasts with the plural form), and to compute agreement, language speakers need to use a range of features such as markers of number and gender, and also person and case.

What the definitions above do not give a clear picture of is what agreement does in different circumstances (different linguistic constructions), nor do they indicate whether identifying the controller is really that straightforward. The general idea is that agreement is riddled with theoretical puzzles within and across natural languages. For that reason, research involving word order in the processing of agreement in a morphological rich language such as Spanish is a productive step to begin disentangling its mechanisms in L2 sentence processing.

**Agreement and word order with advanced learners**

Since agreement is derived from morphological paradigms, we expect advanced learners who have been continuously exposed to these paradigms by means of instruction or direct experience with the target language to get better at computing agreement, especially between the subject and the verb (Cf., Foote, 2011). Again, what is known about agreement in relation to word order in SLA, and more specifically in $O_{cl}VS$ sentences, is very little, since word order has not been fully studied in relation to different within-subjects matching conditions involving the clitic and the verb. As discussed above, most studies in SLA have centered on agreement relationships between the clitic and its antecedent, and agreement computations between a clitic and its antecedent are not of the same nature as those involving subject-verb agreement computations, since verbs and subjects are in a paradigmatic relationship. Reflecting on subject-verb agreement is central to this dissertation for the following reasons: on the one hand, in

$O_{cl}$VS sentences, learners still need to compute agreement with a post-verbal subject but, since object clitics take a sentence-initial position, they compete with the post-verbal subject in establishing agreement with the verb first.

A main question that this chapter explores is whether $O_{cl}$VS sentences are processed differently when the direct object clitic does not "agree" in number with the verb (e.g., *La miran los estudiantes,* her-cl.fem.sing. look at-3rd p. pl.- the students, "The students look at her"). That is, do agreement mismatches between the clitic and the verb trigger a reanalysis of the sentence, which involves the use of L2 cue strength hierarchy? It is important to note here that this question rests on the assumption that advanced learners are able to effectively use number morphology in the clitic and in the verb during real-time sentence processing.

The next sections present a discussion of studies on subject-verb agreement and word order in L1 Spanish and L1 English, as well as English-speaking learners' sensitivity to subject verb agreement violation in canonical sentences in Spanish.

**Subject-verb agreement and word order in L1 research**

How subject-verb agreement is computed has become one of the core questions in psycholinguistics. Structurally, much work in L1 English has been conducted on so-called "attraction error" or "proximity concord" phenomena (i.e., when the verb in the sentence erroneously agrees with the number of an intervening noun) (Quirk & Greenbaum, 1973; Zandvoort, 1961) (e.g.,"*The key to the cabinets are on the table", Bock & Miller, 1991). Whereas most research in this domain has been conducted in speech production, more recently, researchers have also looked at comprehension by comparing agreement attraction mechanisms cross-linguistically, in L1 Spanish and L1

English (Lago, Shalom, Sigman, Lau, & Phillips, 2015).

Research on subject-verb agreement and word order effects has also been conducted to test whether agreement is a "forward" or "backward" process; for example, Haskell and MacDonald (2005) compared L1 English speakers producing indirect questions and direct questions such as: "Can you tell me whether the horses or the clock is/are red?" vs. "Is/are the horses or the clock red?" Their study yielded the finding that English speakers inflected verbs so that they agree with the nearer of the two conjuncts (i.e., proximity effect). But this proximity effect disappeared when the linear order was changed in the subject-before-questions (i.e., indirect questions) (e.g., "*Can you tell me whether the horses or the clock are red?"). In these cases, speakers showed a tendency to produce plural verbs when the nearer noun was singular (PS condition). These results caused Haskell and MacDonald (2005) to ask themselves: "Why should linear order play a role at all?" One of the possible explanations the authors gave, based on the CM, was that some cues may be useful in agreement processing. They considered the *plausibility* of the relationship between the subject and the verb to be a potentially useful cue. Based on the analysis of a writing corpus, they considered that, in the case of subject-before-verb order, a proximity procedure is much less reliable in PS constructions than SP constructions. In this sense, plausibility and word order for English-speaking L2 learners might go in the same direction; that is, in English, because of the overwhelming presence of SVO sentences, it is much more plausible that verbs agree with a sentence-initial element.

A recent eye-tracking study on reading has looked at specific questions in L1 English to further investigate potential linear order effects. Dillon, Levy, Staub and

Clifton (2014) demonstrated a novel effect of linear order arising in sentences with a *wh*-object-V-S structure, such as: "Which flowers is the gardener planting?" They found that the mismatch in number between the fronted *wh*-object and the auxiliary had clear effects on both off-line acceptability judgments and on-line processing. Number mismatch between the fronted object and the auxiliary reduced acceptability of grammatical sentences; this effect was stronger when the object was plural. By contrast, number match between these elements increased acceptability of ungrammatical sentences. Eye-tracking data showed that these effects persisted in online measures past the point of disambiguation (i.e., spillover region). The authors suggested the existence of a *heuristic process* for computing agreement in English: comprehenders seemed to treat a noun in the constituent to the left of the inflected verb as a potential agreement controller (the noun that dictates the number of the verb), regardless of whether it is in the actual structural position of the subject.

Only one study retrieved in the literature has examined agreement computations in relation to word order in L1 Spanish. Antón-Méndez (1996) (cited by Hartsuiker, Antón-Méndez & van Zee, 2001) analyzed SOV sentence production such as *La jardinera los planta*, "The gardener plants them," to determine whether the verb agreed with the preverbal object pronoun. No object attraction errors were observed for this structure. However, Hartsuiker, Antón-Méndez & van Zee (2001) found for L1 Dutch that the mechanism that computes verb agreement is sensitive to number information outside the subject noun phrase. They showed, contrary to Antón-Méndez (1996), that objects exert an attraction effect and that the rate of agreement error increases when the object's grammatical number differs from the subject head noun's grammatical number.

They also claimed that "object attraction is not reserved for those objects that consists of a determiner and a noun: *Pronominal objects* are capable of attraction as well" (p. 563). This supports the idea that the form of pronouns is determined by access to the lexical representation of the antecedent (Cf., lexical hypothesis of pronouns, Meyer & Bloc, 1999).

**Subject-verb agreement in L2 Spanish**

Jiang (2004, 2007) argued that L2 learners cannot integrate number knowledge during online comprehension, as shown by a SPR word-by-word study he conducted amongst ESL (advanced) learners of L1 Chinese. Foote (2011) replicated these studies in English-Spanish bilinguals, arguing that Jiang's results could have been the result of L1-L2 transfer influence, since Chinese rarely makes use of plural morphemes. While English and Spanish differ in morphological richness, in both languages verbs must agree with their subjects in both person and number. The target items in Foote (2011) consisted of grammatical and ungrammatical sentences in which the subject noun and the verb were separated by intervening material (as in Jiang's sentences) (e.g., *El reloj del hombre es/ *son de Suiza*, "The watch of the man is /*are from Switzerland."); she also included sentences where the subject noun and the verb were adjacent (e.g., *Tu padre es/ *son de Texas*, "Your father is /*are from Texas"). None of the comprehension questions that followed the target sentences drew attention to or focused on any ungrammatical elements in the sentences. Results showed that the latter sentences (in which the subject noun and verb were adjacent) received a higher reading time (i.e., a slowdown) in the verb region. This study also showed that late bilinguals were not sensitive to agreement violations when there was intervening material (i.e., a prepositional phrase between the

subject and the verb). This distance effect was not explained by higher or lower working memory capacity (WM), an internal factor that Foote also measured in her study. Her explanation for this result was that participants *expected* a verb to appear after the subject, and sentences with intervening material between the subject and the verb did not fulfill this expectation. Her results did, however, support the idea that late bilinguals (advanced Spanish learners) can integrate number morphology in real-time reading comprehension since they are sensitive to subject -verb agreement violations, at least when the subject and verb are adjacent.

**Summary**

The findings that L1 research, as well as studies conducted amongst English-Spanish bilinguals, provide on subject-verb agreement is useful to broaden our understanding of how agreement mechanisms might work in non-canonical sentences and in the emerging bilingual mind, that is, when L1 and L2 cue hierarchies are triggered depending on the language and the specific input sentence to be processed.

L1 studies on subject verb agreement show that agreement between the subject and verb is complex. Many of these studies have been conducted on natural speech production but they offer valuable insights for reading comprehension as well. What is important to highlight from these studies is that: (a) word order may have particular effects on sentence processing under certain circumstances (e.g., when a constituent that is not the subject is to the left of the verb). These effects consists of (a) misanalyzing a sentence which is grammatical as ungrammatical (i.e., this is a specific type of object attraction error); (b) it seems that pronouns can also exert attraction error (Hartsuiker, Antón-Méndez & Van Zee, 2001). Therefore, the verb might be perceived to agree with the

number of the preverbal object clitic pronoun even though it is not the subject; (c) number mismatch between a fronted object and a verb may reduce acceptability of grammatical sentences (i.e., the sentence sounds "odd"); (d) the result of that oddness may come from relying on certain cues (e.g., plausibility) as proposed by the CM researchers (Haskell & MacDonald, 2005). It is more plausible in English that a verb agrees with a left constituent (L1 transfer effects).

Based on the evidence provided by the studies discussed above, it seems that there is an interaction between word order, agreement mechanisms, and reliance on certain cues (i.e., plausibility). According to the CM, it is expected that if agreement information has more weight in the L2, emerging bilinguals will rely more on clitic and verb morphology than word order as a cue if the L2 cue strength configuration is aligned with L2 cue validity. However, based on the concept of contrastive availability (MacWhinney, 1997), it is also expected that when sentences are less ambiguous, that is, when clitic and verb do not "agree," the information is useful to disambiguate agency assignment in $O_{cl}$VS sentences. A systematic exploration is needed of how word order and agreement cues influence L2 processing (L2 cue hierarchy use) when the language under investigation, Spanish, is morphologically rich.

**Spanish direct object as a special type of clitic**

Although experts in the study of clitics have argued that defining clitics is no easy task – the term itself is sometimes considered to be an "umbrella term," meaning that the category is complex and problematic for researchers – a traditional definition of a clitic is that it is a type of pronoun that behaves like a single-word syntactically, but also behaves like an affix, since it depends, in one way or the other, on an adjacent word (Cf., Zwicky,

1985). The clitics studied in this dissertation are direct object clitics in the third person forms, the properties of which are illustrated in Table 2.

**Table 2: Properties of third person direct object clitics in Spanish**

| **Properties** | **Examples** |
|---|---|
| They are inflected for *gender* and *number*:<br>Masculine-singular: *lo* (him)<br>Feminine-singular: *la* (her)<br>Masculine-plural: *los* (them)<br>Feminine-plural: *las* (them)<br>Bruhn de Garavito (1999) | *A veces **lo** saluda la profesora con algo de indiferencia*. "Sometimes the professor greets him with some indifference."<br>*Por la mañana **las** acompaña el padre a la escuela*. "In the morning the father accompanies them to school." |
| They refer to animate and inanimate nouns | *Ayer vi <u>la película</u> → Ayer **la** vi*<br>"Yesterday, I saw a movie → Yesterday, I saw it"<br>*Ayer vi <u>a María</u> → Ayer **la** vi*<br>"Yesterday, I saw María → Yesterday, I saw her" |
| They have accusative case | *Quien conoce a Susana sabe que todos **la**/*ella aprecian mucho*. "Anyone who knows Susan knows that everyone appreciates her/ *she." |
| They are unstressed, non-salient pronouns that are phonological dependent on the verb (Borer 1984; Suñer, 1988; Everett 1996; Sportiche 1996; Torrego 1995; Uriagereka 1995; Zagona, 2002) | *¿**Lo** estás viendo?* "Are you seeing it?" |
| When the verb is finite, they occupy a preverbal position (i.e., They appear in strict adjacency to the verb in an $O_{cl}$VS structure).<br>Exceptions: they can be attached to the right of imperative, gerund and infinitive forms. | **Pablo quiere **lo**. "* *Paul wants it" |
| They must refer to entities that are already present in the context. They are argued to involve a definite feature, which assures that a unique referent is identified in the context (Arche & Dominguez, 2011). | *¿Me **lo** puedes sostener?* "Can you hold it for me?" (*lo*, "it", refers to an object in the context) |
| No lexical material can intervene between the clitic and the verb (Jaeggli, 1982) | **Lo por tres horas buscó*. " *She looked for three hours for him." |
| They are homophonous with the definite article (e.g., *la*, 'the' sing. fem.) (Ellis, 2006) and with the complementizer of relative pronouns (e.g., *la que*, 'what' sing. fem.) (Lee & Malovrh, 2009) | ***La** chica* "the girl"<br>***La** conozco* " I know her"<br>***Los** gatos* en el patio "the cats in the yard"<br>***Los** alimento* " I feed them" |

| | |
|---|---|
| They may co-occur with phonological independent pronouns with which they agree. (i.e., clitic-doubling) (Cf., Sanchez & Al-Kasey, 1999) | ***La*** *vi* ***a ella***. " I saw her" |

As seen in Table 2, Spanish direct object clitics in the third person exhibit particular phonological, pragmatic, and morphosyntactic characteristics that make them complex in nature. Not coincidentally, these are the forms that have received the most attention in the SLA field. In contrast, English has direct object pronouns that are phonologically independent, always inflected for number, and exceptionally inflected for gender when the direct object pronoun refers to an animate sexualized identity in the singular: *her* and *him*. In this respect, only *her* and *him* carry a conceptual idea of sex.

## Experiment 1

### Rationale for the present study

The rationale for Experiment 1 was: if L2 learners keep relying on word order as a cue for the assignment of semantic functions to NPs in transitive sentences, then we would expect learners to present difficulties processing $O_{cl}$VS sentences in general, regardless of whether or not those sentences convey contrastive agreement between the clitic and the verb. However, if learners can make use of the contrastive value of agreement markers conveyed in the clitic and the verb, then we expect learners to behave qualitatively differently and use L2 cue strategies instead when processing $O_{cl}$VS sentences exhibiting contrastive agreement. If, for example, learners continue relying on word order to process such contrastive sentences, assuming they detect agreement markers in the clitic and the verb, an $O_{cl}$VS sentence conveying contrastive agreement would result in an ungrammatical reading: a sentence such as *Ahora lo están saludando*

*las chicas,* would be interpreted as "* Now he-3rd sing. are- 3rd pl. greeting the girls." Such misanalysis would be shown in higher reading costs (slowdowns) in the verb region and the regions directly afterwards (spill-over regions). $O_{cl}$VS sentences conveying mismatching conditions (when clitic and verb do not agree in number) would create "noise" in processing that would be reflected in higher accuracy rates for the mismatching conditions and would potentially trigger a faster recovery from erroneous mapping throughout the experiment (i.e., during practice). Since this experiment is a training experiment, implicit feedback was provided to all participants so as to create similar learning conditions to those in previous $O_{cl}$VS studies such as Sanz and Morgan-Short (2004) whose findings guided the conceptualization of this dissertation. If feedback is kept constant for all conditions, and contrastive agreement does not play a role in how learners process $O_{cl}$VS sentences, then all conditions should be processed similarly. By contrast, if lack of agreement between the clitic and the verb is qualitatively more informative and beneficial for L2 cue hierarchy activation, then similar accuracy rate improvements in both mismatching conditions are expected.

**Goals of the present study**

The goals of Experiment 1 was to determine whether English-speaking advanced learners of L2 Spanish: (a) rely on word order to interpret $O_{cl}$VS sentences under specific matching conditions (when agreement cues are non-contrastive and less informative, rendering sentences more ambiguous); (b) activate L2 cue hierarchy when they are presented with $O_{cl}$VS sentences conveying contrastive agreement cues (MacWhinney, 1997) (i.e., clitic and verb do not 'agree' in number); and (c) are sensitive to agreement violations in canonical SV sentences as suggested by Foote (2011). This SV sub-

experiment was incorporated in Experiment 1 as a control so as to examine whether emerging bilinguals are able to integrate morphological knowledge in SVs (canonical sentences).

**Research questions (RQs) and Predictions (Ps)**

- RQ1: Do advanced learners of L2 Spanish who are native speakers of English prefer word order for the interpretation of $O_{cl}$VS sentences during reading comprehension, as suggested by Lee and Malovrh (2009) and Malovrh and Lee (2010) for aural stimuli?
- P1: Due to the persistence of word order as a cue, we expect L1 transfer effects to continue at the advanced level in the written modality as well.
- RQ2: Given that advanced learners are sensitive to agreement morphology, can contrastive agreement cues help advanced learners of L2 Spanish who are native speakers of English come to the right interpretation of $O_{cl}$VS sentences?
- P2: When verb and clitic number mismatch (e.g., *Lo*$_{cl\text{-}sg}$ *están*$_{vb\text{-}pl}$ *mirando las chicas*, "The girls are looking at him"), we expect readers to be more likely to arrive at the correct interpretation of experimental sentences, resulting in higher accuracy rates. When verb and clitic match in number (e.g., *Lo*$_{cl\text{-}sg}$ *está*$_{vb\text{-}sg}$*mirando la chica*,"The girl is looking at him"), we expect L2 learners of Spanish who are native speakers of English to prefer word order information for the interpretation of $O_{cl}$VS sentences, resulting in lower accuracy rates in comprehension questions.
- RQ3: Do number mismatches between object clitics and verbs lead to lead to higher processing costs?

- P3: If learners misinterpret the clitic pronoun as the agent of the sentence, and if this creates an expectation that the clitic and verb should agree in number, we expect higher processing costs (i.e., higher RTs) when verb and clitic mismatch in number (e.g., *Lo*$_{\text{cl-sg}}$ *están*$_{\text{vb-pl}}$ *mirando las chicas*, "The girls are looking at him") as opposed to when they match in number (e.g., *Lo*$_{\text{cl-sg}}$ *está*$_{\text{vb-sg}}$ *mirando la chica*, "The girl is looking at him'"). We expect this difficulty to arise at the verb regions or in the regions after it.

## Methods

### Participants

Participants in this study (*N*=38) were native speakers of English (24 female; *mean age*: 18.94) and were advanced learners of L2 Spanish. Proficiency level in L2 Spanish was based on their institutional enrollment in an advanced Spanish course. All participants in this study were enrolled in an Advanced II non-intensive course at Georgetown University and were recruited to participate in this study in exchange for extra credit in their Spanish course. Students were enrolled in these courses either based on results of a fall validation exam, from their performance on a placement exam, or from previous course enrollment. All participants completed the grammar section of the Placement Exam created by the Spanish and Portuguese Department at Georgetown University (See Appendix A). The grammar section of the Placement Exam consists of thirty-seven 4-option multiple-choice items targeting a range of grammatical structures in Spanish. All participants obtained a score of 60% or more in the test (the maximum scored obtained was 86%) (*Mean*= 73 %; SD= 2.71; Skewness= -0.164; *Kurtosis*= -0.60); the value of skewness and kurtosis, which were within the normal range of +/- 2, and the

non-significance of the Kolmogorov- Smirnov test ($D(38)= 0.123$, $p= 0.15$), indicate that the means are normally distributed. In addition, the data presented no outliers: all mean scores fell within +/- 2 SD from the mean. Therefore, no participants were excluded based on this criterion. As to the results on the grammar placement test, one explanation for why learners did not approach 100% in the scores they obtained might be due to the fact that this section of the test conveyed grammatical structures that are difficult for acquisition such as tense, aspect and mood inflection and gender and number concord. Leaners also completed a language questionnaire based on Li, Zhang, Tsai, & Puls (2014) which addressed relevant language learning background variables (see Appendix B). Five participants declared that they spoke languages other than English: Hindi (*n*=1), Amharic (*n*=1), Hebrew (*n*=1), French, Chinese and German (*n*=1), and German (*n*=1). No participant spoke any other Foreign Language. In addition, to ensure that participants stayed on task throughout the experiment, only participants with at least 70% or higher accuracy in their responses on the comprehension questions for filler items were included in the analysis. Overall, from an original sample of 41 participants, three participants were excluded from analysis: two for not being a native speaker of English (*n*=2) and another for not meeting the 70% filler criterion (*n*=1).

## Design

### *$O_{cl}$VS sentences*

$O_{cl}$VS sentences consisted of 24 sentences arranged in a 2 x 2 within-subjects design, with *agreement* (match/mismatch) and *number* (singular/plural) as factors. Table 3 shows a sample of all target items. Accuracy and Reaction Times were the dependent variables.

For the *match* condition, number agreement between the direct object clitic pronoun and verb was manipulated: in the matching conditions, the clitic and the verb (the auxiliary *estar*, "to be") were both either singular or plural (i.e., they agreed in number), while in the *mismatching* conditions, the clitic was singular and the verb plural or, on the contrary, the clitic was plural and the verb singular (i.e., they mismatched in number). Verbs were in present continuous (auxiliary *estar* "to be" + gerund). The regions of interest were the clitic region, the verb region (the auxiliary verb), and the 'spillover' regions: the gerund (verb +1 region) and the determiner region (verb+2 region). $O_{cl}$VS sentence were preceded by a contextual phrase so that participants could establish a reference for the clitic (e.g., for the target sentence provided in Table 3: *El maestro y las chicas están el parque,*"The teacher and the girls are in the park). The true or false question that followed the sentence (e.g., for the target sentence provided in Table 3, the T/F question was: *El maestro está buscando a las chicas.* "The teacher is looking for the girls") assessed whether learners were able to interpret who did the action. The true/ false question pointed to the subject of the sentence, yet it avoided asking the question "who is doing the action" directly, as in previous studies (Malorvh, 2006). Participants received 1 point for correct responses and 0 points for incorrect responses.

***SV sentences***

SV sentences were included as a control sub-experiment to check whether participants were also able to use grammatical information online to detect subject-verb agreement violations in canonical sentences. The rational was that, if learners used agreement morphology to process agreement mismatches in both ungrammatical SV sentences (**El perro duermen*, "The dog sleep") and $O_{cl}$VS sentences conveying

contrastive agreement (*Lo*-cl. sing *están*-v. pl. *buscando las chicas*, "The girls are looking for him"), then they should present a slowdown in the verb region (or spillover region) in both sentence types (canonical and non-canonical).

Previous L1 and L2 researchers employing on-line methodologies to study morphosyntactic violations (e.g., agreement violations between the subject and the verb in SV sentences, Foote, 2011; Jiang, 2004, 2007; Wagers, Lau, & Phillips, 2009) have argued that higher RTs (i.e., slowdowns) in ungrammatical sentences can be interpreted as evidence of participants' sensitivity to agreement violations (i.e., use of morphological markers in real-time). Therefore, 'sensitivity' is the term used in this dissertation when discussing learners' online behaviors toward ungrammatical SV sentences. It could be argued that sensitivity to agreement mismatches also emerges in $O_{cl}$VS sentences conveying contrastive agreement between the clitic and the verb; however, since these sentences are grammatical, form-to-function mappings can be fully assessed. By contrast, taping into forms-to-function mappings in SV ungrammatical sentences is not entirely feasible. For example, in a sentence such as **El perro duermen*, *"The dog sleep," if learners were asked to decide who the subject of the sentence is (the dog or the dogs) the question would be ambiguous and potentially unresolvable, since the nature of the response depends on whether learners look at the subject or the verb in order to answer the question.

For the SV sentences in this study, reaction time was the main dependent variable and the T/F questions checked that participants were paying attention to what they were reading; the comprehension questions for the SV sentences pointed to words that were previously mentioned in the sentence; that is, SV sentences, unlike $O_{cl}$VS sentence, did

not focus on addressing how accurate learners were in their T/F responses. The main goal of these sentences was to check for learners' sensitivity to subject-verb agreement violations online. Therefore, no question pointed to the grammaticality of the sentence. SV sentences consisted of 24 sentences arranged in a 2 x 2 within-subjects design, with *grammaticality* (grammatical/ungrammatical) and *subject number* (singular/plural) as factors, as seen in Table 3. For the grammaticality factor, the agreement relationship was manipulated between the subject noun and the verb: in the grammatical conditions, the subject and the verb were either singular or plural (i.e., they agreed in number), and in the ungrammatical conditions the subject was singular and the verb was plural; or, in the opposing case, the subject was plural and the verb singular (i.e., they mismatched in number), rendering the sentences ungrammatical. Subject and verb were adjacent in all sentences, and verbs were all conjugated in the simple present. The regions of interest were the verb region and the spillover-regions, a preposition (verb +1 region) and the determiner region following the preposition (verb+2 region). (e.g., *En la cama*. "On the bed.")

**Table 3: Sample set of experimental items**

| |
|---|
| **Contextual sentence for $O_{cl}$VS sentences:** e.g., *El maestro y las chicas están en el parque,*"The teacher and the girls are in the park" |
| **$O_{cl}$VS sentence conditions:** |
| Mismatch, Cl-sg;.: *Ahora lo están buscando las chicas con impaciencia*<br>CONDITION A "The girls are now looking for him impatiently" |
| Match, Cl-sg..: *Ahora lo está buscando la chica con impaciencia*<br>CONDITION B "The girl is now looking for him impatiently" |
| Mismatch Cl-pl..: *Ahora los están buscando las chicas con impaciencia*<br>CONDITION C "Now the girls are looking for them impatiently" |
| Match, Cl-pl..: *Ahora los está buscando la chica con impaciencia*<br>CONDITION D "Now the girl is looking for them impatiently" |

**True/ False questions (comprehension) for $O_{cl}$VS sentences**:
*El maestro está buscando a las chicas.* "The teacher is looking for the girls" FALSE
*Las chicas están buscando al maestro.* "The girls are looking for the teacher" TRUE

**SV sentence conditions:**

Ungram, subj. sg.: * *A veces el perro duermen en la cama*
CONDITION A "*Sometimes the dog (sing.) sleep (pl.) in the bed."

Gram, subj.sg.: *A veces el perro duerme en la cama*
CONDITION B "Sometimes the dog (sing) sleeps (sing) in the bed"

Gram, subj. pl.: *A veces los perros duermen en la cama*
CONDITION C "Sometimes the dogs (pl.) sleep (pl) in the bed"

Ungram, subj. pl.: * *A veces los perros duerme en la cama*
CONDITION D "*Sometimes the dogs (plural) sleeps (sing.) in the bed"
True/ False questions for SV sentences:

*El perro duerme en el piso*. "The dog sleeps in the floor". FALSE
*El perro duerme en la cama*. "The dog sleeps in the bed". TRUE

The 48 sentence sets (24 SV and 24 $O_{cl}$VS) and 72 fillers were distributed across four lists in a Latin Square design (120 sentence total).

**Materials**

Sentences included highly frequent vocabulary from textbooks used in low-level courses (e.g., *Vistazos*). The subjects of the contextual sentence for $O_{cl}$VS sentences consisted mainly of common professions ending in *-o, -a, -or, -ora* (e.g., *abogada,*"lawyer" *historiadora,* "historian", *arquitecto,*"architect" ); in a few cases nouns ended in *–ista* (e.g., *tenista,* "tennis player," *futbolista, "*soccer player") or – e (*estudiante*, "student"); some common nouns such as *guía,* "guide" and *mujer,* "woman"; nationalities (e.g., *china,* "Chinese", *italiana*, "Italian", *alemán*, "German") and animals (e.g., *león,* "lion", *perro*, "dog", *orangután,*"orangutan" ) were also included. SV

sentences included intransitive verbs (e.g., *volver,* "come back," *correr,* "run") but not copulas like *ser* (as in two of the experiments in Foote, 2011).

The subjects and direct object clitic pronouns in the target sentences were animate, were either masculine or feminine, and were always counterbalanced so as to be of opposite genders (e.g., masc. subject, fem. direct object clitic). As argued by Sanz (1997, p. 54), counterbalancing the gender of the subject and the clitic direct object pronoun in the target sentence allows researchers to know whether learners are mapping the clitic as the subject or as the object. It also reduces ambiguity in the interpretation. Direct object clitic pronouns referred to one of the noun phrases (NPs) of the contextual sentence (either NP1 or NP2) (e.g., *El maestro y la chica están en el parque*. "The teacher and the girl are in the park"). The gender of the subjects in the contextual sentence was counterbalanced (i.e., masc.-fem; fem-masc.) to guarantee that the clitic and the gender of the subject in the target sentences were counterbalanced. In addition, the clitic referred to either the first subject (NP1) or the second subject (NP2) in a counterbalanced fashion as well.

Following previous literature on agreement and CM studies (Haskell & MacDonald, 2005; Hernandez et al., 1994; Hernandez et al., 2000), and since *estar*, "to be," is a highly frequent verb in the L2 classroom in the lower levels, the progressive present form (*estar*-"to be" + gerund) was chosen for all $O_{cl}$VS target sentences. The gerund's position in the sentence, following the main verb, was seen as advantageous since, as a result, the first spillover region did not coincide with the postverbal subject. A similar choice was made when an adverb was placed at the beginning of all $O_{cl}$VS

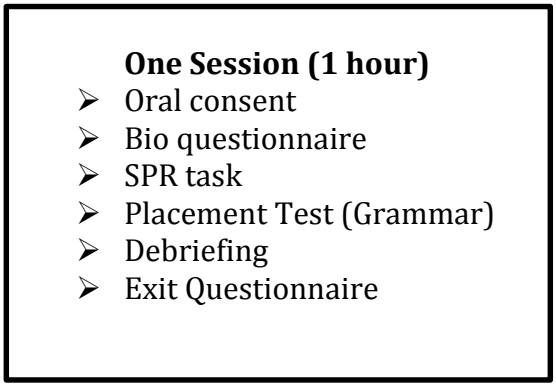
sentences so that the clitic was not put in the position of the first word: there is a tendency for readers to skip over the first word of a sentence (Rayner, 1998).

The fillers for both $O_{cl}$VS sentences and VS sentences consisted of SVO structures using animate or inanimate objects. Sentence were in simple present or in present progressive (as in the target $O_{cl}$VS items). Proper names were also included (e.g., *José,* "Joseph"; *Marta*, "Martha"). The True/False assessment task that followed the fillers consisted of a question pointing to the subject of the sentence (as in the $O_{cl}$VS sentences, true/false questions) or to words that appeared in the sentence (as in the SV sentences, true/false, question). Questions pointed to the way the action was done (e.g., *por enojo* "with anger") or to where the event took place (*por el parque*, "in the park") or to facts conveyed in the sentence (e.g., *Las coordinadoras aprueban los cambios en el program. Las coordinadoras rechazan los cambios*. False, "The coordinators approve the changes in the program. The coordinators reject the changes. False")

**Procedures**

This experiment took place in a silent room (Linguistic Laboratory) where no external noises or distractions could interfere with learner performance on the SPR task. A maximum of three participants were present at the same time in the laboratory. The study consisted of one session and lasted approximately one hour. See Figure 3:

**Figure 3: Description of experimental session**



As seen in Figure 3, after giving their oral consent to the researcher, participants completed a bio questionnaire (Appendix B). Afterwards, sentences were presented on a PC in a self-paced (SPR) word-by-word moving window paradigm (Just, Carpenter, & Woolley, 1982). Item blocking and randomization within lists was automatically managed using the reading-time software, Linger (Doug Rohde, MIT). The order of experimental and filler items was randomized for each participant. In each experimental trial, participants read a sentence with words masked by dashes. When participants pressed the space bar, a word was revealed and the previous word was re-masked. The time spent on each word was measured as the time participants spent between two successive key presses. The contextual sentences that preceded the target $O_{cl}$VS items were also masked, except that, unlike in the target items, participants were able to read the contextual sentence all at once. This contextual sentence disappeared after participants pressed the space bar to reveal the target $O_{cl}$VS sentence (word-by-word). Implicit feedback (correct/incorrect) was provided for every item ($O_{cl}$VS, SV and Fillers). Participants were instructed to read at a natural pace and answer the questions as quickly and accurately as possible by pressing one of two keys ("F" key for True or the "J" key for False) (the selection of these keys followed previous work in psycholinguistics, e.g., Wagers, Lau, & Phillips, 2009). Five practice sentences were presented before the experiment to ensure that learners understood the task (e.g., *La clase de historia tiene muchos alumnos. Ellos tienen diferentes nacionalidades* [contextual sentence]. *Los estudiantes son de una sola nacionalidad*. False, "The history class has a lot of students. The students come from many countries [contextual sentence]. The students come from one country. False"). None of the practice sentences were $O_{cl}$VS sentences, nor were

participants informed that SV sentences contained grammatical errors. Between each block, participants were given three breaks; they were also informed that a break could be taken at any point in the experiment. After completing the experiment, participants took the grammatical section of the Placement Test (Appendix A) and, finally, they were given an exit questionnaire (Appendix C) and were debriefed.

## Analyses

### Accuracy data

In the $O_{cl}$VS sentence experiment, the predictor variables for Accuracy were Condition and Trial order; and the dependent variable was accuracy scores. When Conditions are compared to one another in the agreement manipulation between the clitic and the verb, the condition in which the clitic and verb were both singular (Condition B) (i.e., the intercept) is considered the baseline.

### Latency data

Following recent studies conducted in the field of psycholinguistics (Hofmeister, 2011; Lago et al., 2015), only residual reaction times (i.e., residual RTs) were entered into statistical analysis. Residual RTs were calculated in the following way: a linear model used in R first predicted the expected RTs for each word in function of its length. This model was a linear regression based on the following formula: $y=a+bx$, where $y$ stands for RTs, $a$ stands for the intercept, $b$ stands for coefficient (slope), and $x$ stands for word length (number of characters in a word) (cf. Hofmeister, 2011). Second, the linear model computed the residual RTs by subtracting (i.e., "regressing") the expected RTs that the model predicted from the raw (observed) RT data obtained. Residual RTs were statistically estimated from the entire dataset (target items and fillers)

for all participants. This was done in order to control for differences in length between experimental conditions since in Spanish the plural suffix in the 3rd person plural in the clitic (-*s*) and in the 3rd person plural in the auxiliary *estar* (-*n*) are one character longer than the clitic and the auxiliary verb in the 3rd person singular.

**Statistical procedures**

The accuracy data in Experiment 1 (i.e., True/False questions for the $O_{cl}VS$ sentences) were analyzed using a binomial logistic regression in R, which is an open source programming language and environment for statistical computing (R Development Core Team, 2014). Binomial (or binary) logistic regression models deal with situations in which the observed outcome for a dependent variable (in this case accuracy of correct responses) is a dichotomous variable (correct, incorrect). One of the advantages of a Logistic regression is that it can include more than one predictor variable (independent variables); the model adjusts each predictor in relation to the impact of other predictors (removes confounding effects).

The latency data in Experiment 1 were analyzed using the *lmer* package in R, a general linear mixed model (GLMM), with Condition as the fixed effect and Subjects and Items as the random effects were run on the residual RTs. *Fixed effects* refer to the 'fixed' independent variables intrinsic to the study, and *random-effects* refer to the independent variables randomly sampled from the population of possible values, that are intended to be generalized to the whole population (such as subjects, e.g., learners). RT averages more than 2.5 standard deviations from the participant mean by region and condition were excluded (Ratcliff, 1993).

In fitting a linear model to a set of data, the model calculates a series of "weights"

(also called *coefficient*s), one weight for each independent variable.

The effect sizes used is in this dissertation were determined by the nature of the independent and dependent variables. Cohen's (1988) *d* is generally considered to be the effect size of choice for conveying the magnitude of experimental effects. However, when an experiment has a binary outcome (e.g., correct vs. incorrect) and the data are analyzed with logistic regression analyses, the odds ratio is the effect size generally used in these cases. The odd ratio expresses the magnitude of the effect in probabilities (Feingold, 2012; Fleiss & Berlin, 2009). Therefore, odds ratio were the effect sizes reported for accuracy data and Cohen's *d* were the effect sizes reported for the residual RT data.

Finally, as a general rule of thumb, odds ratios close to 1.0 represent a weak relationship between variables, whereas odds ratios over 3.0 for positive associations (less than one-third for negative associations) indicate a strong relationship (Haddock, Rindskopf, & Shadish, 1998). With regards to Cohen's *d* values, *d*=0.2 is interpreted as a small effect, *d*=0.5 is interpreted as a medium effect, and *d*= 0.8 is considered a large effect (Cohen, 1988).

## Results

### Accuracy in $O_{cl}$VS sentences

**Table 4: Means accuracy $O_{cl}$VS sentences x Condition (overall)**

| Condition | Mean | LowerCI | UpperCI | RT | SEM |
|---|---|---|---|---|---|
| a (sg clitic, pl verb) | 61.40 | 54.94 | 67.48 | 2877.66 | 107.10 |
| b (sg clitic, sg verb ) | 60.08 | 53.61 | 66.22 | 2900.93 | 108.79 |
| c (pl clitic, pl verb ) | 56.14 | 49.64 | 62.42 | 2836.57 | 151.75 |
| d (pl clitic, sg verb) | 58.33 | 51.84 | 64.54› | 2799.24 | 108.22 |

Table 4 shows the mean accuracy scores for $O_{cl}$VS sentences in each

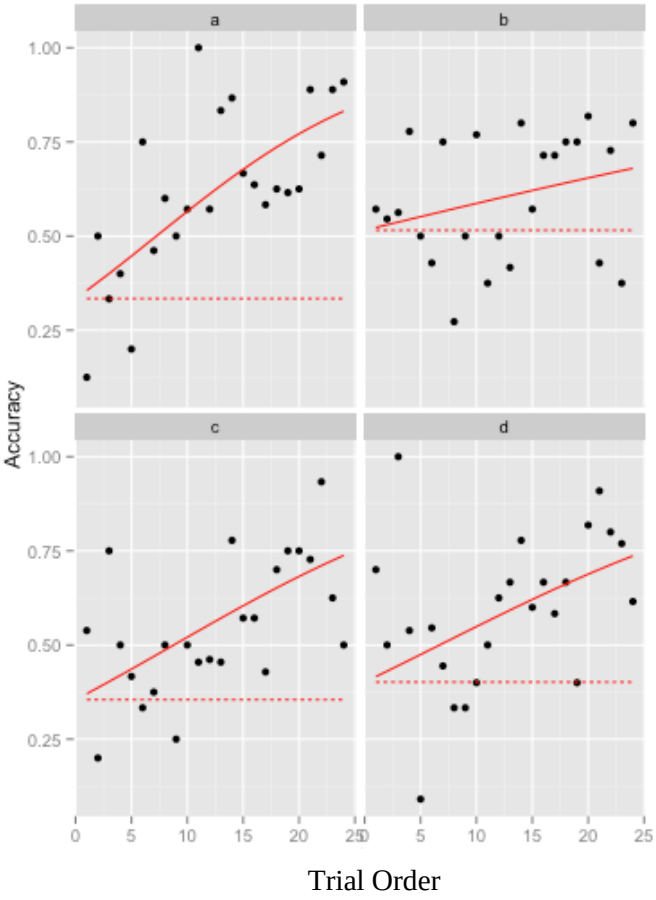
experimental condition. Overall accuracy for all conditions was below 70%, which supports the idea that $O_{cl}$VS sentences are still difficult to process for advanced emerging bilinguals. Accuracy in the mismatching plural verb condition (i.e., Cl sg.- V pl.) (*mean* = 61.40) was numerically higher than in the other conditions. The second condition which followed Condition A (cl. sing., pl. verb) in overall accuracy was Condition B (the baseline). To examine whether these differences were significant and whether any significant interaction emerged from the data, a logistic regression was run with Condition and Trial Order as predictors. The results are shown in Table 5 and represented in Figure 4:

**Table 5: Logistic regression for $O_{cl}$VS sentences (Condition × Trial Order)**

| | $\hat{\beta}$ | St. Error | *z* value | Pr(>\|z\|) |
|---|---|---|---|---|
| (Intercept) | 0.06229 | 0.27059 | 0.230 | 0.8179 |
| Trial Order | 0.02878 | 0.01965 | 1.465 | 0.1430 |
| Condition A | -0.75198. | 0.39572 | 1.900. | 0.0574 . |
| Condition C | -0.65972. | 0.39547 | -1.668. | 0.0953 . |
| Condition D | 0.46085 | 0.38626 | -1.193 | 0.2328 |
| Trial Order: Condition A | 0.06654 | 0.02918 | 2.280 | **0.0226 *** |
| Trial Order: Condition C | 0.03924 | 0.02856 | 1.374 | 0.1696 |
| Trial Order: Condition D | 0.03069 | 0.02785 | 1.102 | 0.2706 |

*Signif. codes*: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

**Figure 4: Accuracy (Trial Order × Condition)**




*Note*: the dots represent the target items, the dotted red lines represent the participant' competency at the beginning of the experiment, and the solid red lines represent the Trial order effects (the improvement over the course of the experiment).

The results from the logistic regression (Table 5) showed that no condition is different from Condition B (the baseline) overall. However a Condition x Trial Order interaction for the mismatching plural verb condition (i.e., Cl sg.- V pl.) ($\hat{\beta}$= 0.066, z= 2.28, $p$= 0.022) was found. As seen in Figure 3, learner accuracy on the comprehension questions improved over the course of the experiment. Importantly, in the mismatching plural verb condition (i.e., Cl sg.- V pl.), participants' improvement was significantly more pronounced than in the other conditions.

The effect size of the improvement over the course of the experiment (i.e., Trial

Order) was calculated for each experimental condition (Condition x Trial Order). Table 6 shows the standard coefficients, which correspond to the effect size in a logistic regression, and the odds ratio, which indexes the proportion by which the accuracy goes up as the order increases.

**Table 6: Effect sizes for $O_{cl}$VS sentences (accuracy)**

| | Std coefficient | Order effects | Odds |
|---|---|---|---|
| a (sg clitic, pl verb) | 0.46 | 0.66 | 1.93 |
| b (sg clitic, sg verb) | 0.20 | 0.20 | 1.22 |
| c (pl clitic. pl verb) | 0.27 | 0.47 | 1.60 |
| d (pl clitic, sg verb) | 0.21 | 0.41 | 1.51 |

As seen in Table 6, the mismatching plural verb condition (Cl sg.- V pl.) was, in fact, the condition with higher effect size (*std. coef.* =0.46, *odds ratio*= 1.93), whereas the baseline condition ( Cl sg.-V sg.) was the condition with the lower effect size (*std. coef.*= 0.20 , *odds ratio*= 1.22 ). The other condition where the verb also mismatched in number with the clitic showed improvement, (Cl pl.- V sg.), but improvement across the experimental trials was not significant, and this condition presented a lower effect size (*std. coef.* =0.21, *odds ratio*= 1.51) than the mismatching condition which was significant (Cl sg.-V pl.). The condition in which clitic and verb were both plural presented a higher effect size (*std. coef.* =0.27, *odds ratio*= 1.60) than the baseline condition and the mismatching singular verb condition (Cl pl.- V sg).

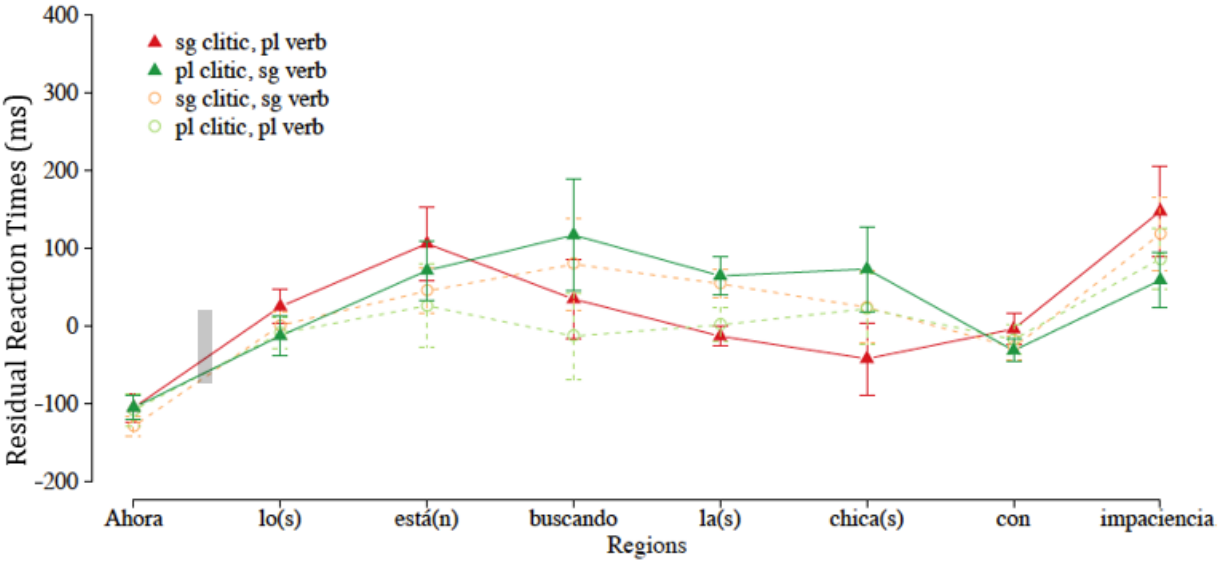
**Reaction Times in $O_{cl}$VS sentences**

Table 7 summarizes the Residual Reaction Times (ms) in each of the regions (words) in the $O_{cl}$VS sentences for all of the conditions; and Figure 5 provides a visual representation of this.

**Table 7: Residual Reaction Times (ms) x Region x Condition ($O_{cl}$VS sentences)**

| | Clitic | Auxiliary (*estar*) | Determiner | Noun |
|---|---|---|---|---|
| a (sg clitic, pl verb) | 24.41 | 105.37 | -13.60 | -42.56 |
| b (sg clitic, sg verb) | 0.47 | 44.52 | 53.73 | 23.49 |
| c (pl clitic. pl verb) | -10.06 | 25.58 | 0.99 | 21.72 |
| d (pl clitic, sg verb) | -13.18 | 70.79 | 64.05 | 72.63 |

**Figure 5: Mean Residual Reaction Times (ms) in each condition ($O_{cl}$VS sentences)**




*Note*: Sample sentence: "Now the girl(s) is (are) looking for him (them) impatiently."

As Figure 5 shows, the conditions in which verb and clitic mismatched presented higher RTs in the verb region (*mean*= 88.08 ms) in comparison with the conditions in which the clitic and verb matched in number (*mean* =35.05 ms). Finally, as shown in Table 6, the condition in which these learners showed significantly better performance in terms of accuracy (cl sg, V pl) also presented a higher RT in the clitic region (24.41 ms)

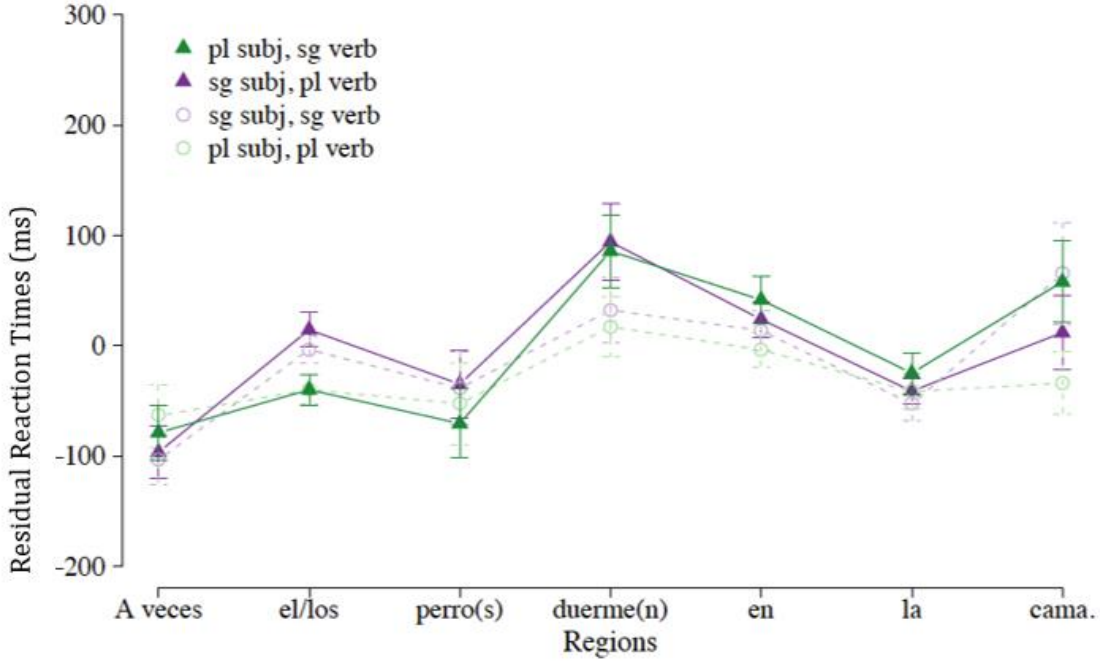
and a lower RT in the spillover region (the determiner of the postverbal subject) (-13.60 ms).

A mixed-effects linear model, with a fixed effect of condition and random intercepts and slopes by subject and item on the residual reading times showed no statistical differences in the RTs in the auxiliary verb region ($\hat{\beta}$= -25.52, $t$= -0.91) for the cl sg-V pl. condition; therefore, the higher processing costs observed in the auxiliary verb region for the condition in which the clitic is singular and the verb plural remain as a tendency in the data, which somehow suggests that this condition seems to be qualitatively more salient than the rest of the conditions.

**Reaction Times in SV sentences**

Figure 6 shows the reaction times for the experimental conditions in the SV sentences, which were included to control learners' ability to integrate agreement morphology during on-line comprehension in canonical word order sentences.

**Figure 6: Mean residual Reaction Times (ms) for SV Sentences**




*Note*: Sample sentence: "Sometimes the dog(s) sleep(s) in the bed."

As seen in Figure 6, learners showed a clear tendency to spend more time reading the verb region for the ungrammatical conditions, although this tendency did not reach significance. In order to shed more light on these results, we calculated the effect sizes for the verb region and the region right after (i.e., spillover 1) both in this study and in Foote's (2011) study, as seen in Table 8.

**Table 8: Effect sizes for ungrammatical sentences in Foote (2011) and the present study**

| Study | Region | Differences in means | SD | Cohen's *d* |
|---|---|---|---|---|
| Foote (2011) (Bilinguals) | Verb | 3 | 88 | 0.04 |
| | Spillover 1 | 79 | 137 | 0.81 |
| This study | Verb (singular) | 62 | 523 | 0.12 |
| | Spillover 1 | 10 | 246 | 0.04 |
| | Verb (plural) | 68 | 502 | 0.15 |
| | Spillover 1 | 45 | 322 | 0.16 |

Table 8 shows that, in comparison to Foote (2011), the differences in variance are larger in this study, suggesting a difference in participant's characteristics and/or differences in the target sentences. Additionally, whereas effect sizes were stronger in the spillover region in Foote's study, this study yielded higher effect sizes in the verb region than in Foote's, suggesting that these learners might also be sensitive to agreement violations. A power analysis we conducted yielded that, in order to have enough power to find the same effect size Foote found in the spillover region, this study would need no less than 614 observations (2.7 times as much data) and no more than 2454 observations (10.7 times as much data).

**Discussion**

The goal of the study was to provide both an online and offline account of how English-speaking advanced learners of Spanish process $O_{cl}$VS sentences conveying different agreement manipulations according to the presence or absence of number agreement between the clitic and the verb.

In relation to learners' accuracy in the interpretation of $O_{cl}$VS structures (RQ1), overall scores showed that reliance on word order when interpreting $O_{cl}$VS sentences is still prevalent among learners in their third year of Spanish coursework at the college level. This means that after years of exposure to the Spanish language, learners still exhibit difficulties interpreting $O_{cl}$VS sentences and that these difficulties can affect the correct interpretation of sentences containing object clitic (i.e., the patient is interpreted as the agent of the action denoted by the verb). These results extend Lee and Malovrh's (2009, 2010) findings in the aural modality. Yet, the tendency for learners to rely on word order does not seem to be entirely fixed: learners performed significantly better in the mismatching plural verb condition across the course of the experiment (i.e., Cl sg., V pl.), suggesting that the contrastive agreement cues between the clitic and the verb helped learners arrive at the correct interpretation of $O_{cl}$VS sentences.

This finding, that the improvement in accuracy was stronger only in the mismatching condition with a singular clitic and a plural verb, supports the idea that not all number contrasts are the same. Specifically, only plural number markers attached to the verb (rather than the clitic) proved helpful or qualitatively more beneficial in L2 cue hierarchy activation. One explanation for this may be related to the fact that clitics are functional morphemes, whereas verbs are lexical morphemes. As a result, detectability of

agreement cues may be enhanced since they are attached to content words. Likewise, the nature of the verb itself, in this case the auxiliary *estar*, "be," may also have played a role. As mentioned earlier, auxiliary verbs are a specific type of verb that learners encounter more frequently in the classroom (e.g., Guntermann, 1992), and this may have enhanced the saliency of the agreement cue, favoring processability. A final explanation is that, as previous research has found (McCarthy, 2008), learners seem to rely on morphological defaults in comprehension; therefore, they will be more predisposed to establish agreement between a plural element in sentence-initial position (in this case, a clitic, *los*-them) and a singular unmarked verb, which, in turn, would lead to a greater number of interpretation errors. This, however, would not seem to apply to the opposite pattern (e.g., *Lo están buscando las chicas*. "The girls are looking for him"), since plural verbs are marked, and consequently, agreement between a singular clitic and a plural verb appears less likely to the learner (or, from a CM perspective, less "plausible," following Haskell & MacDonald, 2003).

The online behavior (RQ 3) did not completely support the findings we obtained for accuracy data; that is, we cannot claim that contrastive agreement entails higher processing costs. However, it is worth noting that learners did show a slight slowdown at the verb region in the mismatch plural verb condition (Cl sg.,V pl.); in this condition, the verb region presented higher RTs in comparison with the conditions in which the clitic and verb matched in number.

Lastly, with respect to the experiment with SV sentences that was included as a control, RT data partially supported Foote's (2011) findings, since learners showed an overall tendency to spend more time on the verb region in the sentences that were

ungrammatical sentences, suggesting that learners were sensitive to subject-verb agreement violations and, in consequence, they were able to use inflectional morphology online. Even though this tendency in the RT data was not significant, the effect size observed in the post-verbal region in our study, the region where Foote (2011) found a grammaticality effect, is consistent with the effect size found in the post-verbal region in Foote's study. This supports the idea, once again, that learners were able to use number morphology online when they encountered agreement mismatches both in canonical and non-canonical sentences. Whereas in SV sentences, learners did not need to realign the L2 cue weights since L1 and L2 strategies converge, in $O_{cl}$VS sentences the use of agreement morphology seems to help learners readjust the L2 cue weights, since maintaining word order as the preferable cue for agenthood would lead to the incorrect interpretation of the sentence. Low power, participant characteristics, and the nature of the stimuli, may contribute to explaining why this tendency was not significant. Specifically, regarding the nature of the stimuli, Foote's ungrammatical sentences consisted of embedded sentences (e.g., *Veo que tu padre es/*son the Texas*, 'I see that your father is/*are from Texas") and one out of two of the experiments she conducted included the *ser* copula (a highly irregular verb in Spanish). Copulas are highly frequent in the input and even more in the classroom; in addition, the singular plural form (*es/son*) look and sound very different when compared to, for example, *come/comen* (he eats/they eat), the verbs used in the present study, and this salient nature of Foote's verbs may have helped learners detect agreement violations more easily, which facilitated their processing.

## Conclusions

The present study adds to previous literature on sentence processing in emergent bilinguals by showing English speaking L2 learners' reliance on the contrastive value of number markers between the clitic and the verb when processing $O_{cl}$VS sentences in Spanish. Specifically, our results show that contrastive agreement helps learners overcome word order bias, in that the L2 weight settings and the L2 cue validity tend to align more easily when the $O_{cl}$VS sentences convey contrastive agreement. These results support the findings in Hernandez et al., (1994) for English-Spanish bilinguals and extend the findings to classroom learners.

Results also show that integration of inflectional morphology during real-time sentence processing depends on the location of the plural morpheme, whether on the clitic or the verb. The question of whether this is due to a higher detectability of agreement cues in content words (i.e., verbs), to the nature of the verb itself, to markedness, or to plausibility seems to be a fruitful line of investigation for future research. For example, comparisons involving different types of verbs (auxiliary vs. non-auxiliary) would shed light on the issue of salience. Likewise, differences in contextual factors, such as study abroad immersion, and the provision or absence of feedback given to learners would also be informative as to the potential for pedagogical interventions to accelerate $O_{cl}$VS sentence processing.

In conclusion, the fact that emerging bilinguals are able to use contrastive agreement during $O_{cl}$VS sentence processing suggests that they do not necessarily behave in terms of non-canonical processing in the same way as learners in the lower levels would when processing non-canonical sentences. Research with different populations,

including emerging bilinguals enrolled in upper level Spanish courses, heritage speakers and Spanish native speakers could be beneficial to better understanding the potential differential effects of individual differences such as language proficiency and experience on the rate at which cue realignment takes place when processing transitive sentences that do not follow the SVO canonical word order in Spanish.

## CHAPTER THREE: Learning Context

### Evidence of morphosyntactic development in immersion contexts: An online and offline account of $O_{cl}$VS sentence processing in L2 Spanish

### Introduction

In recent years, research on the effects of study abroad (SA) on language development has seen a considerable increase in interest (DeKeyser, 2014) along with growth in the enrollment of SA programs, especially in short-term stays overseas (DeKeyser, 2014; Kinginger, 2010). Governmental initiatives encourage participation in Study Abroad (SA) programs since they are seen as vital to responding to future developments (Cf., Manley, 2014) in an ever increasingly globalized market and workforce. Therefore, considered as a whole, Study Abroad (SA) is a phenomenon expanding on several fronts: scholarly, programmatically and governmentally. The present study aims to contribute to the scholarly debate by addressing the main research question guiding this chapter, whether SA contexts influence optimal mappings of form and meaning in $O_{cl}$VS sentences in Spanish.

A recent study (Grey, Cox, Serafini, & Sanz, 2015) has identified several under-researched areas in the SA research agenda that need further investigation. Of interest for this study are: (a) the potential impact that short-term SA programs (currently, the most popular option amongst US college students) may have on L2 development; (b) to what degree level of proficiency moderates specific linguistic L2 skill learning gains, especially since advanced learners are currently underrepresented in the SA research; and (c) which L2 skills other than oral fluency (e.g., Isabelli, 2003; Freed & Segalowitz,

2004) may also benefit from the study abroad experience. Specifically, there is inconclusive evidence on whether immersion experience influences morphosyntactic development. The present study investigates the development of non-canonical sentences in advanced learners who took part in a short-term SA program to address the gaps indicated by recent literature.

The goal of this Study Abroad (SA) study is to examine *in situ*, i.e., in the same environment in which learners are immersed: (1) whether advanced learners of L2 Spanish studying in an intensive short-term study abroad program in Barcelona, Spain showed gains in their ability to interpret $O_{cl}$VS sentences correctly, and in their ability to detect and judge subject-verb agreement violations in SV sentences; (2) whether immersion experience speeds the degree with which learners make form-to-function connections in non-canonical transitive sentences, for which learners need to rely on morphology as a much more valid cue when interpreting $O_{cl}$VS sentences in different agreement conditions; in other words, whether morphological cues are detected and processed more easily so that agency assignment in $O_{cl}$VS sentences becomes less taxing for the learners, as shown by a reduction of time in processing costs at the end of the SA program. The results of this study intend to make valuable contributions to the field of SLA by tapping into learners' processes during real time sentence processing while also offering a more dynamic and detailed analysis as to how accuracy unfolds over the course of the experiments by means of statistical modeling (GLM).

The following sections provide an overview of the growing importance of SA in college education and the current debate in SA research in relation to the benefits study abroad represents for different L2 linguistic skills. The chapter then reviews previous

studies on L2 morphosyntactic phenomena and, specifically, L2 Spanish word order ($O_{cl}$VS sentences) in relation to SA learning contexts.

**Overview of previous SA studies**

L2 skill development in the SA context is an arena of ongoing debate (DeKeyser, 2014; Jackson, 2013). On the one hand, scholars such as DeKeyser (2014) have pointed out that in spite of the almost "magical image" of study abroad as the one and only way to achieve high levels of proficiency, or at least as a dramatic accelerator of linguistic development, there is empirical evidence that contradicts those beliefs (e.g., Collentine, 2004; Diaz-Campos, 2004). Researchers taking this position have argued that Study Abroad (SA) does not necessarily produce "measurable" progress in various areas (e.g., Howard, 2001, Isabelli–García, 2010, DeKeyser 2010, Magnan, 1986, for morphosyntax; Collentine, 2004, for lexical development; Avello & Lara, 2014; Mora, 2012, for phonology). According to DeKeyser (2013), even when clear progress is made, it tends to be in the area of fluency rather than accuracy or complexity (e.g. Freed, 1995; Mora & Valls-Ferrer, 2012; Serrano, Llanes & Tragant, 2011; Valls-Ferrer & Mora, 2014), unless the outcome measure is strongly conducive to monitoring (e.g., multiple-choice grammar tests). Moreover, as mentioned in Chapter One, CM scholars (Bates & MacWhinney, 1981) have found that some individuals who have had been immersed abroad over the course of 30 years, still exhibit patterns of forward transfer or L1 dominance in L2 processing: for example, they continue giving word order an L1 cue weight when interpreting non-canonical sentences.

On the other hand, there is also empirical evidence that Study Abroad (SA) experience benefits L2 linguistic skills other than oral fluency (e.g., Duperron, 2006;

Grey et al., 2015; Guntermann, 1995; Isabelli & Nishida, 2005; Yager, 1998, for grammatical accuracy; Dewey, 2008; Ife, Llanes, Tragant, & Serrano, 2011,Vives, Boix, & Meara, 2000; Milton & Meara, 1995, for lexical accuracy and complexity, Juan-Garau, Salazar-Noguera, & Prieto-Arrane, 2014, for lexico-grammatical competence). After all, the most cited approaches to SLA, including Krashen's Input Hypothesis (1985), Long's Interaction Hypothesis (1996), and Swain's Output Hypothesis (1985), are compatible with a favorable view of immersion contexts as ideal for L2 development (Sanz, 2014). In immersion contexts, learners are exposed to rich oral and written input, and much more discourse-level input; they usually take content courses not focused on language (i.e., implicit context) and these require extensive reading, as well as listening and writing in the target language (Lafford, 2006). Also, since learners have greater exposure to a variety of language events in formal and informal situations, the opportunities to interact (and negotiate for meaning) with locals and peers push learners to move from semantic to syntactic processing. Additionally, authors in psycholinguistics such as Linck, Kroll, & Sunderman (2009) have suggested that immersion can positively impact L2 processing during L2 immersion as a result of L1 inhibition, which is more prevalent in immersion than it is in classroom contexts. In CM terms, the L1 inhibition resulting from the intense exposure to L2 input in the immersion context should help learners to better "fine tune" their L2 cue strength configurations with the L2 cue validity, assigning morphology a higher order in the L2 cue hierarchy and leading to more accurate comprehension of $O_{cl}$VS sentences. This is a topic worth investigating, as so much weight has been placed on the role of the immersion experience on the development of L2 oral skills (e.g., Collentine, 2004; Hernández, 2010; O'Brien, Segalowitz, Freed, & Collentine, 2007;

Segalowitz, Freed, Collentine, Lafford, Lazar, & Díaz-Campos, 2004). In order to assess whether SA contexts have an impact on comprehension, and more specifically sentence comprehension in the written modality, it is necessary to include more comprehension studies in the discussion. This is one contribution this study aims to make.

**Immersion experience and morphosyntactic development in L2 Spanish**

Grey, Cox, Serafini, and Sanz (2015) investigated whether intensive language experience abroad benefited L2 lexical and morphosyntactic development, and whether development was related to learners' cognitive capacity (e.g., WM capacity). Their point of departure was DeKeyser's (1991, 2007, 2010, 2014) claim that, in order to maximize specific language skill proceduralization during a SA experience, L2 learners need a certain level of declarative language knowledge, and for that reason, they chose advanced L2 Spanish learners immersed in a 5-week study abroad program. Grammaticality judgment tasks (GJTs) and lexical decision tasks (LDTs), both in the visual modality were used to address these questions. Pre-post test comparisons of GJT results showed higher accuracy in word order (e.g., **Tengo que millas muchas correr*. * "I have to a lot of miles run") and number agreement items (**La falda es nuevas*, "*The skirt-sg. is new-pl".,) but not on gender agreement (* *El vestido es negra*, "* The dress-masc. is black-fem."); in addition, the authors found decreasing reaction times (RTs) overall, but RTs did not decrease for any of the three targets individually. For the LDT, performance on non-words, but not words, showed improvement, while RTs for both words and non-words decreased. These results led researchers to conclude that short-term intensive L2 exposure abroad significantly influenced morphosyntactic and lexical development.

Grey et al.'s (2015) findings in relation to morphosyntactic development in advanced learners who took part in an intensive short-term SA program provided empirical evidence that: (a) advanced learners improve their accuracy rate on Grammaticality Judgment Tasks (GJTs) which test certain types of word order manipulation (subordinates in this case), (b) overall processing costs (RTs) are reduced when learners process sentences after intensive SA exposure.

Why would immersion abroad, especially in the absence of explicit instruction, result in positive changes in $O_{cl}$VS processing among advanced learners who have been struggling with the structure for years? Reflection on the role of context of learning in the development of preverbal object clitic can be found in Andersen's (1983) and VanPatten's (1990) case studies, as cited by Lee (2003). VanPatten (1990) compared Andersen's data from "Anthony," a teenager who was learning Spanish by speaking with children in his neighborhood (non-instructed context), to data from "David," a monolingual twenty-year-old studying in college (classroom context). In his comparison of the linguistic context in which Anthony and David used clitics, VanPatten noted more differences than similarities (e.g., Andersen's Anthony produced $O_{cl}$VS patterns, never used the wrong person, and used all forms of the clitic but *te*, whereas David did not). One of VanPatten's (1990) main hypotheses regarding the acquisition of clitics in instructed and non-instructive contexts was that differences in performance patterns between classroom and naturalistic learners are not due to a change in underlying processes but due to *quantity and quality of input received*.

Likewise, Lee and Malovrh's (2009) and Malovrh and Lee's (2010) studies within the Input Processing (IP) framework suggested that the SA experience is one external

factor that can help explain nativelike attainment in $O_{cl}$VS structure aural comprehension, especially at the higher levels of profiency. The researchers made this assumption based on bio-data coming from background questionnaires; that is, immersion was not measured as independent variable. The suggestion that proficiency can be an attenuating factor in the extent to which learners take advantage of the favorable conditions provided by immersion contexts is also supported by Torres's (2003) dissertation. Her data consisted of conversational discourse (i.e., oral production) from fifteen L2 learners (L1 English, ten studying abroad and five taking classes in a classroom setting) over a one-semester period. She found that whereas the lower level learners in her study rarely used clitics, intermediate to mid-level learners begin with first person clitics followed by the addition of third person clitics in a more multifunctional manner. Lee and Malovrh (2009), Malovrh and Lee (2010) and Torres (2003) are the only studies in L2 Spanish that made potential connections between $O_{cl}$VS sentence interpretation and the SA abroad experience; however, none of them tested this hypothesis experimentally.

One question that the above studies open is: what are the potential learning gains derived from an implicit and intensive immersion environment for real-time $O_{cl}$VS sentence comprehension? As shown in Chapter Two, it seems that, even though emerging bilinguals continue to rely on word order as the most valid cue when interpreting $O_{cl}$VS sentence structures, contrastive agreement cues seem to help them reorganize the L2 cue hierarchy more promptly; therefore, interpretation errors diminish faster across experimental trials. In the previous chapter, analyses yielded an effect of Trial Order x Condition for the condition (Condition A) in which the clitic and the verb (the auxiliary *estar*, "to be") did not match in number (the clitic was singular and the verb plural.).

Whereas the Experiment 1in Chapter Two examined L2 sentence processing in relation to different matching/mismatching conditions when feedback on the target items was provided, the SA study in this chapter seeks to investigate the interaction between SA experience and L2 morphosyntactic development; that is, whether potential learning gains derived from an implicit and input-intensive immersive environment have an impact on real-time $O_{cl}$VS sentence comprehension in different agreement/disagreement conditions (RTs) as well as on $O_{cl}$VS accuracy rates. It also seeks to investigate whether such learning gains influence learners' sensitivity to subject-verb agreement violations in canonical sentences and their grammaticality judgments on these same items.

**Summary**

Empirical evidence concerning Study Abroad (SA) contexts on L2 development is divergent, but it is critical to take into account that many of the factors that shape Study Abroad (SA) programs themselves are also divergent (Grey et al., 2015). It may seem obvious that not all Study Abroad (SA) programs share the same characteritics (e.g., course quality, opportunities to interact with locals); therefore, it should be expected that programatic variability might also be reflected in the variability of learning outcome (e.g., Llanes, 2011); as a consequence, comparability across-programs and robust conclusions are diminished, again, due to the "aple and oranges" problem (Cf., Norris & Ortega, 2006).

This chapter's study further investigated morphological development in relation to the SA experience when the target items under investigation are $O_{cl}$VS structures, a difficult structure for acquisition that has important consequences for efficient communication: assigning who does what to whom in transitive sentences. Additionally,

as in Experiment 1, the present study investigated whether learners showed sensitivity to subject-verb agreement violations online, so as to confirm whether learners made use of agreement markers during real-time comprehension at the beginning and at the end of the program, in both canonical and non-canonical sentences alike. However, since subject-verb agreement is a computation often difficult to execute perfectly even in an L1 (Hoshino, Dussias, & Kroll, 2009), and the power analysis conducted in Chapter One revealed that power needed to be increased to obtained more conclusive results, two additional changes were included to better analyze the impact of immersion experience on learners' grammatical skills, a Grammaticality Judgment Task (GJT) was added to check whether learners were able to judge agreement computations between the subject and the verb, and whether this ability improved as a result of the immersion experience. As Hoshino et al. (2009) argued, off-line tasks such as GJTs are less cognitively demanding, and, therefore, whereas sensitivity to Subject-Verb agreement violations may not be detected by means of an online task, changes in outcomes, or acquired grammatical knowledge, according to Sagarra (2000), may be observed offline. According to Sagarra (2000, p. 130) GJTs seem to evaluate acquired knowledge. In addition, the number of tokens of canonical and non-canonical sentences was doubled ($k$=48).

Finally, this study intended to provide a good degree of outcome comparability with previous studies (Grey et. al., 2015), which is central for future secondary research (i.e., meta-analysis) (Norris & Ortega, 2000).

**Online and offline measurements in SA contexts**

As LaBrozzi (2009) has pointed out, in order to address real-time morphosyntactic processing, it is necessary to use an appropriate online technique for such purposes. Unlike offline grammaticality judgment tasks or comprehension tasks, which collect information on accuracy or response type, online psycholinguistic procedures used for sentence processing provide real-time information about the speed of response (Jiang, 2014; Ionin, 2013), and, therefore, about the areas that present more difficulty in processing (e.g., disambiguating regions are argued to be more cognitively taxing, Sagarra & Herschensohn, 2010). Whereas offline tasks tap into the state of linguistic knowledge (i.e., they are argued to be *product-oriented* tasks), online procedures (i.e., *process-oriented* tasks) tap into real-time language processing.

SLA scholars have pointed to the importance of combining both offline (e.g., grammaticality judgments) and online (e.g., self-paced reading, eye-tracking) tasks to assess both, grammar knowledge and the ability to implement it in real-time computation (Foucart, 2008; Juffs & Harrington, 1995, 1996; Juffs, 2004; Marinis, Roberts, Felser, & Clahsen, 2005; Sagarra & Herschensohn, 2010, 2011, 2012). By analyzing how accurate learners are in coming to the right form-to-function mappings of the target sentences, together with the time they spend (i.e., latency), L2 processing data allow the researcher to uncover processing pattern behaviors that underlie decision making. This study adheres to that position in the context of Study Abroad (SA) research.

To my knowledge, LaBrozzi's (2009) doctoral dissertation is currently the only study that has been published (LaBrozzi, 2012) which has employed an online procedure (eye-tracking) to compare a SA group vs. a non SA group in order to disentangle whether

the immersion experience impacted how English-speaking learners processed morphological cues (tense markers) in real-time. More specifically, the author was interested in investigating whether learners at the intermediate level of proficiency preferred morphological cues (tense markers) over lexical cues (adverbs) in assigning temporal reference in L2 Spanish.

LaBrozzi's (2009, 2012) findings partially supported the claim that immersion experience affected learners' processing preferences since the online data in this study showed that, even though SA learners began to rely more on morphological cues, they still continued to rely on lexical cues (i.e., adverbs) in order to assign temporal reference. LaBrozzi interpreted this to mean that learners were in a "transitional stage" (LaBrozzi, 2012, p. 238) towards a more native-like pattern. However, since the author did not adopt a pre-post-test design with the same group of learners *in situ,* and conducted the study only after participants had been back from their SA experience for approximately three months, the results of this study cannot discard a potentially larger effect on learners' processing patterns immediately after immersion. Even though it is certainly important to address the potential lasting effect of SA on sentence processing, to advance our understanding of whether the immersion experience has an impact on morphosyntactic development, it is necessary to measure grammatical gains in SA groups (within-subjects design) immediately after immersion as well. In addition, since LaBrozzi's study involved ungrammatical sentences (e.g., **Ayer el estudiante de música graba los discos en un estudio con sus amigos*, * "Yesterday, the music student records the records with her friends") and did not strictly measure learner's form-to-function mappings (many of the comprehension questions targeted a word previously seen in the target sentence, e.g.,

*¿Era un estudiante de química*?, "Was the student a chemistry student?"), future research needs to address, in terms of both accuracy and latency, how processing behaviors relate to learners' form-to-function mappings, as is the case when learners are asked to interpret $O_{cl}$VS sentences in Spanish.

Since self-paced reading (SPR) provides a complete measure of processing across the entire sentence, including immediate disruptions, spillover effects, and wrap-up effects on the final words of sentences (Just & Carpenter, 1980), it therefore allows investigators to respond to a variety of research questions targeting morphosyntactic phenomena at the sentential level. It can easily be administered with the use of a personal computer and open-source software (e.g., Linger). For these reasons, self-paced reading (SPR) seems to be a suitable technique to collect online data.

**Research Questions (RQs)**

**In relation to $O_{cl}$VS sentences:**

RQ1. Does a 5-week L2 intensive language experience abroad lead to higher accuracy in $O_{cl}$VS sentence interpretation overall?

If so,

RQ2: Do accuracy rates differ in relation to different matching/mismatching conditions between the clitic and the verb?

RQ3: Does a 5-week L2 intensive language experience abroad lead to lower processing costs (RTs) overall during real time $O_{cl}$VS sentence processing?

If so,

RQ4: Do processing costs (RTs) differ in relation to different matching/mismatching conditions between the clitic and the verb?

**In relation to SV sentences**

RQ5: Does a 5-week L2 intensive language experience lead learners to higher accuracy in grammaticality judgments tasks (GJTs) on SV sentences conveying subject-verb agreement violations (ungrammatical sentences)?

RQ6: Does a 5-week L2 intensive language experience abroad enhance sensitivity to subject-verb agreement violations (ungrammatical sentences), as measured by the Reaction Times (RTs) in the verb region?

**Predictions (Ps)**

**In relation to $O_{cl}$VS sentences**

P1: Learners will show higher accuracy rates as result of the intensive SA experience.

P2: Based on Experiment 1 results, contrastive agreement cues between the clitic and the verb seem to be more informative for the assignment of semantic functions to NPs in transitive sentences. Therefore, results on the pre-test will likely show a difference between matching and mismatching conditions with higher scores in the mismatching conditions. On the other hand, it is likely that in the post-test, learners will show a more generalized improvement, since, based on the CM, frequency of the input is critical for cue weight alteration to reflect actual L2 cue validity (McDonald, 1987, 1989). And such alteration takes place gradually. The immersion context, in which target structures are highly available in the input, should allow learners to reset their cue weights, and the new cue weights should carry over to all conditions, including the matching conditions, by

the post-test. Therefore, in the post-test results at Week 5, we would expect a general improvement in all agreement conditions.

P3: Based on Grey et al., (2015), advanced learners will exhibit lower online processing costs (lower RTs) overall as result of SA intensive experience.

P4: If processing costs for all sentences decreases, mismatching conditions between the clitic and the verb will exhibit similar online costs to those of the matching conditions (i.e., a more native-like processing pattern).

**In relation to SV sentences**

P5: Based on previous literature, advanced learners will be better at correctly rejecting sentences with subject-verb agreement violations in the GJTs as a result of the SA experience.

P6: After 5 weeks of SA, advanced learners will show higher Reading Times (RTs) in the verb regions where the SV sentences are ungrammatical.

## Methods

### Description of participants and inclusion criteria

22 advanced learners of Spanish participated in this study (*Mean age* = 20.72 years; 4 male; 18 female). Since two participants were native speakers of a language other than English (L1 Chinese and L1 Korean), they were dropped from the original sample; only data from the remaining L1 English speakers (N=20) were submitted for statistical analyses. Participants had received an average of 6 years of instruction in Spanish classroom settings prior to participating in the Barcelona summer abroad program. In order to qualify for this program, learners had to have taken at least two advanced (3rd year) Spanish language classes and have a GPA of 2.7 or higher. Only one

participant had previously participated in a summer abroad program, one year prior to the Barcelona summer program: for all other participants this was their first study abroad experience. Since participants were all advanced learners of Spanish who have completed different advanced courses (e.g., some had completed intensive courses, other non-intensive courses), all participants completed an Elicited Imitation Task (EIT) (see Appendix D, stimuli and coding criteria), (Ortega, Iwashita, Norris, & Rabie, 2002) in an effort to employ more robust proficiency assessment standards and to confirm that learners had homogeneous proficiency (Tremblay, 2011). EITs have been argued to tap into the ability of learners to automatically use core language knowledge, in particular grammar, vocabulary, and phonology, and they are able to assess this ability within a short period of time (Mozgalina, 2015).

Two experienced raters scored learners' performance on the EIT, which consisted of 30-items, and for which the maximum score possible was 120. The two raters gave the same ratings, on average, to 28 out of the 30 items. Disagreements on scoring were resolved by consensus to achieve 100% agreement, and these ratings were used as the final ratings. The minimum scored obtained on the EIT was 60 and the maximum scored obtained was 114 (*Mean*= 91.55; *SD*=15.42; *Skewness*= -0.81; *Kurtosis*=-0.24); the value of skewness and kurtosis, which were within the normal range of +/- 2, and the non-significance of the Kolmogorov- Smirnov test ($D(20)= 1.87$, $p=0.07$) indicate that the means are normally distributed. In addition, the data presented no outliers: the scores of all learners fell into ± 2 SD from the mean. No participants were excluded because of this criterion. Leaners also completed a language questionnaire based on Li, Zhang, Tsai, & Puls (2013) which addressed relevant language learning background variables (see

Appendix B). Two participants declared that they spoke another language: Hindi (*n*=1), and Bengali (*n*=1); and three participants declared that they had studied another foreign language: Japanese (*n*=1), and Portuguese (*n*=2). In addition to the above inclusion criteria, following similar criteria to those used in previous sentence processing research (e.g., Lago et al., 2015), only participants who scored 70% or higher on experimental filler responses for this study were included in the final analysis. This criterion guarantees assessing that all participants stayed on task during the self-paced reading (SPR). No participant was excluded for this criterion. Table 9 summarizes the inclusion criteria for participation in this SA study:

**Table 9: Inclusion criteria for participation in SA experiment**

| Variable | Inclusion criteria |
|---|---|
| (a) Native Language | English |
| (b) GPA | 2.7 or higher |
| (c) Previous coursework | Advanced II |
| (d) Accuracy score on fillers | 70 % or higher |
| (f) Elicited Imitation Task (EIT) | Within ± 2 SD from the mean |

**Characteristic of the SA program**

The Barcelona Summer Abroad Program at GU is a 5-week intensive study abroad program in which students complete 9 units of university coursework and participate in related cultural activities in Barcelona, Spain, and the surrounding regions. Spanish is the only language of communication on the Barcelona program. To participate, learners sign a Language Pledge in which they agree to use Spanish at all times.

Violations of the pledge may result in considerable grade reductions or even suspension from the program.

During the program, students enroll in 3 different content courses (9 credits) in the arts, history, business, linguistics, and/or politics; courses meet for 12 hours a week. All courses are taught entirely in Spanish by GU faculty or faculty from the host institution, Universitat Pompeu Fabra. These courses are conducted as they would be for the L1 speaker population, i.e., students do not receive explicit instruction in grammar or vocabulary as they would in a typical language course. Additionally, students participate in fieldwork activities organized by GU faculty and certified guides for 14 hours a week, and take part in organized conversations in Spanish with local university students on a weekly basis (2-6 hours/week). The home institution's attendance and participation policy is applied to fieldwork, conversation sessions, and coursework. The goals of this program, created in 2007, are: (a) to make L2 input constant, rich and authentic in order to maximize language learning and cultural awareness, (b) to offer learners the opportunity to interact with the local community and faculty members, so that they can be exposed to, and may actively engage in, a broad variety of communicative events (both formal and informal), (c) to promote the art, history and other identitarian expressions of Catalonia through the discussion of primary sources as well as the creation of networks between US students and the local members of the Catalonian community. To achieve these goals, students are encouraged to include examples of discussions with, interactions with, and research with the community into papers and class presentations. Work in the target language is intensive. Coursework/fieldwork begins at 10 am and finishes at 9 pm, four days a week, and lasts from 10 am to 4 pm on Fridays. Finally, to

ensure constant contact with Spanish, further steps are taken: (1) a teaching assistant is made available during the day in the residence hall; (2) one or two teaching assistants lives in the same residence as the students; (3) students do not have contact with other US students who are not part of the program; (4) cell phones are not allowed during classes or fieldwork activities.

**Research design**

The self-paced reading (SPR) conducted in this experiment consisted of two sub-experiments that included the target items ($O_{cl}$VS sentences and SV sentences) and filler sentences. The 96 sentence sets (48 SV and 48 $O_{cl}$VS) and 96 fillers were distributed across four lists in a Latin Square design (192 sentences total). Table 10 provides a sample of the target items.

**Table 10: Sample set of experimental items in the SA experiment**
**$O_{cl}$VS sentence conditions:**

**Contextual sentence for $O_{cl}$VS sentences:** e.g., *La camarera y los cocineros conversan en el comedor*. "The waitress and the cooks chat in the dining room").

**Mistmatch**, Cl-sg;.: *Allí la aburren los cocineros con historias familiares*
CONDITION A "There the cooks bore her with stories about her family."

**Match**, Cl-sg.;.: *Allí la aburre el cocinero con historias familiares*
CONDITION B "There the cook bores her with stories about her family"

**Match** Cl-pl..: *Allí las aburren los cocineros con historias familiares*
CONDITION C "There the cooks bore them with stories about her family"

**Mistmatch** Cl-pl..: *Allí las aburre el cocinero con historias familiares*
CONDITION D "There the cook bores them with stories about her family."

**True/ False questions (comprehension) for $O_{cl}$VS sentences**:
*Los cocineros aburren a la camarera*. "The cooks bore the waitress". TRUE
*Las camareras aburren a los cocineros.* "The waitress bores the cooks". FALSE
**SV sentence conditions:**

**Ungram,** subj. sg.: * *Por la tarde el perro duermen en la cama*
CONDITION A "*In the afternoon the dog (sing.) sleep (pl.) in the bed."

**Gram**, subj.sg.: *Por la tarde el perro duerme en la cama*
CONDITION B "In the afternoon the dog (sing) sleeps (sing) in the bed"

**Gram**, subj. pl.: *Por la tarde los perros duermen en la cama*
CONDITION C "In the afternoon the dogs (pl.) sleep (pl) in the bed"

**Ungram**, subj. pl.: * *Por la tarde los perros duerme en la cama*
CONDITION D "*In the afternoon the dogs (plural) sleeps (sing.) in the bed"

**T/F questions (comprehension) for SV sentences**
*El perro duerme en la cama*. "The dog sleeps in the bed". TRUE
*El perro duerme en el piso*. "The dog sleeps in the floor". FALSE
**Fillers**
*Los médicos examinan a la mujer de ochenta años* .
"The doctors examine the 80 year-old woman."
**T/F question for fillers (comprehension):**
*La mujer tiene ochenta años. False.* "The woman is eighty year old". TRUE
*La mujer tiene setenta y cinco años. False.* "The woman is seventy five year old". FALSE

As seen in Table 10, in this SA experiment, $O_{cl}$VS sentences conveyed a verb in the simple present instead of an auxiliary verb in the present continuous (auxiliary *estar* 'to be' + gerund) as in Experiment 1. The reason for this modification from Experiment 1 materials was to further investigate how learners processed agreement markers attached to frequent verbs in Spanish rather than on an auxiliary verb such *estar*,"to be." This also prevented the repetition of the same verb in all target sentences. In addition, using verbs in the simple present in all target items followed other previous SA studies which employed the present simple in their stimuli (e.g., Grey et al., 2015).

Importantly, the SA experiment in this chapter does not aim to compare the learning gains of two different contexts of learning (i.e., Study Abroad, (SA) versus at Home (AH) settings) since including an AH group as a control or non-treatment group, while it has certain benefits, also has limitations: first, by including an AH group, L2 learning contexts are indefectibly treated as a dichotomy (See Ortega & Byrnes, 2008, Ortega, 2013). Second, traditional between-subjects research designs comparing learning gains between students participating in intensive Study Abroad (SA) programs and

students participating in instructed studies at Home (AH) introduce potential confound variables to the study (e.g., quality of the input received, differences in motivation) since participants cannot be randomly assigned to one of the two groups, as in laboratory experiments (e.g., Sanz, 2014; Grey et al., 2015). Therefore, in an effort to increase the likelihood of reducing the potential of introducing another confound variable for the effects obtained, the present SA study opted for the limitation of not including a control group as a basis for comparison,.

Several steps were taken to make the pre-post test design less susceptible to practice effects: (a) in the post-test, participants were assigned the subsequent consecutive number to that which they had been assigned in the pretest (e.g., if a participant was number 1 in the pretest test, then she was number 2 in the posttest; if the next participant was number 2 in the pre-test, then his number was 3 in the post-test). This procedure was put into place to ensure that no participant read the experimental sentences ($O_{cl}$VS sentences, SV sentences and GJTs) in the same condition she/he read them in the pretest: for example, if a participant read a sentence such as: *Allí la aburre el cocinero con historias familiares*, "There the cook bores her with stories about her family" (Condition B), in the posttest she would read the sentence in any of the other three conditions, e.g., *Allí las aburren los cocineros con historias familiares*, "There the cooks bore them with stories about her family", but never in the same condition that she/he read it in the pre-test.; (b) in addition, the order in which participants read the target sentences was completely different in the post-test. For the GJTs, this procedure meant that, since two of the conditions were grammatical and two of the conditions were ungrammatical, participants may have read a sentence that was ungrammatical in the pre-

test (e.g., * *Por la tarde el perro duermen en la cama*, "*In the afternoon the dog (sing.) sleep (pl.) in the bed.") while the version they read of the same sentence in the post-test was grammatical (e.g., *Por la tarde el perro duerme en la cama* . "In the afternoon the dog (sing) sleeps (sing) in the bed"), or vice-versa. The final possibility was that participants may have read a sentence that was ungrammatical in the pre-test (e.g., * *Por la tarde los perros duerme en la cama*. "*In the afternoon the dogs (plural) sleeps (sing.) in the bed") and also ungrammatical in the posttest, but in a different condition (e.g., * *Por la tarde el perro duermen en la cama*. "*In the afternoon the dog (sing.) sleep (pl.) in the bed." ); (c) the fillers in the pre-test and the post-test (although similar in complexity) were different in each test; (d) a Pearson correlation was conducted to check whether any trial order effects observed in the pre-test strongly correlated with any difference observed in overall accuracy between the pre-test and post-test. The rationale was that, if learners improved across experimental trials in the pre-test, and that improvement explained higher accuracy rates in the post (i.e., training effects), then the improvement observed in the pre-test across experimental trials should correlate with the overall improvements observed in the post-test. The absence of such a correlation would suggest that the differences in accuracy between pre and post were therefore not due to training effects.

***$O_{cl}$VS sentences***

$O_{cl}$VS sentences consisted of 48 sentences arranged according to a 2 x 2 within-subjects design, with *mismatch* (match/mismatch) and *clitic number* (singular/plural) as factors and Reaction Times as the dependent variables. For the mismatch factor, number agreement between the direct object clitic pronoun and verb was manipulated: in the

matching conditions, the clitic and the verb were both either singular or plural (i.e., they agreed in number), while in the mismatching conditions the clitic was singular and the verb plural, or, on the contrary, the clitic was plural and the verb singular (i.e., they mismatched in number). The regions of interest for this experiment were the verb region and the spillover-regions, that is, the determiner region (verb +1 region) and the noun-subject region (verb+2 region). As in Experiment 1, $O_{cl}$VS sentence were preceded by a contextual phrase so that participants could establish a reference for the clitic. The true or false questions that followed all $O_{cl}$VS sentences assessed whether learners were able to interpret who was the agent (see Table 9). Whereas the T/F question pointed to the subject of the sentence, it avoided asking the question "who is doing the action" directly, as was the case in previous studies (Malorvh, 2006). Participants received 1 point for correct responses and 0 points for incorrect responses.

***SV sentences***

SV sentences were included in this study as a control, that is, to check whether leaners were able to integrate agreement morphology in real-time in canonical sentences. Previous research using SPR word-by-word with advanced learners (Foote, 2011) has shown that when SV sentences convey agreement violation between the subject and the verb, the verb region receives higher reading times, which provides evidence that advanced learners are sensitive to subject-verb agreement violations online. Based on the results of Experiment 1, the number of tokens of SV sentences in this study was doubled in order to increase power. In the current design, the SV consisted of 48 sentences arranged according to a 2 x 2 within-subjects design, with *grammaticality* (grammatical/ ungrammatical) and *subject number* (singular/plural) as factors and Reaction Times

(RTs) as the dependent variable. For the grammaticality factor, the agreement relationship was manipulated between the subject noun and the verb: in the grammatical conditions, the subject and the verb were both either singular or plural (i.e., they agreed in number), and in the ungrammatical conditions, either the subject was singular and the verb was plural, or, in the opposing case, the subject was plural and the verb singular (i.e., they mismatched in number), either of which rendered the sentences ungrammatical. Verbs were all conjugated in the simple present. As in Foote (2011), the regions of interest of this sub-experiment were the verb region and the spillover-regions. Since the spillover regions usually consisted of a prepositional phrase (e.g., *en la cama*, "on the bed" as in the example provided in Table 9), the spillover regions to be analyzed were usually a preposition (verb +1 region) and the determiner region following the preposition (verb+2 region). Following each SV sentence, a T/F question checked for comprehension. All T/F questions were related to facts conveyed in the sentence and did not point to the grammaticality of the sentence (see Table 9). Participants received 1 point for correct responses and 0 points for incorrect responses.

***Fillers***

The fillers in this experiment were different from those of Experiment 1, since in in the first experiment the fillers were usually shorter in length. Another change was that, in this experiment, fillers were more informative: for example, some of them were based on comments or blogs from native speakers of Spanish on the web. Unlike the T/F questions in Experiment 1, questions did not ask participants to recognize a word in the filler sentence (e.g., In this phrase the word 'dog' is mentioned); instead, the T/ F questions that followed each filler required learners, for example, to confirm certain

information given in the sentence, such as the place where the action took place, as in the question in Table 10. Participants received 1 point for correct responses and 0 points for incorrect responses.

***Gramaticality Judgment Tasks (GJTs)***

As discussed above, Grammaticality Jugdment Tasks (GJTs) (offline task) were included in the research design to address whether participants were also able to judge ungrammaticality offline and whether their judgment on GJTs improved after immersion. The GJTs consisted of the same SV sentences which were part of the SPR task.

**Materials**

The subjects of the contextual sentence for the $O_{cl}$VS sentences (in the example above: *La camarera y los cocineros conversan en el comedor*. "The waitress and the cooks chat in the dining room") always ended in *–a,* fem. or *–o,* masc.; or *–ora,* fem. or *–or*, masc. There were no items ending in *–ista*, or common nouns such as *mujer*, "woman" or *hombre*, "man" in these sentences, as in Experiment 1. That is, gender affixes were in accordance with the biological sex of their referent (i.e., they conveyed semantic gender, Corbett, 1988). The antecedents usually referred to professions: *programador*, "programmer", *doctora,*"doctor" *maestro*, "teacher", or nationalities, *argentino,* "Argentinian", *chino*, "Chinese." The subjects and direct object clitic pronouns in the target sentences were both animate, and were either in masculine or feminine, but they always had different genders. As argued by Sanz (1997, p. 54), counterbalancing the gender of the subject and the clitic direct object pronoun in the target sentence allows the researcher to know whether the learner is mapping the clitic as the subject or as the object. It also reduces ambiguity in the interpretation. Direct object

clitic pronouns referred to one of the subjects in the contextual sentences (e.g., in *La camarera y los cocineros conversan en el comedor*. "The waitress and the cooks chat in the corridor", the clitic *la*, "her," in the target sentence refers to the first subject (NP1) of the contextual sentence). The gender of the subjects in the contextual sentence was counterbalanced (i.e., masc.-fem; fem-masc.) to guarantee that the clitic and the gender of the subject in the target sentences were also counterbalanced. Whether the clitic referred to the first subject (NP1) or the second subject (NP2) was also counterbalanced. Likewise, an adverb (e.g., *a veces*, "sometimes") began the $O_{cl}$VS sentence to avoid placing the clitic in the position of the first word of the sentence since there is a tendency for readers to skip over the first word of a sentence (e.g., Rayner, 1998).

The verbs in SV sentences consisted of frequent intransitive verbs (e.g., *correr,* "run") but not of copulas (*ser*, "to be"), as in two experiments Foote (2011) conducted.

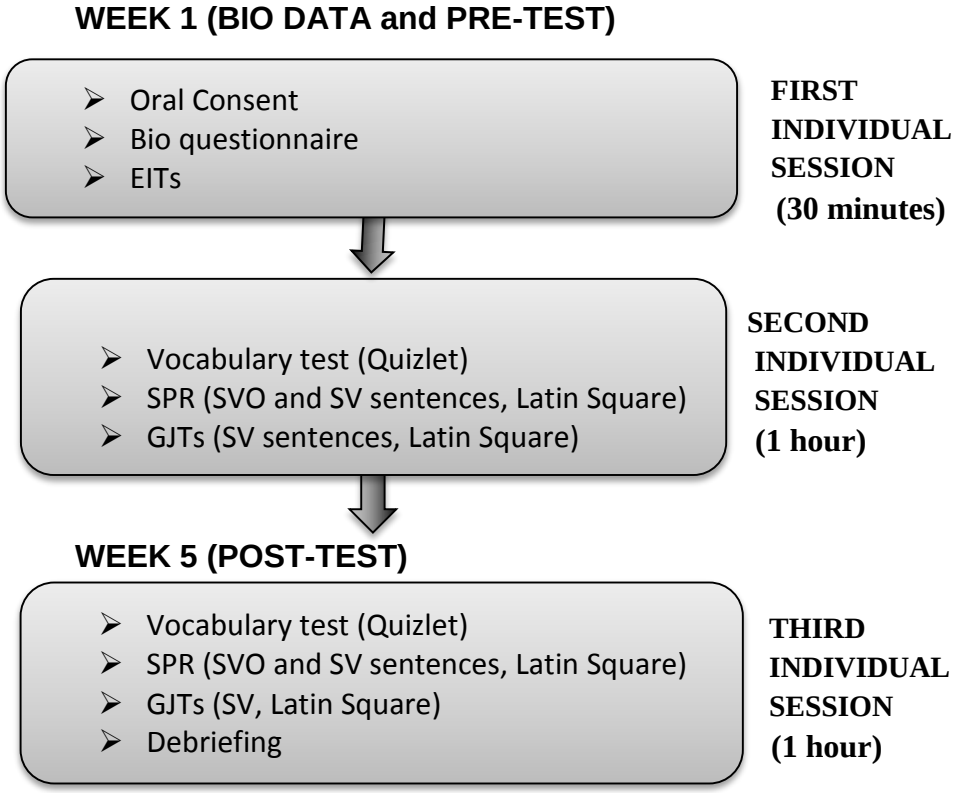
Finally, fillers consisted of SVO sentences using animate or inanimate objects. Sentences were in simple present (as in the target $O_{cl}$VS items). Proper names were also included (e.g., *Gabriel*, *Sandra*). The T/F questions that followed each filler pointed to: (a) the subject of the sentence (as in the $O_{cl}$VS sentence True/False questions), (b) the place where the event took place (*por la casa*, "in the house"), or (c) a fact conveyed in the sentence. Two sets of fillers were created, one for the pretest and one for the posttest. This was done so as to maintain learner interest and also to avoid inducing familiarity with the experimental task. Both sets of fillers consisted of SVO sentences that were comparable in complexity; that is, only small changes were made.

**Procedures**

Participants were invited to voluntarily participate in this study in exchange for extra credit on the Barcelona program courses. The study consisted of three sessions that were scheduled with the researcher on an individual basis. The first session and the second session took place within the first week of the students' arrival in Barcelona (i.e., Week 1), and the third session took place in the last week of their stay (Week 5). The first session lasted approximately 30 minutes. In this first session, students' gave their oral consent and then completed the bio questionnaire and the EIT. Within a framework of one or two days afterwards, students then met with the researcher for the second session, which lasted approximately 1 hour. In the second session, students completed: (i) a quiz consisting of vocabulary words from the target sentences and fillers: mainly on the professions. This vocabulary quiz ($k$= 36) was administered with Quizlet, a free online software that creates reversible electronic vocabulary flashcards in Spanish and in English (see Appendices E and F). Participants were instructed to look at the word in Spanish and translate the word into English; if they did not know the word, they were instructed to flip the card and read it in English; they could also ask the researcher questions about the words they still did not understand, if needed. After the vocabulary quiz, the participants then proceeded with the SPR and, finally, the GJT (pre-test). After 5 week of immersion, students then met with the researcher for the third and final session. In this third session, which lasted approximately 1 hour, student completed the vocabulary quiz again, which contained the same vocabulary words from the target items as well as the new vocabulary words from the second set of fillers. At the end of the

vocabulary quiz, participants completed the SPR task, the GJT and were then debriefed at the end of the session. Figure 7 describes the three individual sessions.

**Figure 7: Description of the three experimental sessions (pre and post-test)**




Participants completed all three sessions in a silent room using the researcher's PC. The SPR word-by-word (Just, Carpenter, & Woolley, 1982) was administered using Linger (Doug Rohde, MIT), a reading time software that automatically manages item blocking and randomization within lists. The order of experimental and filler items was randomized for each participant, and none of the experimental sentences appeared consecutively. In each experimental trial, participants read a sentence with words masked by dashes. When participants pressed the space bar, a word was revealed and the previous word was re-masked (following the same procedure discussed in Chapter One).

Six practice sentences (e.g., *Esta mañana José y Alexandra se han levantado de*

*muy mal humor [contextual sentence). Por eso Alexandra trata mal a José por una tontería. José trata mal a Alexandra por una tontería. False*) were presented before the experiment to ensure that learners understood the task. None of the practice sentence consisted of $O_{cl}$VS sentences, nor were participants informed that SV sentences had grammatical errors.

The contextual sentences that preceded the target $O_{cl}$VS items (e.g., *La camarera y los cocineros conversan en el comedor*. "The waitress and the cooks chat in the dining room") were also masked, except that, unlike the target items, participants were able to read the contextual sentence all at once since the goal of the SPR was to measure the RTs of the regions of interest in the target sentences and not in the contextual sentence. This contextual sentence disappeared after participants pressed the space bar so as to avoid participants' regressing to the contextual sentence; the experiment sought to measure the RTs on each of the words of the target sentences, amongst which were the region of interest (e.g., verb region). After the contextual sentence disappeared the target $O_{cl}$VS sentences were revealed word-by-word. Participants were instructed to read at a natural pace and answer the questions as quickly and accurately as possible by pressing one of two keys (the "F" key for True or the "J" key for False) (the selection of these keys followed previous work in psycholinguistics, e.g., Wagers, Lau, & Phillips, 2009). Between each block, participants were given three breaks; they were also informed that a break could be taken at any point of the experiment.

After the participants completed the SPR task, the GJTs were then administered, also using Linger. Participants read the same SV sentences they had read before in the SPR (in the same condition), although in a different random order. Participants were

provided with two sentences as examples to guarantee that they understood the task. None of the two sentences comprised a violation of subject-verb agreement (e.g., **La chica tiene ojos-masc muy bonitas-fem.* "The girl has beautiful - fem. eyes - masc."). During this task, participants were instructed to read the sentences and, after the sentence was displayed, to judge whether the sentence was acceptable or unacceptable in Spanish as quickly as possible by pressing the same keys they pressed in the SPR task. This was done so as not to arbitrarily change the procedure: this time, the "F" key meant the sentence was Acceptable and the "J" key was for Unacceptable. The GJTs generated one score: accuracy on identifying sentences as acceptable or unnaceptable.

Feedback was not provided for any of the experimental items in the SPR ($O_{cl}$VS, SV and Fillers) nor was it provided in the GJTs. This is one principal difference between this study and Experiment 1. Since the SA context is considered to be an implicit environment and this experiment was not a training experiment, provision of feedback inhibited the testing of potential learning gains derived from the implicit SA experience.

**Analyses**

**Accuracy data**

In the $O_{cl}$VS sentence experiment, the predictor variables for Accuracy were Condition and Trial order and Pre-Post (Time); and the dependent variable was accuracy scores. When Conditions are compared to one another in the agreement manipulation between the clitic and the verb, the condition in which the clitic and verb were both singular (Condition B) (i.e., the intercept) was considered the baseline.

**Latency data**

Following the methodological procedure of recent studies conducted in the field of psycholinguistics (Hofmeister, 2011; Lago et al., 2015), only residual reaction times (i.e., residual RTs) were entered into statistical analysis. Residual RTs were calculated in the following way: a linear model used in R first predicted the expected RTs for each word in function of its length. This model was a linear regression based on the following formula: $y=a+bx$, where *y* stands for RTs, *a* stands for the intercept, *b* stands for coefficient (slope), and *x* stands for word length (number of characters in a word) (cf. Hofmeister, 2011). Second, the linear model computed the residual RTs by subtracting (i.e., "regressing") the expected RTs that the model predicted from the raw (observed) RT data obtained.). Residual RTs were statistically estimated from the entire dataset (target items and fillers) for all participants. This was done in order to control for differences in length between the plural and singular forms of the clitic and the verb in the different experimental conditions: in Spanish, the suffix in the clitic in the 3rd person plural (-*s*) is one character longer than its counterpart in the singular; and the suffix in the verb in the 3rd person plural for the present tense (-*n*) is also one character longer than the 3rd person singular. Residual RT averages more than ± 2.5 SD from the participant mean by region and condition were excluded (Ratcliff, 1993).

**Statistical procedures**

The statistical analysis for accuracy data, the correct responses for the T/F comprehension questions targeting the agent for the $O_{cl}$VS sentences, was carried out using a binomial logistic regression in R (R Development Core Team, 2014). Two independent logistic regressions were conducted on the accuracy data (dependent

variable) of the pre-test and post-test, with Condition and Trial Order as fixed factors (independent variables). In addition, a logistic regression was conducted on the aggregated accuracy data of both tests with Condition, Trial Order, and Pre vs. Post as fixed factors. The analysis for Condition was carried out first for Condition B (the baseline); the other conditions were then analyzed as to how those other conditions differed from Condition B.

The latency data were analyzed using the *lmer* package in R. A series of general linear models (GLMs) were run on the residual RTs in each of the sentence regions (words) in order to calculate main effects and potential interactions. On the residual RT data of the pre-test and the post-test, Condition and Trial Order were the fixed factors of the GLMs conducted on each test independently. On the aggregated residual RT data, Condition, Trial Order and Pre vs. Post were the fixed factors of the GLMs conducted on both tests. RT averages more than 2.5 standard deviations from the participant mean by region and condition were excluded (Ratcliff, 1993).

## Results

### $O_{cl}$VS sentences

#### *Accuracy in $O_{cl}$VS sentences*

Table 11 shows the mean accuracy overall on the T/F questions in the pre-test and post-test in each of the $O_{cl}$VS experimental conditions. The T/F questions measured whether learners could identify the agent of $O_{cl}$VS sentences; therefore, they indicate the percentage of learners' correct form and meaning mappings for $O_{cl}$VS sentences.

**Table 11: Overall accuracy for $O_{cl}$VS sentences in the pre-test and post-test**

| | Condition | Mean | *SD* | Lower CI | Upper CI |
|---|---|---|---|---|---|
| Pre | A | 0.43 | 0.04 | 0.36 | 0.51 |
| **Post** | A | **0.59** | 0.03 | 0.52 | 0.66 |
| Pre | B | 0.42 | 0.04 | 0.35 | 0.49 |
| **Post** | B | **0.56** | 0.03 | 0.49 | 0.63 |
| Pre | C | 0.35 | 0.04 | 0.28 | 0.43 |
| **Post** | C | **0.53** | 0.03 | 0.46 | 0.60 |
| Pre | D | 0.41 | 0.04 | 0.34 | 0.49 |
| **Post** | D | **0.53** | 0.03 | 0.46 | 0.60 |

As seen in Table 11, on the pre-test, overall accuracy for all conditions was below 50% (Min: 35 % for Condition C, cl-pl., verb-pl.; Max: 43 % for Condition A, cl-sing, verb-pl.). In the post-test all conditions showed overall improvement, even though the improved score was right at or below 60% accuracy (Min: 53% for both Condition C (cl-pl., verb-pl.) and Condition D (cl-sing, verb-sing); Max: 59% for Condition A (cl-sing, verb-pl.)). As in Experiment 1, the overall accuracy results support the idea that $O_{cl}$VS sentences are still difficult to process for emerging bilinguals. Additionally, both Condition A (cl. sing., verb pl.) and Condition B (cl. sing., verb sing.), were the two conditions that exhibited higher accuracy scores in the pre-test and the post-test, occupying first and second place positions respectively.

There are two key differences between the designs of the present SA experiment and Experiment 1. First, participants in the SA study did not receive feedback on their responses; therefore, the only information to guide them in their progress was the wealth of input available in the immersion context. Even though changes in design do not allow for a direct comparison, a look at the results from both studies tells us that: in Experiment 1, in which participants completed the SPR once, they scored between 56% and 61% in

overall accuracy, a higher percentage compared to the results obtained in the pre-test of this SA study, which did not provide feedback on accuracy responses. Furthermore, the scores from the group receiving feedback were comparable in range to those from the post-test in this SA study. This suggests that (a) feedback and/or verb type may have influenced learner responses in Experiment 1, and (b) SA experience and feedback seem to be beneficial factors for both improvement on and correct interpretation of $O_{cl}VS$ sentences by English-speaking advanced learners. In other words, while the results obtained here apply to sentences conveying frequent verbs in the simple present (*mirar*: "look"; *ayudar*, "help"), instead of the auxiliary *estar*, "to be," which means that learners were exposed to a more varied type of $O_{cl}VS$ sentences, the common denominator between these two studies is that the mismatching plural verb condition (i.e., Condition A, cl-sing- verb- pl.) continued to elicit the highest score in overall accuracy (*mean* = 43% in the pre-test; *mean*=59% in the post-test) in comparison to the overall accuracy in Condition B (the baseline). Also, Condition B, (the baseline) (cl., sing, verb sing.), was the condition that followed Condition A in *overall* accuracy both in Experiment 1and this SA study.

In order to check whether the differences between the accuracy scores obtained in the pre-test (Week 1) and the post-test (Week 5) were significant, as well as to check for possible interactions between (a) pre and post (Time), (b) agreement in experimental conditions, and (c) trial order in the experiment (i.e., Trial Order), a logistic regression with Condition, Trial order and Pre-Post as predictors was run. In addition, the coefficients in the logistic regression were standardized in order to allow for direct comparison between this experiment and other similar experiments, regardless of the

number of experimental tokens. The results of the logistic regression for $O_{cl}$VS sentence are presented in Table 12.

**Table 12: Logistic regression for $O_{cl}$VS sentences**

| | $\hat{\beta}$ | Std. Error | z value | Pr(>\|z\|) |
|---|---|---|---|---|
| **Condition B** | -0.0461 | 0.1035 | -0.45 | 0.6557 |
| **Trial Order Std** | 0.1014 | 0.1046 | 0.97 | 0.3321 |
| **Condition A vs. Condition B** | 0.0663 | 0.1496 | 0.44 | 0.6575 |
| **Condition C vs. Condition B** | -0.2111 | 0.1487 | -1.42 | 0.1558 |
| **Condition D vs. Condition B** | -0.0891 | 0.1473 | -0.60 | 0.5453 |
| **Pre-test vs. Post-test** | 0.2909 | 0.1035 | 2.81 | **0.0049**** |
| **(Trial Order Std) × (Condition A vs. Condition B)** | 0.3206 | 0.1520 | 2.11 | **0.0349*** |
| **(Trial Order Std) × (Condition C vs. Condition B)** | 0.0684 | 0.1489 | 0.46 | 0.6457 |
| **(Trial Order Std) × (Condition D vs. Condition B)** | 0.2028 | 0.1485 | 1.37 | 0.1722 |
| **(Trail Order Std) × (Pre-Post vs. Post-test)** | -0.0173 | 0.1046 | -0.17 | 0.8687 |
| **(Condition A vs. Condition B) × (Pre-test vs. Post-test)** | 0.0968 | 0.1496 | 0.65 | 0.5176 |
| **(Condition C vs. Condition B) × (Pre-test vs. Post-test)** | 0.1018 | 0.1487 | 0.68 | 0.4934 |
| **(Condition D vs. Condition B) × (Pre-test vs. Post-test)** | -0.0383 | 0.1473 | -0.26 | 0.7949 |
| **(Trial Order Std) × (Condition A vs. Condition B) × (Pre-test vs. Post-test)** | -0.1677 | 0.1520 | -1.10 | 0.2700 |
| **(Trial Order Std) × (Condition C vs. Condition B) × (Pre-test vs. Post-test)** | -0.1091 | 0.1489 | -0.73 | 0.4636 |
| **(Trial Order Std) × (Condition D vs. Condition B) × (Pre-test vs. Post-test)** | -0.0624 | 0.1485 | -0.42 | 0.6743 |

*Signif. codes*: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

**Coefficients' operationalizations:**

- **Trial order:** accuracy improvement across experimental trial units (i.e., $O_{cl}$VS sentences).
- **Condition A vs. Condition B:** $O_{cl}$VS accuracy overall of Condition A against of the $O_{cl}$VS accuracy overall of condition B (the baseline).
- **Condition C vs. Condition B:** $O_{cl}$VS accuracy overall of Condition C against of the $O_{cl}$VS accuracy overall of condition B (the baseline).
- **Condition D vs. Condition B:** $O_{cl}$VS accuracy overall of Condition D against of the $O_{cl}$VS accuracy overall of condition B (the baseline).

- **Pre vs. Post:** $O_{cl}VS$ accuracy in the baseline condition (Condition B) between the Pre-test and the Post-test, with effects in the direction of post-test minus pre-test.
- **(Trial order) x (Conditions A vs. Condition B):** Experimental trial order effects in Condition A with respect to Condition B (the baseline) in the experiment *overall*.
- **(Trial order) x (Conditions C vs. Condition B):** Experimental trial order effects in Condition C with respect to Condition B (the baseline) in the experiment *overall*.
- **(Trial order) x (Conditions D vs. Condition B):** Experimental trial order effects in Condition D with respect to Condition B (the baseline) in the experiment *overall*.
- **(Trial order) x (Condition A vs. Condition B) x (Pre vs. Post):** Experimental trial order effects in Condition A with respect to Condition B, and in relation to the accuracy differences between the Pre-test and the Post-test.
- **Trial order) x (Condition C vs. Condition B) x (Pre vs. Post):** Experimental trial order effects in Condition C with respect to Condition B, and in relation to the accuracy differences between the Pre-test and the Post-test.
- **Trial order) x (Condition D vs. Condition B) x (Pre vs. Post):** Experimental trial order effects in Condition D with respect to Condition B, and in relation to the accuracy differences between the Pre-test and the Post-test.

As seen in Table 12, the logistic regression yielded a significant difference between the pre-test scores and post-test scores; that is, the overall accuracy for Condition B (the baseline) in the post-test was significantly higher ($\hat{\beta}$= 0.2909, z= 2.81, $p$= 0.0049). Importantly, the interaction coefficients in the rest of the conditions were not significant ((Condition A vs. Condition B) × (Pre-test vs. Post-test), $\hat{\beta}$= 0.0968, z= 0.65, $p$= 0.5176; (Condition C vs. Condition B) × (Pre-test vs. Post-test), $\hat{\beta}$= 0.1018, z= 0.68, $p$= 0.4934; (Condition D vs. Condition B) × (Pre-test vs. Post-test), $\hat{\beta}$= 0.0383, z= -0.26, $p$= 0.7949) indicating that learners significantly improved in all four conditions in the post-test.

Effect sizes calculations show that participants were over twice as likely to receive better scores on the post-test, as the odds of participants' being more accurate on the post-test than on the pre-test increased ($\hat{\beta}$=0.2909; *odds*=2.66)

Figure 8 represents learners' estimated accuracy across the experimental trials in the pre-test and post-test respectively.

**Figure 8: $O_{cl}VS$ accuracy over the course of the experiment in the pre-test and post-test**

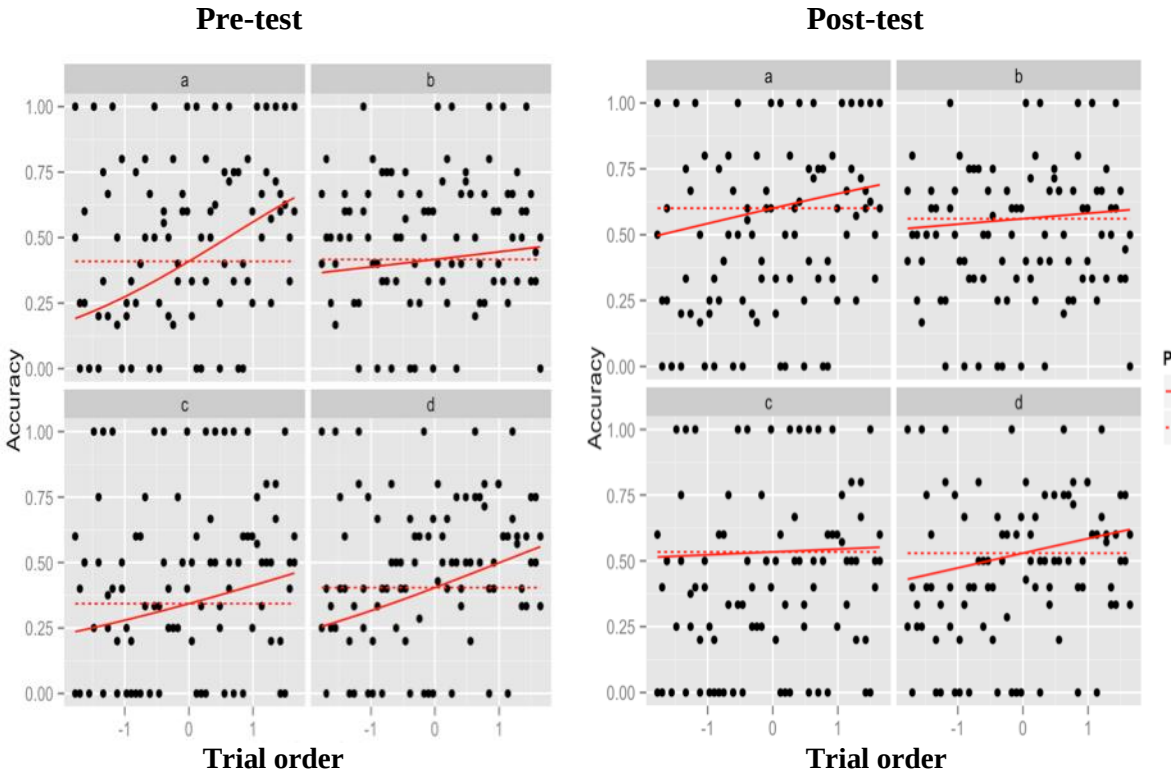



Figure 8 shows that not only did learners improve in the post-test, but, in the pre-test, learners seem to have improved faster in Condition A (cl.sing., verb pl.). To check whether there was an interaction between Condition and Trial Order in the pre-test, two general linear models were independently conducted on the accuracy data of the pre-test and post-test independently, as shown in Table 13 and Table 14:

**Table 13: Logistic regression for $O_{cl}VS$ sentences in the pre-test**

| | $\beta$ | St. E | z value | Pr(>\|t\|) |
|---|---|---|---|---|
| Condition B (Intercept) | 0.3370 | 0.1507 | -2.24 | 0.0253 |
| Trial order std. | 0.1187 | 0.1547 | 0.77 | 0.4429 |
| Condition A vs. Condition B | -0.0305 | 0.2206 | -0.14 | 0.8901 |
| Condition C vs. Condition B | -0.3129 | 0.2213 | -1.41 | 0.1574 |
| Condition D vs. Condition B | -0.0508 | 0.2168 | -0.23 | 0.8147 |

| | | | | |
|---|---|---|---|---|
| **Trial Order Std. × (Condition A vs. Condition B)** | **0.4883** | **0.2287** | **2.14** | **0.0327*** |
| Trial Order Std. × (Condition C vs. Condition B) | 0.1776 | 0.2237 | 0.79 | 0.4273 |
| Trial Order Std. × (Condition D vs. Condition B) | 0.2652 | 0.2218 | 1.20 | 0.2317 |

**Table 14: Logistic regression for $O_{cl}$VS sentences in the post-test**

| | *β* | St. E | z value | Pr(>\|t\|) |
|---|---|---|---|---|
| Condition B (Intercept) | 0.2447 | 0.1419 | 1.72 | 0.0846 |
| Trial order std. | 0.0841 | 0.1407 | 0.60 | 0.5498 |
| Condition A vs. Condition B | 0.1631 | 0.2020 | 0.81 | 0.4195 |
| Condition C vs. Condition B | 0.1092 | 0.1987 | -0.55 | 0.5824 |
| Condition D vs. Condition B | 0.1274 | 0.1995 | -0.64 | 0.5232 |
| Trial Order Std. × (Condition A vs. Condition B) | 0.1530 | 0.2004 | 0.76 | 0.4452 |
| Trial Order Std. × (Condition C vs. Condition B) | -0.0407 | 0.1965 | -0.21 | 0.8360 |
| Trial Order Std. × (Condition D vs. Condition B) | 0.1404 | 0.1977 | 0.71 | 0.4776 |

Table 13 confirms what Figure 8 shows visually: learners in the pre-test significantly improved across experimental trials in Condition A. The same interaction was found in Experiment 1, even though in that experiment, unlike in this SA study, learners received feedback on their responses. Table 14 shows that no interaction between Condition and Trial Order was found in the post-test.

Finally, to statistically verify that the pre-test versus post-test differences were not due to a task training effect after exposure to the SPR task in the pre-test (i.e., familiarity with the SPR task), a Pearson's correlation was run for each participant in order to check whether their rate of improvement across experimental trials in the pre-test (i.e., how

much they learned from the SPR in the pre-test) related to the observed learners' overall improvement in the post-test. Therefore, one variable of the correlation was the coefficient for order effect on accuracy in the pre-test (estimated without Condition as a predictor) and the other variable was the coefficient for the pre-post-test effect on overall accuracy. No correlation was found between these two variables ($r(16) = .49$, $p = -0.078$) suggesting that the amount of learning found over the course of the experiment in the pre-test did not seem to predict the learning gains found in the post-test, (i.e., results suggest a lack of test effects). In addition, even though learners showed a trial order effect on accuracy in the pre-test, the accuracy rate overall at Week 1 was below 50 %; therefore, it seems unlikely that the trial order effect found in the pre-test would have significantly influenced learners' performance at Week 5.

With regards to the degree to which the effects in Condition A, C, and D differed with respect to those in Condition B (cl-sing, verb-sing.; the baseline) in the experiment *overall* (i.e., taking into account pre-test and post-test accuracy scores as agreggated data), the interaction between Condition A × Trial order ($\hat{\beta} = 0.3206$, $z = 2.11$, $p = 0.0349$, *odds*=2.74) suggests that learners improved almost three times (2.74) faster in Condition A (cl-sing., verb-pl.) than they did in Condition B (cl-sing., verb-sing), the baseline, a much more ambigous condition in which agreement cues are not informative.

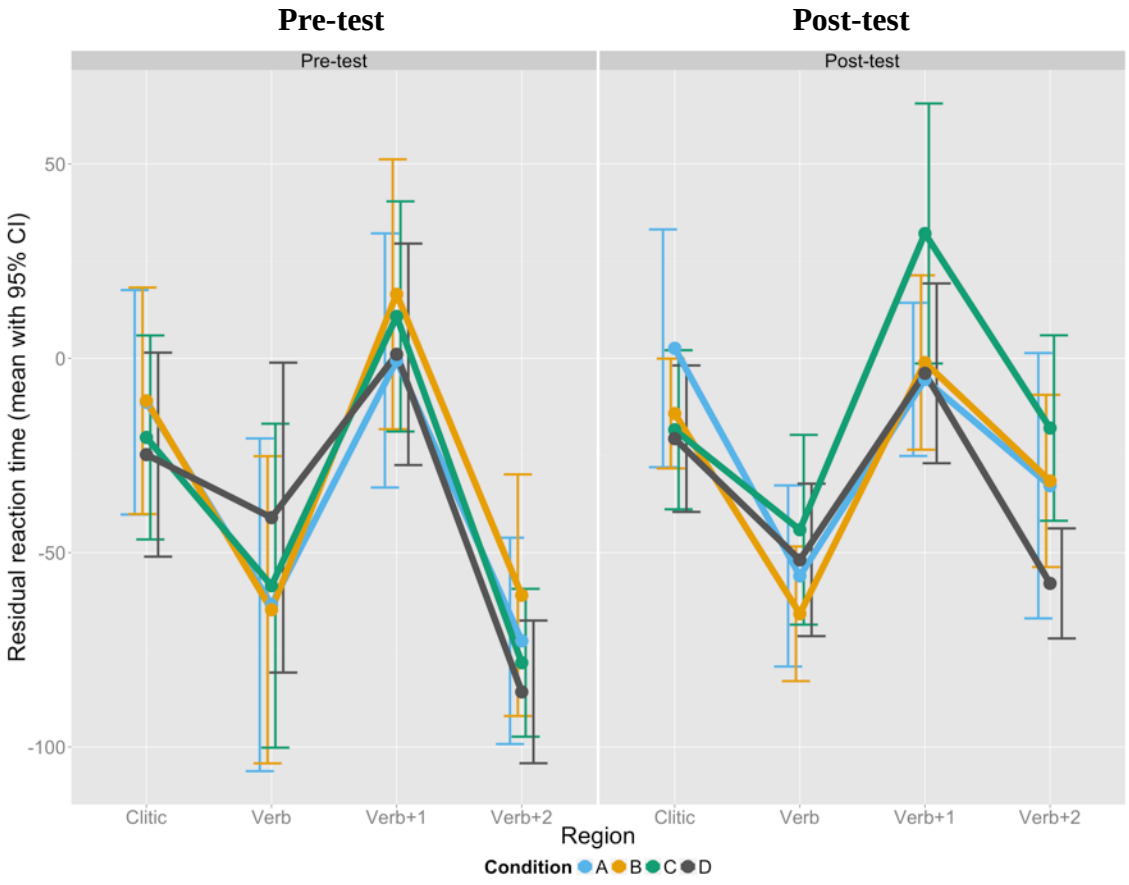
To summarize, in the post-test (Week 5), learners significantly improved in overall accuracy in $O_{cl}VS$ sentences in all conditions. Participants were over twice as likely to attain better scores at the end of the SA program. Lastly, when results in the pre-test and in the post-test were analyzed together, Condition A (cl. sing., verb pl.) emerged as the condition that yielded higher accuracy scores as the experiment progressed (i.e., an

order effect was found in this condition in the experiment overall). This result supports the findings that Experiment 1 yielded.

***Latency in $O_{cl}VS$ sentences***

Residual Reaction Time (RT) data was analyzed in this section in order to identify differences in $O_{cl}VS$ sentence processing costs at Week 1 (pre-test) and at Week 5 (post-test) of the program. Figure 9 illustrates the mean average in milliseconds (ms) for each of the regions (words) in the $O_{cl}VS$ sentences in all conditions.

**Figure 9: Mean residual RTs overall for all conditions in the pre-test and post-test**




Visually, Figure 9 shows that $O_{cl}VS$ sentences in the pre-test and post-test present a similar processing pattern, which resembles a V shape. Overall, participants spent more

time in the clitic region and in the spillover region (i.e., the determiner of the post-verbal subject, verb +1 region) than in the verb region. This contrasts with the pattern that was observed in Experiment 1, which resembled the opposite: a ∧ shape. Figure 9 illustrates the mean of the residual RTs in each of the regions of interest and conditions *overall*.

The residual RT data in each region of interest, clitic region, verb region and both spillover regions was submitted to further statistical analyses to detect any significant changes over the course of the SPR task and over time from Week 1 to Week 5 (pre-test and post-test). Four linear models were conducted independently for each of the regions of interest with Condition, Trial order and Pre-Post as fixed factors. Table 15 provides the results of the linear models conducted on each region of interest.

**Table 15: Linear models for the clitic region, the verb region, and spillover region (verb + 1) & (verb + 2)**

| | CLITIC REGION | | | | VERB REGION | | | |
|---|---|---|---|---|---|---|---|---|
| | $\widehat{\beta}$ | St. E | *t* value | Pr(>\|z\|) | $\widehat{\beta}$ | St. E | *t* value | Pr(>\|z\|) |
| **Condition B** | -12.65 | 8.93 | -1.42 | 0.1569 | -64.76 | 11.53 | -5.62 | **0.0000 ***** |
| **Trial Order Std** | -13.73 | 9.05 | -1.52 | 0.1297 | -41.45 | 11.67 | -3.55 | **0.0004 ***** |
| **Condition A vs. Condition B** | 8.27 | 12.63 | 0.66 | 0.5124 | 5.78 | 16.30 | 0.35 | 0.7229 |
| **Condition C vs. Condition B** | -7.02 | 12.63 | -0.56 | 0.5784 | 13.42 | 16.29 | 0.82 | 0.4099 |
| **Condition D vs. Condition B** | -9.94 | 12.64 | -0.79 | 0.4314 | 18.91 | 16.31 | 1.16 | 0.2462 |
| **Pre-test vs. Post-test** | -1.56 | 8.938 | -0.18 | 0.8610 | -0.96 | 11.53 | -0.08 | 0.9330 |
| **(Trial Order Std) × (Condition A vs. Condition B)** | 0.18 | 12.71 | 0.01 | 0.9886 | 6.11 | 16.37 | 0.37 | 0.7091 |
| **(Trial Order Std) × (Condition C vs. Condition B)** | 3.35 | 12.65 | 0.27 | 0.7906 | -3.95 | 16.31 | -0.24 | 0.8085 |
| **(Trial Order Std) × (Condition D vs. Condition B)** | -6.80 | 12.79 | -0.53 | 0.5947 | 17.37 | 16.50 | 1.05 | 0.2927 |
| **(Trail Order Std) × (Pre-test vs. Post-test)** | 12.27 | 9.05 | 1.36 | 0.1754 | 41.74 | 11.67 | 3.58 | **0.0004 ***** |
| **(Condition A vs. Condition B) ×(Pre-test vs. Post-test)** | 8.46 | 12.63 | 0.67 | 0.5027 | 3.95 | 16.30 | 0.24 | 0.8086 |
| **(Condition C vs. Condition B) ×(Pre-test vs. Post-test)** | 2.90 | 12.63 | 0.23 | 0.8184 | 8.13 | 16.29 | 0.50 | 0.6175 |
| **(Condition D vs. Condition B) ×(Pre-test vs. Post-test)** | 3.48 | 12.64 | 0.28 | 0.7826 | -5.05 | 16.31 | -0.31 | 0.7566 |
| **(Trial Order Std) × (Condition A vs. Condition B) × (Pre-test vs. Post-test)** | -18.60 | 12.71 | -1.46 | 0.1436 | -3.37 | 16.37 | -0.21 | 0.8369 |
| **(Trial Order Std) × (Condition C vs. Condition B) × (Pre-test vs. Post-test)** | 18.88 | 12.65 | 1.49 | 0.1357 | -6.87 | 16.31 | -0.42 | 0.6734 |
| **(Trial Order Std) × (Condition D vs. Condition B) × (Pre-test vs. Post-test)** | 9.77 | 12.79 | 0.76 | 0.44 | -18.89 | 16.50 | -1.14 | 0.2524 |

Continued. (Table 15)

| | SPILLOVER REGION (Verb +1) (determiner post-verbal subject) | | | | SPILLOVER REGION (Verb +2) (noun post-verbal subject) | | | |
|---|---|---|---|---|---|---|---|---|
| | $\widehat{\beta}$ | St. E | *t* value | Pr(>\|z\|) | $\widehat{\beta}$ | St. E | *t* value | Pr(>\|z\|) |
| **Condition B** | 7.74 | 10.26 | 0.75 | 0.4508 | -38.01 | 8.9625 | -4.24 | **0.0000 *** ** |
| **Trial Order Std** | -20.28 | 10.40 | -1.95 | **0.0513*** | -17.47 | 9.0694 | -1.93 | **0.0541** |
| **Condition A vs. Condition B** | -10.67 | 14.50 | -0.74 | 0.4620 | -11.25 | 12.6747 | -0.89 | 0.3746 |
| **Condition C vs. Condition B** | 13.32 | 14.50 | 0.92 | 0.3584 | -3.30 | 12.6747 | -0.26 | 0.7944 |
| **Condition D vs. Condition B** | -9.14 | 14.51 | -0.63 | 0.5287 | -30.80 | 12.6747 | -2.43 | **0.0152 *** |
| **Pre-test vs. Post-test** | -8.84 | 10.26 | -0.86 | 0.3892 | 8.26 | 8.9625 | 0.92 | 0.3564 |
| **(Trial Order Std) × (Condition A vs. Condition B)** | 11.67 | 14.59 | 0.80 | 0.4238 | -11.69 | 12.7204 | -0.92 | 0.3579 |
| **(Trial Order Std) × (Condition C vs. Condition B)** | -4.38 | 14.52 | -0.30 | 0.7628 | 0.65 | 12.7124 | 0.05 | 0.9589 |
| **(Trial Order Std) × (Condition D vs. Condition B)** | 16.287 | 14.68 | 1.11 | 0.2676 | 8.89 | 12.8312 | 0.69 | 0.4882 |
| **(Trail Order Std) × (Pre-test vs. Post-test)** | 14.194 | 10.40 | 1.36 | 0.1725 | 22.93 | 9.0694 | 2.58 | **0.0100 *** |
| **(Condition A vs. Condition B) ×(Pre-test vs. Post-test)** | 6.35 | 14.50 | 0.44 | 0.6615 | 6.98 | 12.6747 | 0.55 | 0.5816 |
| **(Condition C vs. Condition B) ×(Pre-test vs. Post-test)** | 19.87 | 14.50 | 1.37 | 0.1709 | 21.55 | 12.6747 | 1.70 | 0.0891 |
| **(Condition D vs. Condition B) ×(Pre-test vs. Post-test)** | 6.37 | 14.51 | 0.44 | 0.6605 | 3.46 | 12.6747 | 0.27 | 0.7847 |
| **(Trial Order Std) × (Condition A vs. Condition B) × (Pre-test vs. Post-test)** | -1.73 | 14.59 | -0.12 | 0.9056 | -14.76 | 12.7204 | -1.16 | 0.2459 |
| **(Trial Order Std) × (Condition C vs. Condition B) × (Pre-test vs. Post-test)** | 7.05 | 14.52 | 0.49 | 0.6274 | -11.35 | 12.7124 | -0.89 | 0.3720 |
| **(Trial Order Std) × (Condition D vs. Condition B) × (Pre-test vs. Post-test)** | -9.54 | 14.68 | -0.65 | 0.5158 | -17.25 | 12.8312 | -1.34 | 0.1788 |

***Processing costs in $O_{cl}$VS sentences in the experiment overall***

Table 15 shows that for the experiment *overall* (i.e., taking into account the RTs in the pre-test and post-test) there is an effect of trial order, a speedup in Condition B (cl. sing., verb sing., the baseline); this speedup can be seen in the verb region and the spill over regions since all estimates are negative (*verb region*: $\hat{\beta}$= - 41.4593, *t*= - 3.55, *p*= 0.0004; determiner region: $\hat{\beta}$= - 20.2885, *t*= -1.95, *p*= 0.00513, post-verbal subject region: $\hat{\beta}$= - 17.4755, *t*= -1.93, *p*= 0.0541). This indicates that the more $O_{cl}$VS sentences learners read in Condition B the faster they read them across experimental trials.

On the other hand, in the verb region, Condition A (cl.sing., verb-pl.) and Condition D (cl. pl., verb-sing.), the two conditions that exhibited contrastive agreement, were read, overall, slightly slower than the baseline (Condition B) across experimental trials,. As seen in Table 15, the estimates for both Condition A and Condition D are positive, which indicates a slight slowdown (Trial order std × (Condition A vs. Condition B), $\hat{\beta}$= 6.11, *t*= 0.37, *p*= 0.2927; Trial order std × (Condition D vs. Condition B), $\hat{\beta}$=17.37, *t*=1.5, *p*= 0.2927). In regard to Condition C (cl. pl. verb-pl.), this last condition, in which agreement information was not useful, was read slightly faster than Condition B across experimental trials (Trial order std × (Condition C vs. Condition B), $\hat{\beta}$= -3.95, *t*= -0.24, *p*= 0.8085).

In summary, in the experiment overall, the tendency was that in the two conditions in which agreement information was not useful (i.e., $O_{cl}$VS sentences conveying non-contrastive agreement) learners read the verb region faster, and in the two conditions in which agreement information was useful (i.e., $O_{cl}$VS sentences conveying

contrastive agreement) learner read the verb region slightly more slowly. However, none of these differences in reading pace (Condition A, D and C in comparison with Condition B) were significant.

***Processing costs in $O_{cl}$VS sentences in the post-test***

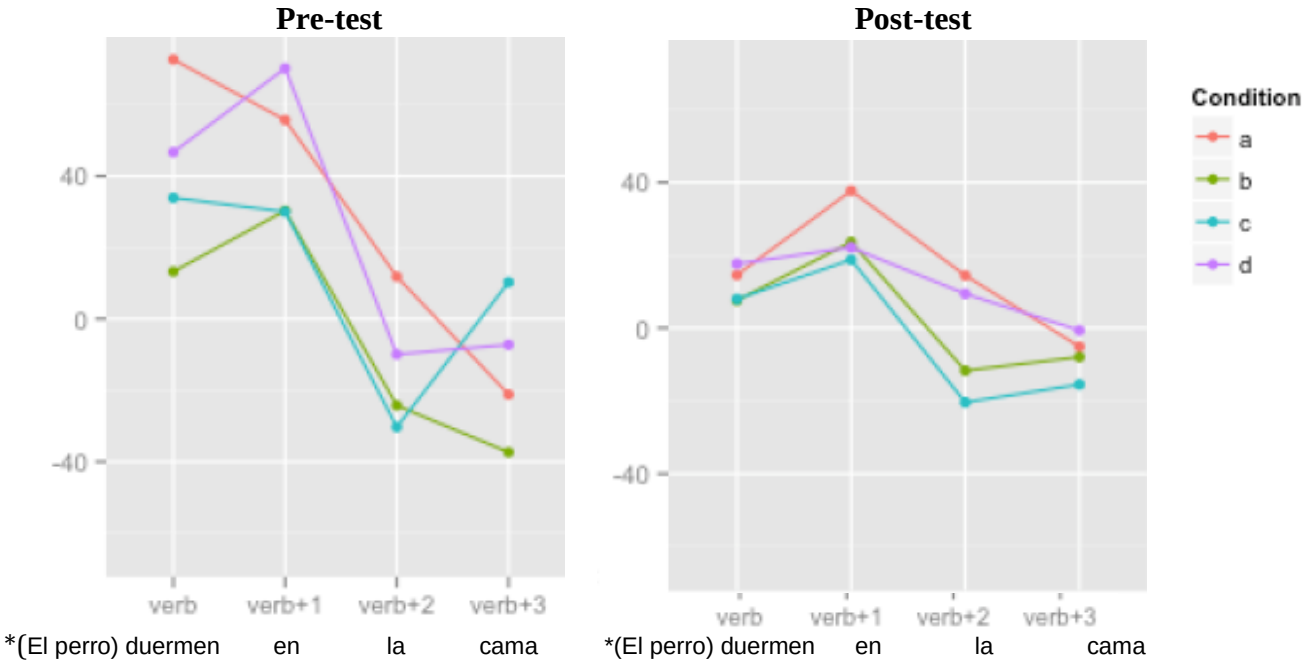
Table 15 shows that, regarding the processing costs of the $O_{cl}$VS sentences in the *post-test*, participants showed a slowdown across experimental trials in the verb region ($\hat{\beta}$= 41.7492, *z*= 3.58, *p*= 0.004) and especially in the noun of the *post-verbal subject* ($\hat{\beta}$= 22.9391, *z*= 2.58, *p*= 0.010) in Condition B (the baseline, cl. sing., verb pl.). The slowdown in these two specific regions, which are highly informative for disambiguating *who does what to whom* as they progressed through the experiment, suggests that learners' attention focused on these disambiguating cues. Lastly, Table 15 shows that participants tended to read $O_{cl}$VS sentences in Condition B (as whole units) at a slightly faster rate in Week 5, since the estimates in the clitic, verb and verb+1 regions are negative. This tendency was not significant.

**SV sentences**

***Latency in SV sentences***

Figure 10 illustrates the mean residual RTs in the verb region and spillover regions after the verb in the SV sentences (e.g., **El perro duermen en la cama*, "*The dog sleep on the bed"), the sub-experiment that was included in the SPR task to test sensitivity to subject-verb agreement violations online, similar to that of Experiment 1. The major difference between this sub-experiment and the one conducted in Experiment 1 was that participants in this study did not receive feedback on the T/F comprehension question as the participants in Experiment 1. Also the number of tokens was doubled *(k*=48) in relation to Experiment 1.

**Figure 10 : RT data for SV sentences**




As seen in Figure 10, there seems to be an effect of order for both the pre-test and the post-test (i.e., participants read sentences faster across experimental trials), although this effect seems to be less marked in the post-test. A linear model run on the residual RT data with Condition, Trail order and Pre and Post as fixed factors showed that, overall, the verb region and the spillover region (verb+1) in Condition B were read faster over the course of the experiment: these two results approached significance ($\hat{\beta}$= -32.81, $t$= -17.67, $p$=0.063 for Trial order in the verb region in Condition B; $\hat{\beta}$= -22.63, $t$= -7.52, $p$=0.079, for trial order in the verb +1 region in Condition B). These results suggest that as they progressed through the experiment learners skipped the verb region when reading an SV sentence whose subject in sentence-initial position was a full NP.

Regarding the ungrammatical sentences (Condition A and Condition D), although no significant differences were found in the verb region either in the pre-test or in the

post-test, learners spent more time in the verb region of the two ungrammatical conditions. In the pre-test, Condition A (**El perro duermen*, "*The dog sleep") received higher RTs in comparison to the other ungrammatical conditions, suggesting that this ungrammatical condition in particular might have been more 'disruptive' than its counterpart, condition D (* *Los perros duerme*, "The dogs sleeps"), which exhibited lower RTs in the verb region.

In the post-test, differences between grammatical and ungrammatical conditions were even smaller in the verb region, although the two ungrammatical conditions continued to receive slightly higher RTs in this region. Also, in the post-test, learners tended to spend more time in the spillover region following the verb (verb+1) in Condition A (**El perro duermen en la cama*, "*The dog sleep on the bed"). However, these differences were not significant.

As will be discussed in the next section on GJTs, concluding that learners somehow became 'less sensitive' to subject-verb agreement violations online in the post-test seems counterintuitive, considering that they significantly improved in their judgment of ungrammatical sentences offline. One explanation to account for the lack of significance in the verb region of ungrammatical SV sentences may be related to learners' tendency to progressively skip the verb region of the SV sentences (whose subjects consisted of full NPs in sentence-initial position); finally, the relatively small *N* size of this sample may also account for these results. As discussed earlier in respect to Experiment 1, in order to (dis)confirm the tendencies found here, this sub-experiment needs to be replicated with a larger group of learners, and other verbs such as copulas should also be included.

### Grammaticality Judgement Tasks (GJTs)

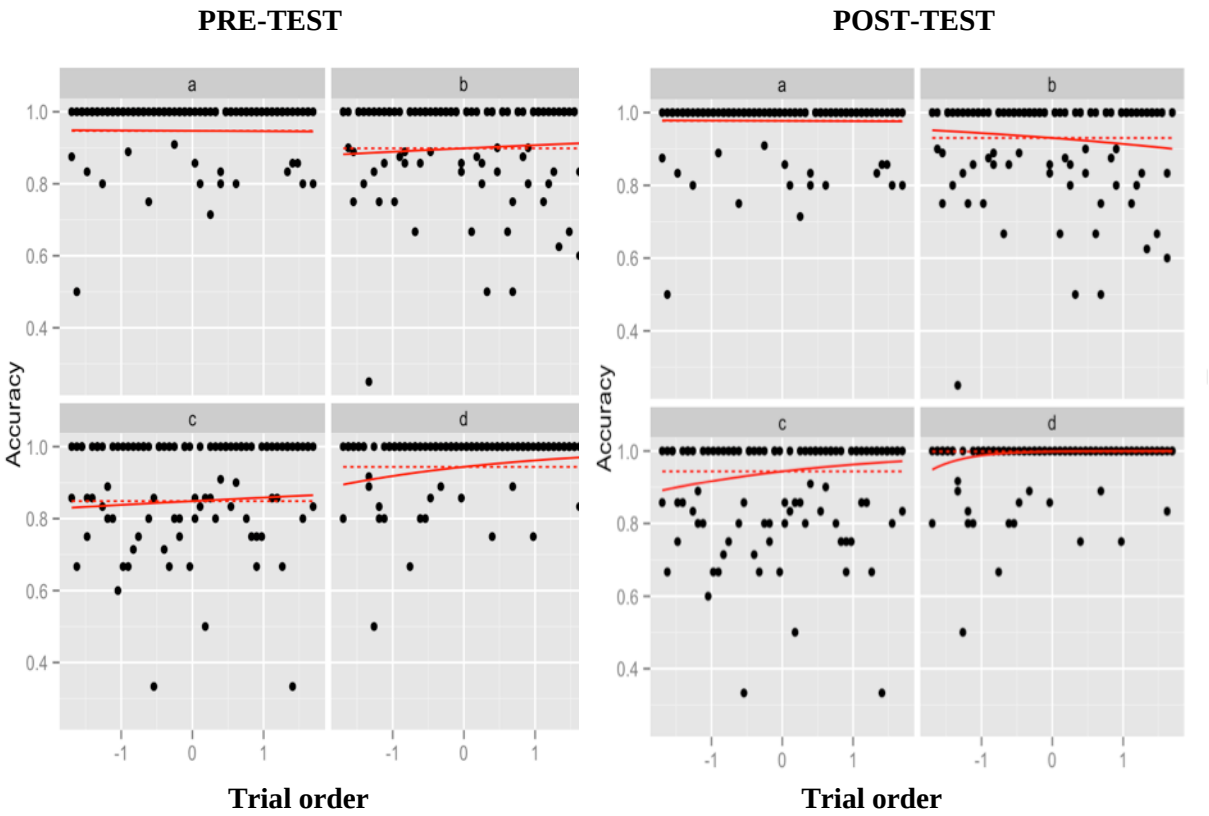
Since in Experiment 1, it was not possible to provide conclusive evidence that learners were sensitive to Subject-Verb agreement violations when processing canonical SV sentences on-line, due to a lack of enough power, in this study, an offline test was included to ensure that learners were able to use grammaticality knowledge offline before starting the SA program and, more importantly, to measure whether learners' ability to judge grammaticality improved as result of immersion. It is important to note here that even though the GJTs were untimed, learners were encouraged to decide, as soon as possible, whether the sentence was grammatical or ungrammatical. Since learners closely followed these instructions, participants may have accessed implicit knowledge when performing these GJTs.

Table 16 shows the mean accuracy for each of the GJT conditions (grammatical and ungrammatical) and Figure 11 visually represents how learners responded to GJTs across experimental trials for the same SV sentences they read in the SPR task. Both pre-test and post-test are compared.

**Table 16: Mean accuracy for Grammaticality Judgment Tasks (GJTs)**

| | Condition | Mean | *SD* | Lower CI | Upper CI |
|---|---|---|---|---|---|
| Pre | A | 0.95 | 0.01 | 0.91 | 0.97 |
| **Post** | **A** | **0.98** | **0.01** | **0.95** | **0.99** |
| Pre | B | 0.90 | 0.02 | 0.85 | 0.93 |
| **Post** | **B** | **0.93** | **0.02** | **0.89** | **0.96** |
| Pre | C | 0.85 | 0.02 | 0.80 | 0.89 |
| **Post** | **C** | **0.94** | **0.01** | **0.90** | **0.96** |
| Pre | D | 0.94 | 0.01 | 0.90 | 0.96 |
| **Post** | **D** | **0.99** | **0.01** | **0.97** | **1.00** |

**Figure 11: Accuracy for GJTs in the pre-test and post-test**




*Note:* Condition A (**El perro duermen*. "The dog sleep") and Condition D (* *Los perros duerme.* **"The dogs sleeps"*) are the two ungrammatical sentences.

As seen in Table 16 and in Figure 11, at Week 1 of the program, learners were skillful in their grammaticality judgments in the pre-test in judging both whether a sentence was grammatical or ungrammatical (the overall accuracy in all conditions are at or above 85 %, min= 85 % Condition C (S.pl., V pl.) and max= 95 % Condition A (*S.sg., V.pl)). The correct judgments on ungrammatical conditions were slightly higher than in the grammatical conditions.

At Week 5 of the program, learners improved not only in how accurately they judged grammatical sentences, but especially in how accurately they judged ungrammatical sentences: the overall accuracy in Condition A is 98 % (*S.sg., V.pl) and in Condition D is 99 %. (*S.pl., V.sg.) This suggests that learners found grammaticality

easier to judge at week 5 of the SA program.

In order to check whether the differences between the accuracy in grammaticality judgments obtained in the pre-test (Week 1) and the post-test (Week 5) were significant, two logistic regression with Trial Order and Condition as predictors were run on the pre-test and post-test accuracy data in the GTJS. Table 17 shows the results of these logistic regressions in the pre-test and the post-test respectively.

**Table 17: Logistic regressions for pre-test and post -test in the GJTs**

| **PRE-TEST** | $\widehat{\boldsymbol{\beta}}$ | **St E.** | **z value** | **Pr(>\|z\|)** |
|---|---|---|---|---|
| **Condition B** | 2.1804 | 0.2047 | 10.65 | 0.0000*** |
| **Trial Order** | 0.1004 | 0.2076 | 0.48 | 0.6286 |
| **Condition A** | 0.7031 | 0.3430 | 2.05 | **0.0404 ** ** |
| **Condition C** | 0.4567 | 0.2673 | -1.71 | 0.0875 |
| **Condition D** | 0.3011 | 0.3366 | 0.89 | 0.3709 |
| **(Trial Order Std) × (Condition A vs. Condition B)** | 0.1222 | 0.3473 | -0.35 | 0.7250 |
| **(Trial Order Std) × (Condition C vs. Condition B)** | 0.0210 | 0.2703 | -0.08 | 0.9382 |
| **(Trial Order Std) × (Condition D vs. Condition B)** | 0.3011 | 0.3366 | 0.89 | 0.3709 |

| **POST- TEST** | $\widehat{\boldsymbol{\beta}}$ | **St E.** | **z value** | **Pr(>\|z\|)** |
|---|---|---|---|---|
| **Condition B** | 2.5901 | 0.2456 | 10.54 | 0.0000*** |
| **Trial Order** | -0.2279 | 0.2398 | -0.95 | 0.3421 |
| **Condition A** | 1.1681 | 0.4815 | 2.43 | **0.015 ** ** |
| **Condition C** | 0.2336 | 0.3711 | 0.63 | 0.5290 |
| **Condition D** | 3.8866 | 2.0467 | 1.90 | **0.0576**** |
| **(Trial Order Std) × (Condition A vs. Condition B)** | 0.1928 | 0.4946 | 0.39 | 0.6966 |
| **(Trial Order Std) ×** | 0.6544 | 0.3618 | 1.81 | 0.0705 |

| | | | | |
|---|---|---|---|---|
| **(Condition C vs. Condition B)** | | | | |
| **(Trial Order Std) × (Condition D vs. Condition B)** | 2.3252 | 1.5322 | 1.52 | 0.1291 |

As seen in Table 17, in the pre-test, only Condition A (* S.sg., V. pl.) ($\hat{\beta}$= 0.7031, z= 2.05, *p*= 0.404, *odds*=2.03) was significantly different from Condition B (the baseline), which means that these participants were better at judging grammaticality in that specific condition. This supports McCarthy's (2008) idea that it is more likely to learners to be more predisposed to establish agreement between a plural subject and a singular unmarked (Condition D) verb, than the opposite (Condition A). On in terms of the CM, such combination would be less 'plausible' for L2 learners (Haskell & MacDonald, 2003).

By contrast, Table 17 shows that, in the post-test, both Condition A (* S.sg., V. pl.) ($\hat{\beta}$= 1,1681, z= 2.43, *p*= 0.015, *odds*= 3.33) and Condition D (* S.pl., V. sg.) ($\hat{\beta}$= 3.8866, z= 1.90, *p*= 0.0576, *odds*= 10.15) were significantly different from Condition B in the post-test, which means that learners were able to correctly judge both ungrammatical conditions. Effects sizes (odds) were much higher in Condition D than Condition A, suggesting a slight better improvement for Condition D at Week 5.

Finally, while in the $O_{cl}$VS sentences there was no need to control for handedness since they keys for True and False responses targeting agenthood assignment were counterbalanced so that 50 % of the correct responses were on the left side of the keyboard and 50 % of the correct responses were on the right side of the keyboard, for the GJTs, since participants always needed to use their right hand to indicate that a sentence was non-acceptable (this option was placed to the right of the keyboard), further steps were taken to control for potential effects of handedness on learners responses on

the GJTs. From a total of 20 participants, 3 participants declared themselves to be left-handed (S3, S7, and S12). A close look at the mean accuracy of these three participants reveled that the three of them were extremely accurate in judging grammaticality, as seen in Table 18:

**Table 18: Mean accuracy on GJTs for left-handed participants**

| | Condition | Mean | *SD* |
|---|---|---|---|
| Pre | A | 0.97 | 0.02 |
| **Post** | **A** | **0.97** | **0.02** |
| Pre | B | 0.94 | 0.03 |
| **Post** | **B** | **0.97** | **0.02** |
| Pre | C | 0.80 | 0.06 |
| **Post** | **C** | **0.88** | **0.05** |
| Pre | D | 0.97 | 0.02 |
| **Post** | **D** | **1.00** | **0.00** |

Importantly, Table 18 shows that the 3 left-handed participants exhibited the same tendencies observed in the group overall: they showed improvements in the post-test, and most especially in Condition D (*S.pl., V.sg.). In other words, left-handed participants were successful at judging sentences ungrammatical, which required them pressing a button to their right.

Moreover, a general linear model was run on the accuracy scores for each subject taking into account whether the correct responses were on the left or on the right side of the keyboard; the coefficients ($\hat{\beta}$), which indicate how participants performed on the GJTs, confirmed that left-handed participants were very accurate when the correct answers were on the right, that is, when the sentence was ungrammatical, (S3=$\hat{\beta}$= 18.43,

S12= $\hat{\beta}$= 2.24; S7= $\hat{\beta}$= 0.74) (positive coefficients show that participants exhibited a marked tendency to respond correctly when the correct responses were on the right). These results confirm once again that for participants to be left-handed was not detrimental to their performance on the GJT, and that high accuracy on grammaticality judgments was not related to participants' handedness.

**Discussion**

The goal of this study was to investigate from an online and offline account the influence of SA experience on L2 morphosyntactic development, and, more precisely, on how advanced learners of Spanish process $O_{cl}$VS sentences in Spanish after a 5-week immersion program abroad.

The results of this study bring new empirical evidence showing that morphosyntactic development can be achieved at the end of an intensive short-term SA program, as supported by previous research (Duperron, 2006; Guntermann, 1995; Grey et al., 2015; Isabelli & Nishida, 2005; Yager, 1998). Contrary to what some scholars have argued (e.g., DeKeyser, 2013) *measurable* learning gains in L2 skills other than oral fluency can be derived from the SA experience. This conclusion is based on the significant improvements the learners in this study have shown in the post-test when interpreting $O_{cl}$VS sentences and in their ability to correctly judge a sentence ungrammatical in both ungrammatical conditions (Condition A and Condition D). Both findings suggest that learners were able to better assign semantic functions to non-canonical transitive sentences (i.e., decide correctly 'who does what to whom') at the same time that they were more sensitive to different types of subject-verb agreement violations between the subject and the verb, at least offline. On the other hand, the

changes that were observed in processing costs suggest that learners read slightly faster in Condition B (the baseline, cl.sing., verb sing.), although this change was not significant. Most importantly, latency data also showed that leaners had a clear tendency to give more attention to the verb region and, especially, to the post-verbal subject in the baseline condition as the experiment progressed, which suggests that they might found these regions to be more relevant for assigning agenthood.

With regard to the specific research questions that guided this study, accuracy data have shown that, on the post-test in Week 5, emerging bilinguals were able to interpret $O_{cl}$VS sentences easier than they were able to on the pre-test (RQ1). Even Condition B (cl-sing, verb-sing), the most ambiguous condition in the $O_{cl}$VS sentence stimuli, showed significant improvements, suggesting that SA experience can influence L1 processing strategies to move them in the direction of the L2 processing strategies, thereby mitigating L1 transfer effects, as hypothesized by Lee & Malovrh, 2009; Malovrh & Lee, 2010; Torres, 2003. In CM terms, the framework on which this dissertation is based, the results of this study suggest that SA experience helps learners "reweight" word order as a reliable cue: at the end of the SA program, learners exhibited a readjustment of their L2 cue configurations in accordance with L2 cue validity, meaning that morphology may have been processed as a more reliable cue than word order when the learners assigned agency to $O_{cl}$VS sentences.

As discussed in Chapter One, full differentiation between L1 and L2 cue strength configurations is ultimately expected in L2 development. The question asked in this chapter was whether immersion experience can accelerate that process and help learners maximize form and meaning mapping in $O_{cl}$VS sentences The data from this study

suggest that this might be the case, and this seems to be true for sentences in all agreement conditions (RQ2).

These results also point to the role of L1 inhibition in the triggering of L2 cue hierarchy. As Linck, Kroll, & Sunderman (2009) suggested, the constant use of the L2 in an immersive environment which is rich in input seems to make a difference in how quickly emerging bilinguals "fine tune" their L2 cue strength configurations to L2 cue validity. In the case of this study, due to the program's characteristics, learners were in continuous contact with the target language: either taking content courses, participating in fieldwork activities or conversational interactions with locals in the community, or developing community-based projects. Therefore, the intensity of the input, as well as the number of opportunities leaners had to meaningfully engage in communicative exchanges with native speakers, seems to have been a crucial factor in preventing their L1 processing bias from influencing form-to-function mappings in non-canonical transitive sentences.

Finally, with respect to Condition A, the only condition in which participants in Experiment 1 showed significant accuracy improvements across experimental trials – even though the learners in this SA study also showed improvements *in their accuracy overall* (in relation to an average in accuracy scores between the pre-test and the post-test) – they also exhibited qualitatively different processing patterns in Condition A. Overall, it was the condition in which they improved the fastest across trials. This, then, suggests, as did Experiment 1, that contrastive agreement is beneficial in helping learners align their processing strategies to Spanish cue strength configurations.

This study hypothesized that changes in accuracy would be accompanied by

changes in latency in the processing of $O_{cl}$VS sentences, with learners showing overall gains in efficiency. Quantitatively, overall RT data in this study show that in Week 5, emerging bilinguals were able to read $O_{cl}$VS sentences slightly faster than in the pre-test, although this difference was not significant (RQ3). However, interesting qualitative differences emerged. Participants in the post-test showed a statistically significant tendency to spend more time on the verb region and, even more, in the post-verbal subject region over the course of the experiment in the baseline condition (Condition B, cl-sing; verb-sing.), which was the most ambiguous condition. The verb region conveys important morphological information to disambiguate $O_{cl}$VS sentence interpretations, so this might have explained the change. By contrast, conditions conveying contrastive agreement information between the clitic and the verb did not receive higher RTs in the verb region on the post-test, suggesting that on the post-test, learners seem to have looked at the verb region more consistently, regardless of whether it conveyed contrastive agreement or not. Considering the experiment overall (i.e., taking into account the pre-test and the post-test RTs) learners showed a tendency to read the verb region across experimental trials slightly faster in the two conditions in which agreement was not useful, although this tendency was not significant (RQ4).

GJTs were included in order to ensure that, in the eventuality that sensitivity to subject-verb agreement did not emerge in on-line measurements (RT data), it would still be possible to determine whether learners used grammatical knowledge so as to judge sentences conveying subject-verb agreement violations offline.

In relation to learners' ability to judge subject-verb agreement violations in SV sentences, the results from the GJTs showed that participant's grammaticality judgments

significantly improved by the end of the SA program (RQ5). These results support Grey et al.'s (2015) findings that GJTs improve after exposure to intensive and implicit input abroad. In that study, the authors found higher accuracy for GJTs on word order in the visual modality (e.g., **Tengo que millas muchas correr*. * "I have to a lot of miles run"). The findings in the present study extend those reported by Grey et al., (2015) to include a new target structure and theoretical framework. The present study also coincides with Grey et al., in showing that sensitivity to ungrammaticality offline interacts with the quality of the input.

With regard to the online data for SV sentences (testing sensitivity to subject-verb agreement violations in SV sentences online), no effect of grammaticality was shown in the residual RTs on the verb region or spillover region (RQ6). However, there was a tendency to spend more time on the verb when the verb was ungrammatical in Condition A and Condition D in the pre-test. This tendency lost force in the post-test, although the spillover region received slightly higher RTs in Condition A in the post-test. It may be the case that RTs in the SV sentences overall were influenced by the speed with which learners tended to read the verb region over the course of the experiment in these particular sentences. More power is needed to determine how learners process SV sentences with subject-verb agreement violations on-line. This is a limitation that the present SA study shares with Experiment 1.

**Conclusions**

The central finding in this SA study, conducted *in situ* with advanced learners participating in a short 5-week SA program, is that immersion experience can benefit the readjustment of L2 cue hierarchy to better correspond to L2 cue validity in the processing

of Spanish $O_{cl}VS$ sentences, a structure which is difficult to acquire and process for English-speaking advanced learners of Spanish. This finding has the potential to advance SLA research on word order and other morphosyntactic phenomena in relation to SA learning contexts. The results of this study support the idea that the SA experience is an important factor that can impact morphosyntactic development. However, the online data in the SV ungrammatical sentences reported here showed that some morphological phenomena (e.g., sensitivity agreement in SVcanonical sentences) are harder to grasp. Overall, the findings in this study contradict LaBrozzi (2012), who suggested that SA experience does not seem to influence L2 processing towards more target-like patterns (in his study, the author was interested in studying the processing of morphological and lexical cues to assign verbal tense in L2 Spanish). LaBrozzi's work is the only previous study in the SA research literature that has used an online measurement to test the SA experience on sentence processing in the written modality, although the author did not adopt a pre-post-test design and did not conduct the study *in situ*, as did the present study. By contrast, LaBrozzi (2012) collected data from participants after participants had been back from their SA experience for approximately three months. More studies using online methodology *in situ*, such as SPR or eye-tracking are needed to investigate how SA exposure shapes L2 processing in different L2 learners populations such as advanced learners and also heritage speakers studying abroad.

## CHAPTER FOUR: Feedback

### Assessing instructional factors in advanced learners' $O_{cl}VS$ sentence processing: Does feedback make a difference?

### Introduction

Scholars in the field of Second Language Acquisition (SLA) have argued that corrective feedback plays a facilitative role in the acquisition of L2 forms that are difficult to learn through input alone, for example, forms that are perceptually non-salient (e.g., Leeman, 2003; Long, 1992; Long & Robinson, 1998). However, L2 feedback research conducted from an interactionist perspective has also suggested that corrective feedback, specifically in the form of recasts, does not seem to be equally valuable for all L2 forms, all aspects of language, or all levels of proficiency. Feedback may be less influential for morphosyntactic phenomena, for more complex linguistic target items, for forms that are in the early stages of development, or for developmentally unready L2 learners (e.g., Carpenter, Jeon, MacGregor, & Mackey, 2006; Han, 2002, 2008; Lyster, 2004; Mackey, Gass, & McDonough, 2000; Long, Inagaki, & Ortega, 1998; Mackey & Philp, 1998).

Therefore, even though there is a certain degree of consensus that corrective feedback is beneficial for L2 development, an idea that has been supported by several meta-analyses conducted in this domain (see Li, 2010 for research on feedback and computer-delivered feedback; Lyster & Saito, 2010 for research on feedback in the classroom setting; Mackey & Goo, 2007 for research on language interaction; Norris & Ortega, 2000 for effectiveness of feedback as a type of L2 instruction; Russell & Spada,

2006 for grammar-oriented feedback), SLA researchers have not yet arrived at conclusive answers regarding which types of feedback are most effective (e.g., implicit vs. explicit), for which areas of language it is most effective (e.g., lexis vs. morphosyntax), or at which stages of learning (early vs. late) it is most helpful to L2 learners. For this reason, Li (2010) has argued that, since "the effect of corrective feedback has been established," now it is time for researchers to "embark on the mission of investigating the factors constraining its effectiveness" (p. 349). More specifically, Li's (2010) meta-analysis has revealed that, in order to identify potential moderating variables, further research needs to be conducted with: (a) L2 learners of higher proficiency, (b) feedback implemented in the computer mode, (c) linguistic phenomena involving L1 transfer effects, and (d) complex linguistic target structures.

The study in this chapter is designed to move in the direction indicated by Li (2010). Whereas numerous studies have been conducted on feedback's effectiveness, especially on the role of the recast on uptake during conversational exchanges (for overviews of interactional feedback see Ellis & Sheen, 2006; Long, 2007; Mackey, 2007; Nicholas, Lightbown, & Spada, 2001), further research is still needed to continue advancing our understanding of how computer-delivered feedback affects learners' outcomes in Computer-Assisted Language Learning (CALL) environments (see Heift, 2004; Bangs, 2002; and Felix 2002), especially when learners process complex target structures. With this goal in mind, the present study looks at whether $O_{cl}$VS sentence processing in the written modality by advanced English-speaking learners of Spanish is affected by computer-delivered feedback.

The target structures examined in the present study, $O_{cl}$VS sentences in Spanish, "La mira el estudiante" (her-clitic.fem.sing. look at-3rd p. sing. the student, "The student looks at her"), are relevant to the study of the potential role of computer-delivered feedback on sentence processing since such structures are highly frequent in the target language, yet their persistent misinterpretation leads to communication breakdowns for L2 learners (i.e., failing to establish the correct form and meaning mappings by interpreting the patient of the $O_{cl}$VS sentence as the agent). In spite of these structures' importance for effective communication, previous studies have shown (e.g., Lee & Malovrh, 2009; LoCoco, 1987; Montrul, 2010; Malovrh & Lee, 2010; Sanz & Morgan-Short, 2004; VanPatten 1984; VanPatten & Houston, 1998; VanPatten & Borst, 2012), and the results obtained in Experiment 1 and the SA study (Experiment 2) of this dissertation have suggested, that L2 learners at different levels of proficiency continue exhibiting difficulties in determining 'who does what to whom' in non-canonical transitive sentences. From the perspective of the Competition Model (CM), the framework adopted in this dissertation, English-speaking learners tend to overwhelmingly rely on word order as the most valid cue for agency assignment in L2 Spanish, since word order is the most valid cue for agenthood in English. The phonotactic characteristics of clitics – which are low in saliency, both visually and auditorily, which have multiple forms (*la*-her, *lo*-him, *me*-me, *no*s-us, etc.) corresponding to one function (patient), and which are dependent on preceeding NPs for meaning –conspire to make $O_{cl}$VS structures difficult to process even for advanced learners.

Sanz and Morgan-Short (2004) and Moreno (2007) are the only studies to have examined $O_{cl}$VS sentence processing by English-speaking learners of Spanish and to

have also used computer-delivered feedback on learners' responses (item-by-item). Whereas Moreno's (2007) study consisted mainly of production tasks accompanied by think-aloud protocols, Sanz and Morgan-Short (2004) studied $O_{cl}$VS sentences in both the aural and written modality following the guidelines of Input Processing (IP).

In their work, Sanz and Morgan-Short (2004) compared, in a pre-test/post-test design, the performance of four experimental groups resulting from the combination of two independent variables: [+/- prior explanation] and [+/- explicit feedback]. During the treatment, first- and second- year English-speaking L2 learners of Spanish heard and read a set of $O_{cl}$VS sentences and subsequently performed a visual agent-choice task in which they were presented with two pictures, and they were asked to select which one better represented the agent. Results from the assessment phase (post-test) showed that all four experimental groups made significant improvements. Therefore, the authors concluded that, in the context of these structured-input activities were created to push learners to process the target structures and alter their L2 processing strategies (e.g., Cadierno, 1995; Farley, 2001; VanPatten & Cadierno, 1993; VanPatten & Oikkenon, 1996; VanPatten & Sanz, 1995), explicit information did not add improvement to the learning outcomes. However, Sanz and Morgan-Short (2004) acknowledged that, since they did not compare the implicit feedback condition ("OK"/"Sorry, try again!") with a condition in which learners received no information regarding the correctness/incorrectness of their responses, it would be worthwhile for future studies to investigate whether the provision of yes-no feedback makes a difference in the processing of non-canonical sentences, especially when the L2 input is structured so as to be useful to the learner. Additionally, the authors of that study argued that there is a need to further address how the provision

of feedback may affect sentence processing on-line, that is, as processing unfolds over time.

The research design of the present study has taken into consideration the directions and gaps pointed out by Sanz and Morgan-Short (2004). First, the present study employed a within-subjects design to address how learners process $O_{cl}$VS sentences that exhibited different agreement manipulations between the clitic and the verb: two conditions presented contrastive agreement cues between the clitic and the verb, and the other two did not. Contrastive number morphology between the clitic and the verb was predicted to aid advanced learners in circumventing their tendency to rely on word order when processing $O_{cl}$VS sentences. Processing was measured offline and online by analyzing, respectively, the accuracy rates and processing costs (Reaction Times, RTs) yielded by a word-by-word Self-Paced Reading task (SPR). Second, this study compared two experimental groups of advanced English-speaking learners, ([+ Feedback]) and ([- Feedback] in order to assess potential differences in the learners' accuracy rates and processing costs in the four agreement conditions listed above. The [+ Feedback] group received computer-delivered feedback on their responses while the [- Feedback] group did not; that is, feedback was manipulated as a between-subjects variable.

Moreover, the present study's design is distinct from other L2 feedback studies conducted in L2 Spanish, which have mainly employed pre-test and post-test designs (e.g., Bowles, 2005; Moreno, 2007; Rosa & Leow, 2004; Sanz & Morgan-Short, 2004). By contrast, in this study, accuracy on T/F comprehension responses was analyzed both overall and over the course of the experiment; in this way, it was possible to compare the rate of improvement across experimental trials in each of the within-subjects agreement

conditions as well as between the experimental groups. Previous Computer Assisted Language Learning (CALL) studies, such as Heift (2001, 2002, 2004, 2010) and Sasaki (1998) have also analyzed accuracy improvement across experimental trials to assess the effectiveness of computer-delivered feedback on language learning. Whereas Heift (2001, 2002, 2004, 2010), who followed an interactionist approach to CALL, examined the influence of computer-delivered feedback in promoting immediate uptake in L2 German, Sasaki (1998) examined, from a Competition Model (CM) perspective on CALL, the influence of computer-delivered feedback on L2 learners' adjustment of cue strength in non-canonical sentences in L2 Japanese.

In sum, the goal of the study in this chapter is to address potential interactions between agreement cue manipulations between the clitic and the verb in $O_{cl}$VS sentences, the provision or absence of implicit feedback on learners' accuracy responses (True/False comprehension questions assessing $O_{cl}$VS sentence agency assignment), and the improvement (item order) across experimental trials. As noted earlier, $O_{cl}$VS online sentence processing in Spanish has been addressed in a previous L2 feedback study (Moreno, 2007), which employed think-aloud protocols. While Moreno's use of think-aloud protocols allowed researchers access to participants' self-reports on their cognitive processes, the methodology used in this study, word-by-word Self Paced Reading (SPR), promises to shed new light on potential processing patterns taking place as reading takes place over time, that is, in each of the regions of the sentence. As Sagarra (2007) stated in regard to the SPR technique: it allows researchers to "measure attention as processing unfolds, and its use continues a line of research which is much needed in SLA" (p. 134).

Finally, by evaluating from both an offline and online account the effectiveness of computer-delivered feedback on a complex structure amongst advanced learners, this study proposes to make contributions to the field of CALL research. As discussed by Chapelle (2001, 2009), following Doughty (1987), learner-computer interactions can inform researchers of learners' processes and strategies that are relevant to the perspective of a particular theory. At the same time, theory needs to play a central role in designing CALL research, which will ultimately inform practice. In this study, the role of computer-delivered feedback on sentence processing is analyzed from a Competition Model perspective, and the theoretical question to be answered is whether learners need feedback to alter their current (L1-based) cue weights or whether structured input alone is sufficient for this adjustment to take place.

## Background

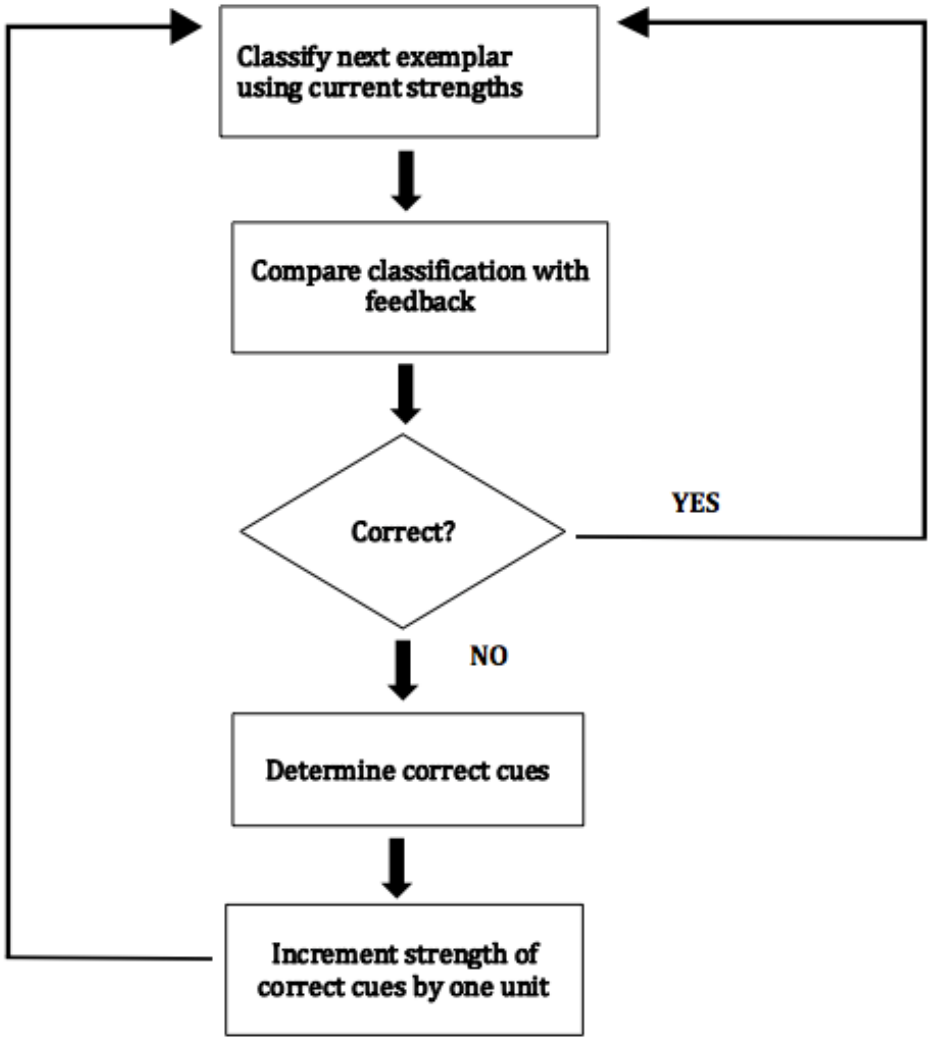
### The role of feedback from a Competition Model (CM) perspective

As discussed above and in previous chapters, the Competition Model predicts that L2 learners will transfer the cue strength from their first language into the second language. With exposure to the L2, learners need to evolve a sense as to which cues are the strongest and most reliable in order to establish correct form-meaning connections. In addition, MacWhinney (2012) argued that negative feedback can draw learners' attention to invalid form-meaning mappings, helping them "fine tune cue weights" (p. 18).

Of interest for this study, McDonald (1987, 1989) proposed an extension of the CM by including the role of feedback as an essential part of cue strength reweighting in L1 and L2 acquisition, as shown in Figure 12. McDonald (1989) called this learning mechanism "learning-on-error" since only feedback applied to errors would lead to cue

strength alteration.

**Figure 12: A mechanism for learning-on-error to determine L1 and L2 cue-strengths (McDonald, 1987, 1989)**




Based on the learning-on-error mechanism, in childhood, learners enter the mechanism with an initial cue weight of zero, having no information upon which to base their categorizations (e.g., agenthood, patienthood, etc.). Therefore, they first make guesses as to the correct categorization, and as time progresses, they compare their

categorization with feedback available in the environment, for example, to an actual external situation, to information received when the sentence is rephrased, or to overt corrections. If feedback indicates that a categorization is correct, or if feedback is not available, learners' cue strengths are not adjusted. But if feedback indicates that an incorrect categorization has been made, the strengths of the cues that would have given a correct categorization are increased by a small amount. From that point on, the new cue weights are used for judging new exemplars (i.e., input sentences). Thus, according to McDonald (1989), "cues are only adjusted on exemplars that are miscategorized" (p. 377-378). Moreover, in order for cue weights to continue readjusting, the cue must first be detectable, since the acquisition of form-to-function mappings seems to be influenced by several different factors, including perceptibility (Cazden & Brown, 1975; MacWhinney, 1978; MacWhinney, Pleh, & Bates, 1985, Slobin, 1973). This acquisitional process continues until all relevant cues have gained sufficient strength. According to McDonald (1987, 1989), the mechanism applies to L2 processing as well.

When referring to cue weight readjustments, McDonald (1987, 1989) distinguished between two types of sentences. *Non-conflict sentences* are those sentences in which all cues agree, and cue strengths are added up in order to determine the final interpretation. Non-conflict sentences are correctly interpreted because all cues cooperate; therefore, it does not matter which set of cues the learner uses, since all cues will lead to the correct form-to-function mappings. But languages, especially those which are morphologically flexible, also include *conflict-sentences* in which cues disagree, and cue strengths compete until a final interpretation is made. Conflict sentences are

interpreted correctly when the strongest cue is the one that is most often correct in a given conflict situation.

Whereas measuring cue validity in both non-conflict and conflict sentences yields *overall cue validity* (which determines the order of cue acquisition in a particular language), by contrast, assessing validity through conflict sentences alone yields a "more specific" type of cue validity: *conflict validity*. Maximum correct interpretation occurs when learners have command of all relevant cues and the learner's cue strength is set to conflict validity rather than to overall validities.

With respect to why agency assignment is problematic for English speakers learning a more richly inflected language, McDonald (1989) argued that since conflict with SVO order in English is either "not allowed" or "resolved in favor of the word order cue," the conflict validity of SVO word order in English is 100%. In addition, SVO word order is also the cue with the highest overall cue validity and, therefore, it is the first cue acquired by English native speakers (p. 382-383).

When it comes to second language acquisition, the model predicts that, for L2 language processing, the first cue to be significantly mapped to a function will be the one with highest overall cue validity in the L1 (Kilborn & Cooreman, 1987; McDonald, 1987). But increasing exposure to the target language should cause cues strengths to gradually shift in the direction of the conflict sentence validities in the L2. The speed with which this shift takes place depends on the number of conflict sentences encountered. That is why frequency is central to setting the cue strengths correctly in specific conflict cases.

McDonald (1987), for example, found that with increasing exposure to the second language, English-Dutch adult bilinguals gradually shifted from first to second language cue weights when they were asked to assign the actor role in Dutch, a more richly inflected language than English. McDonald did not directly assess the role of feedback or the error-on-learning mechanism in her study, but Sasaki (1998), who will be discussed in the next session, did test McDonald's (1987, 1989) proposal.

The study in this chapter tests whether feedback on invalid form-to-function mappings makes a difference in learners' $O_{cl}VS$ sentence processing, as predicted by the CM and extensions of the CM such as the error-on-learning mechanism proposed by McDonald (1987, 1989). In addition, the present study assesses whether, in the presence of such feedback, certain agreement combinations – contrastive agreement cues between the clitic and the verb, which are more informative than non-contrastive agreement cues and, thus, potentially more 'detectable,' assuming English-speaking advanced learners of Spanish are sensitive to such agreement mismatches – lead to significantly higher accuracy rates.

**The effect of computer-delivered feedback on morphosyntactic development**

Sasaki (1998) tested the pedagogical feasibility of McDonald's (1987, 1989) proposal through the use of feedback in a computer-controlled learning environment. Sasaki's study consisted of four sessions: pre-test, Feedback I, Feedback II, and post-test. During the two Feedback sessions, intermediate English-speaking learners of L2 Japanese ($N$=9) listened to a series of canonical and non-canonical sentences conveying case markers (active and causative sentences). Following each sentence, participants were shown a set of six pictures of animals and were asked to select which picture represented

the “doer” of the action. Learner responses were then followed by feedback (“That’s right!”/”That’s wrong!”) which, in addition, was accompanied by contrasting sound effects for correct / incorrect answers. Reaction Time (RT) data were gathered to examine the latency of learners’ responses. After the feedback messages were presented on the screen, the same set of pictures was displayed again with the correct choice circled. Participants heard the same set of sentences ($k$=24) in both feedback sessions.

Sasaki’s prediction, following McDonald’s (1987, 1989), was that learners should make increasingly fewer interpretation errors over the course of the feedback sessions. Accuracy results demonstrated that while agency assignment was problematic for learners in the pre-test due to the influence of word order, after provision of feedback, the intermediate learners made significant improvements in their interpretation of all types of sentences assessed. These results confirmed that feedback had an effect on L2 learners’ adjustment of cue strength.

With regard to learners’ response-time latency, the analysis of the RT data showed an effect of order in the non-canonical sentences: learners exhibited a slowdown in their responses for both types of non-canonical sentences, and also in their incorrect responses. However, overall, learners’ latency generally shortened throughout the feedback training sessions. This last finding indicated that, as an effect of feedback, learners became faster in making form-to-function connections. Sasaki suggested that it would be desirable in future studies to use a comparison group that receives no feedback in order to tease apart its effects on sentence processing from other factors. The study in this chapter uses such a comparison group in its design.

Another limitation of Sasaki's (1998) study, which the present study intends to address, is that the researcher measured on-line processing by means of overall response-time latencies. As discussed in Chapter One, overall response-time latencies are a useful source of information for sentence processing; however, unlike word-by-word SPR data, they do not allow researchers to follow potential patterns that emerge as sentence processing unfold over time, that is to say, in each region of the sentence. For that reason, the present study includes a SPR methodology, word-by-word to discuss possible emerging patterns in the RT data.

**Computer-delivered feedback characteristics and operationalizations**

As noted by Sachs (2011), the seminal work of Carroll, Swain, and Roberge (1992), and Carroll and Swain (1993), the two highly controlled laboratory studies conducted in the early nineties which have had a significant influence on posterior L2 feedback research, already started to use feedback in a similar manner to how it has been done in recent CALL studies, that is, on a one-on-one and item-by-item basis.

Computer-delivered feedback *per se* has been characterized by many scholars (Heift 2004; Chapelle, 2009; and Sachs, 2011) who have used computer-delivered feedback in CALL studies, as qualitatively different from the feedback provided in conversational exchanges. For example, several studies have also referred to computer-delivered feedback as 'individualized', 'consistent' and 'intense' (e.g., Han, 2002; Sagarra, 2007; Sanz, 2004; Nagata, 1993; Nagata & Swisher, 1995); and its use has been argued to improve a study's internal validity as well as its replicability (Rosa & Leow, 2004).

Additionally, even though it seems that computer-delivered feedback may be

perceived as "overwhelming" at times (Nagata & Swisher, 1995), it has been argued that, overall, computer-delivered feedback promotes positive attitudes in L2 learners. For example, Nagata and Swisher (1995) reported that the participants in their study found computer-delivered feedback helpful. Also, Despain (1997), who studied the effectiveness of computer-delivered feedback on listening comprehension in a group of English-speaking learners of L2 Spanish, reported that computer-delivered feedback promoted a positive attitude towards the exercises provided to the learner and towards language learning in general.

Finally, with regard to computer-delivered operationalizations in CALL studies, Heift (2004) has argued that, because of the mode of delivery, feedback in CALL cannot be 'identical' to the feedback provided orally in the classroom. In a similar vein, Sachs (2011) pointed out that the definitions of the terms 'explicit' and 'implicit' feedback, as used in discrete-item CALL studies, differ critically from how they are understood in more communicative contexts. That is, in L2-classroom and interaction research "the term *explicit* usually means that attempts are made to draw learners' attention to non-target-like aspects of their production in such a way that they are likely to realize they are being corrected," and, by contrast, "the term i*mplicit* means that learners are not overtly informed of the unacceptability of their utterances and, therefore, may not necessarily recognize any corrective intent." (p.16) In CALL research, on the other hand, where practice takes place item-by-item, "learners' recognition of the corrective nature of the computer-delivered feedback is often essentially taken for granted" (p. 16). As a consequence of these differences in operationalizations, Sachs (2011) noted that 'implicit' error messages in CALL – that is, simply informing learners of their

correctness/incorrectness – might be counted as 'explicit' in the context of interaction research, since the messages' corrective intent would be considered to be recognizable by the learner; whereas, the contrary situation, an 'implicit' recast in interaction, might be counted as 'explicit' in the context of discrete-item CALL practice since this recast would be considered to provide additional information beyond simple acceptance or rejection of a learner's L2 interpretation or production.

Therefore, whereas in the context of CALL research, previous studies in L2 Spanish using computer-delivered feedback yes-no feedback has been labeled as implicit since no metalinguistic explanation was provided to the learners (e.g., Sanz & Morgan-Short, 2004; Rosa & Leow, 2004, Moreno, 2007), the present study refers to the feedback provided after each experimental trial as [+ feedback]. This feedback consisted of informing participants of the correctness or incorrectness of their responses: no additional information was provided.

**Previous studies on $O_dVS$ sentence processing in L2 Spanish using computer-delivered feedback**

With the exception of a few studies (e.g., Despain, 1997; Sagarra, 2007), most studies conducted in L2 Spanish have made use of computer-delivered feedback (e.g., Camblor-Portilla, 2007; Bowles, 2005; Rosa, 1999; Rosa & Leow, 2004; Rosa & O'Neill, 1999; Sagarra & Abbuhl, 2013) to study its effectiveness in relation to different types of feedback (implicit vs. explicit). Some of these studies (e.g., Rosa & Leow, 2004; Sanz & Morgan-Short, 2004; Stafford, Bowden, & Sanz, 2012) have also combined the manipulation of feedback with other pedagogical variables, such as the presence or absence of grammatical explanation prior to practice. Comparing the effectiveness of

feedback type (implicit vs. explicit) is out of the scope of the present study; therefore, this section will be limited to discussing the main findings of the two previous studies (Sanz & Morgan-Short, 2004 and Moreno, 2007) which have investigated the processing of $O_{cl}VS$ sentences in L2 Spanish using computer-delivered feedback.

As discussed earlier, Sanz and Morgan-Short (2004) examined beginner English-speaking learners' ability to assign semantic roles correctly in VS sentences (e.g., *Lo besa la chica*, 'Him kisses the girl' = 'The girl kisses him'). Four experimental groups were presented with different combinations of two independent variables: [+/- prior explanation] and [+/- explicit feedback]. The treatment activities, which included 56 items, involved aural and written interpretation at the sentence level (i.e., identifying which of two pictures represented the meaning of a given sentence) and written interpretation at the text level (e.g., reading a short article and deciding which of two titles was appropriate for the article). All participants were given feedback on each response (i.e., either "OK" or "Sorry, try again!"). The test phase involved both interpretation and production. For the 15 interpretation items (10 of which involved the linguistic target), participants were asked to choose which of two pictures represented the meaning of a sentence. One of the production tests consisted of similar activity: For each of 15 items (11 targets), the participants were asked to fill in a blank with the appropriate pronoun needed to express 'who was doing what to whom' in a given picture. The second production test involved watching a silent video that portrayed several interrelated events and then retelling the story in writing, a narrative which naturally called for the use of pronouns. Results showed that all of the groups showed statistically significant improvement on both the interpretation and the production tests, with no significant

differences amongst the conditions. The authors argued that the task itself (i.e., structured input) seems to have led to acquisition (p. 43). However, since no comparison group was created in that study to contrast implicit feedback to no feedback, the researchers called for future research to further explore whether structured input alone produced the benefits. Sanz and Morgan-Short explained that the differences between their results and those of other studies in SLA (e.g., Carroll & Swain, 1993; Nagata, 1993, Nagata and Swisher, 1995, Rosa, 1999) may have been due to the nature of the input, the degree of task-essentialness in the implicit feedback condition, or the fact that they provided feedback on the input processing of complex linguistic target instead of feedback on oral production.

In Moreno's (2007) computer-based experiment, beginner English-speaking learners of L2 Spanish were randomly assigned to four treatment conditions that conveyed different combinations of the variables [+/- task essentialness] and [+/- explicit feedback]. All of the participants produced concurrent think-aloud protocols in the treatment and assessment phases (pre-test, post-test, and delayed-post-test). The researcher tested whether type of feedback had a differential effect on the learners' recognition, and oral and written production of $O_{cl}VS$ sentences. The treatment task consisted of describing pictures ($k$=16) in which one of the persons depicted was performing an action that clearly affected another person. Learners in the task-essential conditions played a game that involved making linguistic choices as they traveled down a path. When the learners chose the correct object clitic to describe the picture, they were allowed to continue on their journey, and the computer completed the sentence for them, adding the verb and subject. In the implicit feedback group, learners were informed only

about the correctness of their choices. In the explicit feedback group, they were given a metalinguistic explanation every time they clicked on an answer; this was done for both correct and incorrect responses so as to ensure that all learners received the same amount of information. In both groups, when the learners' answers were wrong on a first attempt, they were prompted to try again. After that, the computer provided the correct answer. The recognition test involved choosing which sentence could be used to describe a picture (participants were given 3 different options or 'none of the above'). In the production tests learners needed to describe pictures conveying an actor (s) and a patient(s) in written or orally.

Moreno (2007) found that explicit feedback was as beneficial as implicit feedback. One of the limitations of the study, however, is that the instructions given to learners in the picture recognition test may have been unclear. In the sample item provided in the study's report, for example, the picture showed two men (Juan and Luis) and two women (Eva and Ana), and next to the women were two circles. It was not clear whether the circles were pointing at the subject or at the object of the sentence, and learners' instructions were not provided either; therefore, the possibility exists that, even if learners had a solid understanding of the target items, they still may not have known whether they should select "*Los odian Eva y Ana*" (Eva and Ana hate them-pl,masc.) or "*Las odian Juan y Luis*" (John and Luis hate them-pl. fem.). In addition, according to the experiment's description, the researcher also manipulated subject and object clitic number, since pictures conveyed, for example, a plural subject and a singular object, a plural subject and a plural object, or a singular subject and a singular object. However, number manipulation was not systematically studied by the author, and, since it was not

considered as an independent variable, whether learners' outcomes were affected by this manipulation was not statistically analyzed nor was it reported. Finally, the number of treatment items (*k*=16) was small in the study.

**The present study**

Based on the previous literature, gaps and discussion in the above sections, these are the research questions that guide the present study.

**Research questions (RQs) and Predictions (Ps)**

- RQ1: Does accuracy rates in the experiment overall differ between a group which received computer-delivered feedback on accuracy responses and a group which did not receive feedback?
- P1: Based on the CM (MacWhinney, 2011) and extensions of the CM (McDonald, 1987, 1989) and the empirical evidence provided by recent meta-analysis which have assessed the effects of computer-delivered feedback on language learning (Li, 2010), it is expected that the group receiving computer-delivered feedback will outperform the group that does not receive such feedback.
- RQ2: Do learners in the [+feedback] and the [- Feedback] groups exhibit similar accuracy rates, overall and across experimental trials, in the matching/mismatching conditions between the clitic and the verb?
- P2: Based on the results from Experiment 1, for which feedback was provided to learners and the pre-test of the Study Abroad experiment (Experiment 2) for which feedback was not provided, it is expected that: (a) both experimental groups will show higher overall accuracy rates in Condition A (cl. sing,, verb pl.), the contrastive agreement condition that seems to be more useful for L2 learners,

(b) the [+ Feedback] group is likely to improve more quickly across experimental trials in the sentences where clitic and verb mismatch in number, this probability is even higher for Condition A (cl. sing., verb pl.), the condition in which learners in Experiment 1 showed *significant* improvements across experimental trials, (c) the [- Feedback] group is likely to show a slower rate of improvement across experimental trials due to the absence of feedback.

- RQ3: When interpreting $O_{cl}$VS sentences, do learners who receive computer-delivered feedback exhibit the same processing costs as learners who do not receive feedback? And if so, do processing costs between the two experimental groups differ in the $O_{cl}$VS sentences where clitic and verb match/mismatch in number?
- P3: The experimental groups will not show the same processing costs overall. If feedback causes learners to spend more time in the verb region as suggested by the results in Experiment 1, higher RTs overall are expected in the verb region for the [+ Feedback] group than in the [- Feedback] group. With regard to the matching/mismatching conditions, if learners misinterpret the clitic pronoun as the agent of the sentence, and if this creates an expectation that the clitic and verb should agree in number, learners in both groups should exhibit more processing difficulty (i.e., higher RTs) when verb and clitic mismatch in as opposed to when they match in number. This difficulty is expected to arise at the verb regions or in the subsequent regions for both experimental groups [+ /- feedback]. However, this slowdown is expected to be even more marked in the [+ Feedback] group since feedback is likely to draw learners' attention to the verb region. Finally,

regarding order effects in the RT data, that is the speed in processing costs across experimental trials, based on Sasaki (1998), it is expected that the provision of feedback will trigger an overall speedup as the experiment progresses in the [+ Feedback] group, in comparison to the [- Feedback] group.

## Methods

### Description of participants

Participants in this study (*N*=90) were advanced learners of L2 Spanish who were native speakers of English (48 female; *mean age* = 19.72 years). Proficiency level in L2 Spanish was based on participants' institutional enrollment in an advanced Spanish course. All participants in this study were enrolled in an Advanced II non-intensive course at Georgetown University and were recruited to participate in this study in exchange for extra credit in their Spanish course. Students were enrolled in these courses either based on results of a fall validation exam, from their performance on a placement exam, or from previous course enrollment. All participants completed the grammar section of the Placement Exam created by the Spanish and Portuguese Department at Georgetown University. The grammar section of the placement exam (*k*=37) consisted of four-option multiple-choice items targeting a range of grammatical structures in Spanish (*Mean*= 70%, *SD*=3.15; Skewness= 0.03 *Kurtosis*=-0.77); the value of skewness and kurtosis, which was within the normal range of +/- 2, and the non-significance of the Kolmogorov- Smirnov test (*D*(90)= 0.08, *p*= 0.14) indicate that the means are normally distributed. In addition, the data presented no outliers: all mean scores fell within two standard deviations from the mean. Leaners also completed a language questionnaire based on Li, Zhang, Tsai and Puls (2013). Five participants declared that they spoke

languages other than English: L1 Chinese (*n*=1), L1 French (*n*=1), L1 Korean (*n*=1), L1 Malagasy (*n*=1), and L1 Rumanian (*n*=1). Four participants declared that they had studied other Romance Languages at the beginner level: Italian (*n*=2), French (*n*=1), Latin (*n*=1). In addition, to ensure that participants stayed on task throughout the experiment, only participants with at least 70% or higher accuracy in their responses on the comprehension questions for filler items were included in the analysis. Overall, from an original sample of 97, seven participants were excluded from the analysis for not being native speakers of English (*n*=5), for not falling within ± 2 SD from the mean criteria in the grammatical section of the Spanish Placement Exam (*n*=1), and for not meeting the 70% filler criterion (*n*=1).

### Research design

This experiment manipulated:

- [Number] (singular/plural) in the clitic and in the verb, and [Agreement] (match/mismatch) as factors, which yielded four within-subjects conditions arranged in a 2 x 2 design (Latin square): two were matching conditions between the clitic and the verb: [cl.sing., verb.sing], [cl.pl., verb pl.] and two were mismatching conditions between the clitic and the verb: [cl. sing., verb pl.], [cl. pl. verb sing]. The matching/mismatching conditions were labeled A, B, C and D. Matching conditions were Condition B [cl. sing., verb sing] and Condition C [cl. pl., verb pl.]. Mismatching conditions were Condition A [cl. sing., verb pl.] and Condition D [cl. pl. verb sing]. (See Table 13. The $O_{cl}VS$ sentences were delivered by means of a word-by-word Self-Paced Reading (SPR) methodology.

- [Feedback] (Feedback/No Feedback) as a between-subjects variable, which yielded two experimental groups: [+ Feedback] (*n*=45) and [- Feedback] (*n*=45).

The word-by-word Self-Paced Reading (SPR) which participants in the [+ Feedback] and [- Feedback] performed, consisted of 32 $O_{cl}$VS sentences (8 per agreement condition) and 48 fillers in the present simple, which were distributed across four lists (Latin Square). Table 19 shows an example of each of the experimental items.

**Table 19 Sample set of $O_{cl}$VS sentences (Latin Square design)**

| |
|---|
| **$O_{cl}$VS sentence conditions:**<br>**Contextual sentence for $O_{cl}$VS sentences:** e.g., *La camarera y los cocineros conversan en el corredor* "The waitress and the cooks chat in the hallway.") |
| **$O_{cl}$VS sentences (main sentence):** |
| **Mistmatch**, Cl-sg;.: *Allí la aburren los cocineros con historias familiares*<br>CONDITION A "There the cooks bore her with stories about her family." |
| **Match**, Cl-sg.;.: *Allí la aburre el cocinero con historias familiares*<br>CONDITION B "There the cook bores her with stories about her family" |
| **Match** Cl-pl..: *Allí las aburren los cocineros con historias familiares*<br>CONDITION C "There the cooks bore them with stories about her family" |
| **Mistmatch** Cl-pl..: *Allí las aburre el cocinero con historias familiares*<br>CONDITION D "There the cook bores them with stories about her family." |
| **Example of True/ False questions for $O_{cl}$VS sentences (comprehension):**<br>*Los cocineros aburren a la camarera*. "The cooks bore the waitress". TRUE<br>*Las camareras aburren a los cocineros.* "The waitress bores the cooks". FALSE |

**Fillers**
**Contextual sentence for the filler item**:
Esta semana hay finales en la universidad
"This week there are finals at the univesity"

**Filler itmes (main sentence):**
Los alumnos están cansados de estudiar para los exámenes.
"The students are tired of studying for the exams"

**Example of T/F questions for fillers (comprehension):**
*Los estudiantes están cansados de estudiar para los exámenes.* "The students are tired of studying for the exams'.TRUE
*Los estudiantes están contentos de estudiar para los exámenes*."The students are happy to study for the exams". FALSE

---

## Materials

### *$O_{cl}$VS sentences*

$O_{cl}$VS sentences consisted of transitive sentences, which were preceded by a contextual sentence that contained a suitable referent for the clitic. In this study, taking into account that fillers were preceded by contextual sentences, so as to make the experiment more manageable, the number of target items per agreement condition was slightly smaller than number of target items in the SA study (Experiment 2),: in the SA study the number of target items per agreement condition was 12, and in this experiment it was 8.

The subjects of the contextual sentence for the $O_{cl}$VS sentences (in the example above: *La camarera y los cocineros conversan en el corredor*. "The waitress and the cooks chat in the halleway") always ended in either: *–a*, fem.; *–o*, masc.; *–ora*, fem.; *–or*, masc. That is, gender affixes agreed with the biological sex of their referent (i.e., they conveyed semantic gender, Corbett, 1988).

The antecedents usually referred to professions: *doctora,*"doctor" *maestro*, "teacher", or nationalities, *argentino,* "Argentinian," *chino*, "Chinese." The length of the names of professions in the contextual sentence did not exceed four syllables. The subjects and direct object clitic pronouns in the target sentences were both animate, and were either masculine or feminine, but they always had different genders. As argued by Sanz (1997, p. 54), counterbalancing the gender of the subject and the clitic direct object pronoun in the target sentence allows the researcher to know whether the learner is mapping the clitic as the subject or as the object. This means that, if the direct object-clitic is masculine and the subject is feminine, the correct response to a question targeting the agent of the sentence will necessarily be in the feminine form. By contrast, if learners select an NP in the masculine form as the agent, then this selection serves as evidence that they are mapping the clitic as the agent of the sentence. In addition, counterbalancing gender reduces ambiguity in the interpretation, since the clitic unequivocally refers to a masculine or feminine referent. Therefore, in this experiment, direct object clitic pronouns referred to one of the subjects in the contextual sentences (e.g., in *La camarera y los cocineros conversan en el corredor*. "The waitress and the cooks chat in the hallway," the clitic *la*, "her," in the target sentence refers to the first subject (NP1) of the contextual sentence). The gender of the subjects in the contextual sentence was counterbalanced (i.e., masc.-fem; fem-masc.) to guarantee that the clitic and the gender of the subject in the target sentences were also counterbalanced. Whether the clitic referred to the first subject (NP1) or the second subject (NP2) was also counterbalanced. Likewise, an adverb (e.g., *a veces*, "sometimes") began the $O_{cl}$VS sentence to avoid

placing the clitic in the position of the first word of the sentence since there is a tendency for readers to skip over the first word of a sentence (e.g., Rayner, 1998).

The length of all $O_{cl}$VS sentences was controlled for; that is, each sentence conveyed a clitic, a verb in simple present, a subject, and a prepositional phrase (PP) of the same length (e.g., *en la casa*, "in the house"). Sentences were composed of vocabulary items taken from textbooks used in lower level courses (e.g., *Vistazos*). Following Sagarra (2007), the vocabulary frequency was determined by the content covered in the textbooks, which contain lexical lists at the end of each chapter. In addition, the frequency of verbs and the nouns was checked using LEXESP (*Léxico informatizado del español*) (Sebastián Gallés, Antonín, Carreira Valiña, & Cuetos Vega, 2000). Verbs never started with s- to avoid synalephas with plural clitics, and the noun of the prepositional phrase never repeated the same gender of the post-verbal subject: for example, if the gender of the post-verbal subject was masculine (*el cocinero*, "the cook"), therefore, the noun in the prepositional phrase was feminine (*con historias*, "with stories").

Finally, each $O_{cl}$VS sentence was followed by a True/False question that probed participants' interpretation of the previous sentence by either conveying a correct agent assignment or an incorrect one (See Table 13).

***Fillers***

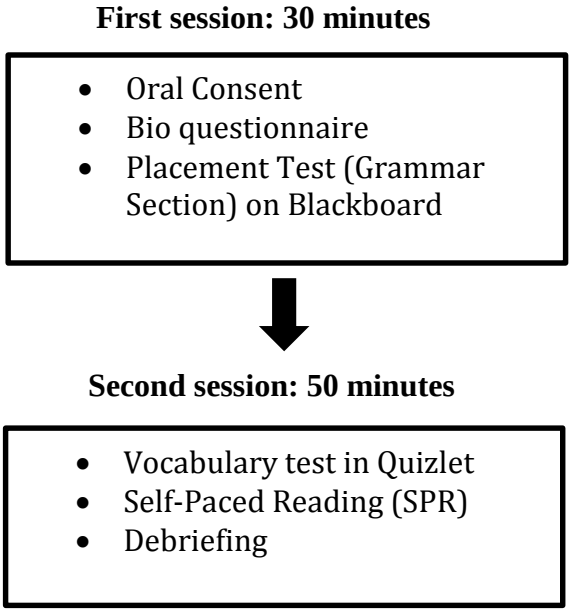
Fillers consisted of sentences conveying the verb *ser* (used to describe general charactistics of a noun, for example, "*ser trabajador"*, "to be hardworking'); *estar* (used to describe temporary states, for example, "*estar triste*," "to be sad"); and *tener* (in expression such as "*tener hambre*","to be hungry," "*tener sueño*," "to be sleepy") in the

simple present. These are verbs that are usually practiced together in textbook excercises and in the classroom because their use tends to be difficult for learners, especially the verbs *ser* and *estar* (to be). Most of the T/F comprehension questions for *ser* and *estar* sentences conveyed an antonym or indicated the opposite of what was conveyed in the sentence.This pattern in the fillers, as well as the inclusion of a contextual sentence for filler items, are two of the major improvements in this study from Experiment 1. The reason for these changes was to ensure that fillers were actual distractors. Additionally, fillers were created so as to be engaging and related to student's lives or to every-day topics. The subjects of the fillers were either animate or inanimate. When they were animate, they conveyed professions, nationalities, and a few proper names. Finally, the length of the fillers as well as the frequency of the vocabulary employed were controlled for. The vocabulary used in the fillers was extracted from the students' textbook for the lower levels, *Vistazos*. If a word that was not listed in this textbook, it was a cognate of a word from the learners' L1 (e.g., *esporádico*, sporadic).

**Procedures**

This experiment took place in a silent room (Linguistic Laboratory) where no external noises or distractions could interfere with learner performance on the SPR task. A maximum of three participants were present at the same time in the laboratory. The study consisted of two sessions: the first session lasted approximately 30 minutes, and the second session lasted approximately 50 minutes. See Figure 13:

**Figure 13: Description of two experimental sessions**




As seen in Figure 13, after giving their oral consent to the researcher, participants completed a bio questionnaire (Appendix B) and the Placement Test (grammar section) (Appendix A). In a second session, students completed a Vocabulary Quiz comprising the verbs, professions, nationalities, and a few prepositional phrases in the target sentences and fillers, so as to guarantee students' familiarity with the vocabulary. Following Sagarra (2007), ensuring that learners know the form and meaning connections of the vocabulary conveyed in the sentences helps to "avoid latencies in the reading times due to unfamiliarity with the experimental stimuli" (p. 139). This vocabulary quiz (*k*= 123) was administered with Quizlet, a free online software that creates reversible electronic vocabulary flashcards in Spanish and in English (see Appendix E). Participants were instructed to look at the word in Spanish and translate the word into English; if they did not know the word, they were instructed to flip the card and read it in English; they could also ask the researcher questions about words they still did not understand, as needed.

Additionally, learners completed a questionnaire in which they were asked to (a) self-assess the difficulty of the vocabulary they just saw in the Quizlet, (b) write down the words they did not know or did not remember well on the answer sheet (Quizlet allows learners to click on a star icon to mark the cards (words) they do not know/remember, so they can retrieve them very easily) (Appendix F). This information was valuable for the researcher, since it allowed to address how learners perceived the vocabulary and also to check which words presented more difficulties. This was also a major improvement in the procedures in comparison to Experiment 1.

The answer sheet that participants completed after taking the vocabulary quiz yielded that: 11 participants (12 %) said that vocabulary was "very easy"; 57 participants (63 %) said the vocabulary was "easy"; and 22 participants said that the difficult of the vocabulary was "moderate" (22%) The average number of words participants did not know was 9 words (out of 123 tokens). Importantly, most of the words participants were unsure of came from the fillers and consisted of nouns (e.g., *peluquero*, "hairdresser"), prepositional phrases (e.g., *a sus espaldas*, "behind someone's back") or adjectives (e.g., *nocivo, "*harmful*"*). This indicated that, by and large, participants knew the form-meaning connections in the verbs and professions conveyed in the $O_{cl}$VS sentences. None of the participants expressed that the vocabulary was difficult or very difficult. Finally, most of the participants who selected "moderate" as an answer wrote that they just needed "to refresh" some vocabulary that they had studied in the past.

After the vocabulary quiz, the participants then proceeded with the word-by-word Self-Paced Reading Task (SPR)(Just, Carpenter, & Woolley, 1982) in an individual PC. Item blocking and randomization within lists was automatically managed using the

reading-time software, Linger (Doug Rohde, MIT). The order of experimental and filler items was randomized for each participant. In each experimental trial, participants read a sentence with words masked by dashes. When participants pressed the space bar, a word was revealed and the previous word was re-masked. When participants pressed the space bar again, another word was revealed and the previous word was re-masked (following the same procedure discussed in Chapter One).

Six practice sentences (e.g., *Esta mañana José y Alexandra se han levantado de muy mal humor* [contextual sentence]. *Por eso Alexandra trata mal a José por una tontería.* [True / False question] *José trata mal a Alexandra por una tontería*. [False]) were presented to learners before the experiment began in order to ensure that learners understood the task. None of the practice sentence consisted of $O_{cl}$VS sentences.

The contextual sentences that preceded the target $O_{cl}$VS items (e.g., *La camarera y los cocineros conversan en el corredor*. "The waitress and the cooks chat in the hallway") were also masked, except that, unlike the target items, participants were able to read the contextual sentence all at once since the goal of the SPR was to measure the RTs of the regions of interest in the target sentences and not in the contextual sentence. This contextual sentence disappeared as soon as the participants pressed the space bar so as to avoid participants' regressing to the contextual sentence; the experiment sought to measure the RTs on each of the words of the target sentences, amongst which were the region of interest (e.g., verb region). After the contextual sentence disappeared, the target $O_{cl}$VS sentences were revealed word-by-word. Participants were instructed to read at a natural pace and answer the questions as quickly and accurately as possible by pressing one of two keys ("F" key for True or the "J" key for False) (the selection of these keys

followed previous work in psycholinguistics, e.g., Wagers, Lau, & Phillips, 2009). Between each block, participants were given three breaks; they were also informed that a break could be taken at any point of the experiment.

The [+ Feedback] group received yes/no feedback every time they answered the T/F questions on both the target items and the fillers: the message generated by the computer was in the target language so as to prevent switching from the L2 to the L1. Following Sanz and Morgan-Short (2004), when participants were correct, the message that appeared on the screen was (*Bien!* "Good!), and when participants made a mistake, the error message said: (*Lo siento!* "Sorry"). The [- Feedback] group did not receive feedback on any accuracy responses.

**Analyses**

**Accuracy data**

For both experimental groups, the statistical analysis of the correct responses for the T/F comprehension questions targeting the agent for the $O_{cl}$VS sentences was carried out using a binomial logistic regression in R (R Development Core Team, 2014). The predictor variables were Condition and Trial order, and the dependent variable was the accuracy scores. Each agreement condition between the clitic and the verb (Condition A, C and D) was compared against Condition B (cl-sing, verb-sing.), which served as the baseline (i.e., the intercept). That is, the analysis was carried out first for Condition B, and then the other conditions were analyzed so as to see if and how those conditions differed from the Condition B baseline. The coefficients in the logistic regressions conducted were standardized in order to allow for direct comparisons with other similar studies (See Norris & Ortega, 2000 with regard to standardization).

**Latency data**

Following the methodological procedure of recent studies conducted in the field of psycholinguistics (Hofmeister, 2011; Lago et al., 2015), only residual reaction times (i.e., residual RTs) were entered into statistical analysis. Residual RTs were calculated in the following way: a linear model used in R first predicted the expected RTs for each word in function of its length. This model was a linear regression based on the following formula: $y=a+bx$, where *y* stands for RTs, *a* stands for the intercept, *b* stands for coefficient (slope), and *x* stands for word length (number of characters in a word) (cf. Hofmeister, 2011). Second, the linear model computed the residual RTs by subtracting (i.e., "regressing") the expected RTs that the model predicted from the raw (observed) RT data obtained.). Residual RTs were statistically estimated from the entire dataset (target items and fillers) for all participants.. This was done in order to control for differences in length between the plural and singular forms of the clitic and the verb in the different experimental conditions: in Spanish, the suffix in the clitic in the 3rd person plural (-*s*) is one character longer than its counterpart in the singular; and the suffix in the verb in the 3rd person plural for the present tense (-*n*) is also one character longer than the 3rd person singular. Residual RT averages more than ± 2.5 SD from the participant mean by region and condition were excluded (Ratcliff, 1993).

**Statistical procedures**

The statistical analysis for accuracy data, the correct responses for the T/F comprehension questions targeting the agent for the $O_{cl}$VS sentences, was carried out using a binomial logistic regression in R (R Development Core Team, 2014). Two independent logistic regressions were conducted on the accuracy data (dependent

variable) for each of the experimental groups [+/- Feedback], with Condition and Trial Order as fixed factors (independent variables). In addition, a logistic regression was conducted on the aggregated accuracy data of both groups with Condition, Trial Order, and Feedback vs. No Feedback as fixed factors. The analysis for Condition was carried out first for Condition B (the baseline); the other conditions were then analyzed as to how those other conditions differed from Condition B.

The latency data were analyzed using the *lmer* package in R. A series of general linear models (GLMs) were run on the residual RTs in each of the sentence regions (words) in order to calculate main effects and potential interactions. On the residual RT data of the [- Feedback] and [+ Feedback] group, Condition and Trial Order were the fixed factors of the GLMs conducted on each group independently. On the aggregated residual RT data, Condition, Trial Order and Feedback vs. No Feedback were the fixed factors of the GLMs conducted on both groups. RT averages more than 2.5 standard deviations from the participant mean by region and condition were excluded (Ratcliff, 1993).

## Results

### Accuracy in $O_{cl}$VS sentences

Table 20 shows the mean accuracy overall on the T/F questions in the pre-test and post-test in each of the $O_{cl}$VS experimental conditions. The T/F questions measured whether learners could identify the agent of $O_{cl}$VS sentences; therefore, they indicate the percentage of learners' correct form and meaning mappings for $O_{cl}$VS sentences.

**Table 20: Overall accuracy for $O_{cl}$VS sentences with feedback and no-feedback**

| | Condition | Mean | *SD* | Lower CI | Upper CI |
|---|---|---|---|---|---|
| No Feedback | A | 0.32 | 0.01 | 0.30 | 0.34 |
| **Feedback** | A | **0.63** | 0.01 | 0.61 | 0.65 |
| No Feedback | B | 0.30 | 0.01 | 0.28 | 0.32 |
| **Feedback** | B | **0.64** | 0.01 | 0.62 | 0.66 |
| No Feedback | C | 0.25 | 0.01 | 0.23 | 0.27 |
| **Feedback** | C | **0.59** | 0.01 | 0.57 | 0.61 |
| No Feedback | D | 0.23 | 0.01 | 0.22 | 0.25 |
| **Feedback** | D | **0.59** | 0.01 | 0.57 | 0.61 |

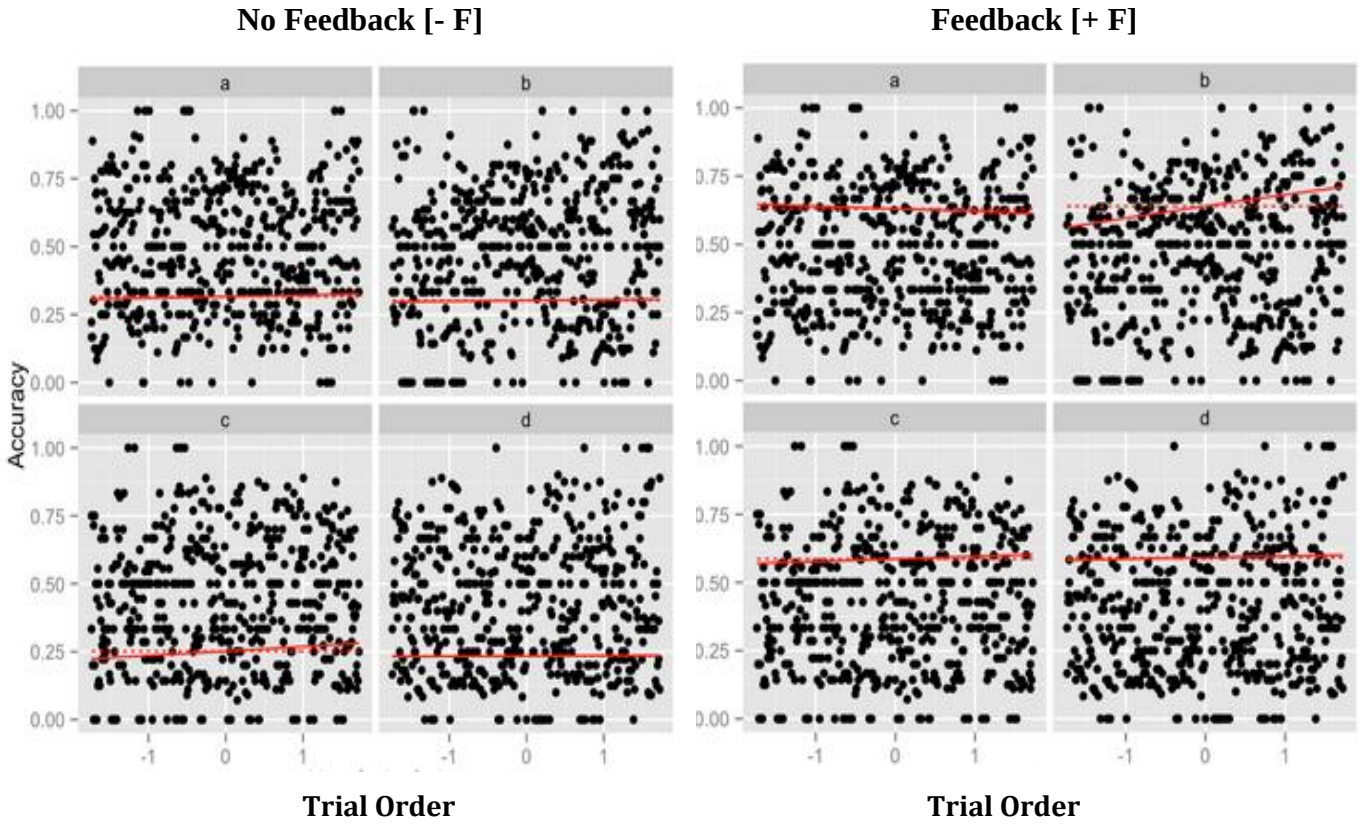
As seen in Table 20, in the [- Feedback] group, the overall accuracy for all conditions was right at or below 30% (Min: 23% for Condition D, cl.pl., verb sing.; Max: 32% for Condition A, cl. sing, verb-pl.). The [+ Feedback] group clearly outperformed the no feedback group, showing an improvement of 100% or greater in all experimental conditions: Min: 59% for both Condition C (cl. pl., verb pl.) and Condition D (cl. pl., verb sing); Max: 64% for Condition B (the baseline, cl. sing., verb. sing.); Condition A (cl. sing, verb-pl.) received a similar overall accuracy of 63%.

Results are consistent with the other two experiments conducted in this dissertation: Experiment 1 and SA study (Experiment 2). Therefore, the overall accuracy of this experiment provides additional empirical evidence supporting the idea that, even at an advanced level of proficiency, $O_{cl}$VS sentences are difficult to process for emerging bilinguals. To remind the reader, in Experiment 1, in which all participants received feedback on their responses, the overall accuracy for $O_{cl}$VS sentences conveying the auxiliary verb *estar*, "to be" as the main verb was between 56% and 61%, a range consistent with the results obtained in this study. In addition, for the participants in the

SA study (Experiment 2), an experiment for which the participants did not receive feedback on their responses, the overall accuracy in the *Pre-test* was between 35% and 43%. These figures are also consisted with the present study in that both the SA group and the no-feedback group which were reported on here exhibited an accuracy performance below 50%, even though the SA accuracy scores in the pre-test were slightly better than the ones obtained here for learners in the [- Feedback] group.

Figure 14 represents learners' estimated accuracy across the experimental trials in the [- Feedback] and [+ Feedback] groups, No Feedback and Feedback respectively.

**Figure 14: $O_{cl}VS$ accuracy over the course of the experiment in the No Feedback and Feedback groups**




As seen in Figure 14, the differences in accuracy rates between the [- Feedback] group and the [+ Feedback] group are notable. Condition B (the baseline) in the [+

Feedback] is the only condition where there was an improvement across experimental trials.

A logistic regression was conducted independently for each experimental group, [- Feedback] and [+ Feedback], to examine how participants' performance changed as the experiment progressed. Results are shown in Table 21 and Table 22.

**Table 21: Logistic regression for $O_{cl}VS$ sentences in the [- Feedback] group**

| | $\beta$ | St. E | z value | Pr(>\|t\|) |
|---|---|---|---|---|
| Condition B (Intercept) | -0.8379 | 0.0482 | -17.37 | **0.0000*** ** |
| Trial order std. | 0.0143 | 0.0492 | 0.29 | 0.7718 |
| Condition A vs. Condition B | **0.0661** | 0.0684 | 0.97 | 0.3339 |
| **Condition C vs. Condition B** | -0.2514 | 0.0711 | -3.53 | **0.0004*** ** |
| **Condition D vs. Condition B** | -0.3474 | 0.0714 | -4.87 | **0.0000*** ** |
| Trial Order Std. × (Condition A vs. Condition B) | 0.0120 | 0.0687 | 0.17 | 0.8618 |
| Trial Order Std. × (Condition C vs. Condition B) | 0.0734 | 0.0729 | 1.01 | 0.3142 |
| Trial Order Std. × (Condition D vs. Condition B) | -0.0068 | 0.0707 | -0.10 | 0.9230 |

**Table 22: Logistic regression for $O_{cl}VS$ sentences in the [+ Feedback] group**

| | $\beta\hat{}$ | St. E | z value | Pr(>\|t\|) |
|---|---|---|---|---|
| Condition B | 0.5745 | 0.0435 | 13.22 | **0.0000*** ** |
| Trial order std. | 0.1867 | 0.0450 | 4.15 | **0.0000*** ** |
| Condition A vs. Condition B | -0.0424 | 0.0606 | -0.70 | 0.4841 |
| **Condition C vs. Condition B** | -0.2262 | 0.0601 | -3.76 | **0.0002*** ** |
| **Condition D vs. Condition B** | -0.2055 | 0.0604 | -3.40 | **0.0007*** ** |
| Trial Order Std. × (Condition A vs. Condition B) | -0.2283 | 0.0623 | -3.66 | **0.0002*** ** |
| Trial Order Std. × (Condition C vs. Condition B) | -0.1477 | 0.0604 | -2.45 | **0.0145*** |
| Trial Order Std. × (Condition D vs. Condition B) | -0.1642 | 0.0611 | -2.69 | **0.0071**** |

***No Feedback group***

As seen in Table 21, for the [- Feedback] group, the coefficient obtained in Condition B (the baseline, cl. sing., verb sing.) shows that the participants in this group significantly performed *below* chance (($\hat{\beta}$= -0.8379, z= -17.37, *p*= 0.0000). By transforming this coefficient into probabilities, following the formula: exp ($\hat{\beta}$)/ (1+ exp ( $\hat{\beta}$ )), it was found that the probability for learners in the [- Feedback] group's performing correctly in Condition B was 30%.

In this [- Feedback] group, participants performed slightly better in Condition A (cl. sing., verb pl.) ($\hat{\beta}$= 0.0661, z= 0.97, *p*= 0.3339) than in Condition B (the baseline, cl. sing., verb sing.). By converting this coefficient into probabilities, it was obtained that the probability of learners in the [- Feedback] group's performing accurately in Condition A was 51%. This difference is numerically higher although it was not significant.

In the other two conditions, Condition C (cl. pl., verb pl.) and Condition D (cl. pl., verb sing), the participants in the [- Feedback] group performed significantly worse than in Condition B, the baseline ($\hat{\beta}$= -0.2514, z= -3.53, *p*= 0.0004 for Condition C and $\hat{\beta}$= -0.3474, z= -4.87, *p*= 0.0000 for Condition D).

Additionally, no order effects were found across experimental trials for accuracy in the [- Feedback] group, which means that learners did not improve over the course of the experiment in any of the four experimental conditions.

Further comparisons may be noted here between the results obtained in the [- Feedback] group in this study and those obtained in the *pre-test* of the SA study [Experiment 2], where feedback was also not provided. In both studies, Condition A (cl. sing, verb pl.) was the condition in which learners showed better overall performance,

although this difference was not significant. The main difference between these two studies is that whereas in the pre-test, learners in the SA study *did* improve across experimental trials in Condition A (cl. sing., verb pl.), learners in the [- Feedback] group in this chapter's study did not. This suggests that learners' characteristics in the two groups (the [- Feedback] group and the SA learners) were somehow different; this brings attention to the role individual differences might have on morphosyntactic development.

***Feedback group***

With regard to the [+ Feedback group], Table 22 shows that Condition B (cl. sing., verb sing.), the baseline, was the condition in which learners performed the best ($\hat{\beta}$= 0.5745, z= 13.22, *p*= 0.0000). Even though this study predicted that Condition A was going to exhibit higher accuracy scores overall, the result that Condition B was more accurate than Condition A, while not predicted, is not completely unexpected either in terms of *overall* accuracy, since, in the previous experiments (Experiment 1) and the SA study (Experiment 2), Condition B (cl. sing., verb. sing.) has closely followed Condition A in mean accuracy, and no study in this dissertation has yielded that the overall accuracy in Condition A was significantly different from Condition B (the baseline). In addition, in the SA study, Condition B showed significant improvements in the post-test. With regard to the [- Feedback] group in the present study, the performance in Condition B for the [+ Feedback] group was above chance: the probability of a learners' being accurate in Condition B was 64%.

Learners' performance in Condition A (cl. sing., verb pl.), on the other hand, was not as strong in comparison with Condition B ($\hat{\beta}$= -0.0424, z= -0.70, *p*= 0.4841). The probability of getting the sentences in this condition correctly was 49%. However, the

differences in performance between Condition A and Condition B were not significant. By contrast, learners did perform significantly worse in Condition C and Condition D than in Condition B, the baseline, ($\hat{\beta}$= -0.2262, z= -3.76, *p*= 0.0002 for Condition C and $\hat{\beta}$= -0.2055, z= -3.40, *p*= 0.0007 for Condition D). The probability of coming to the right interpretation of $O_{cl}$VS sentences in Condition C and D, in both cases, was around 44%. Lastly, unlike the [-Feedback] group, the [+ Feedback] group significantly improved across experimental trials, since an order effect was found for Condition B ($\hat{\beta}$= 0.1867, z= 4.15, *p*= 0.0000). This means that participants in the [+ Feedback] group not only improved in accuracy but also improved faster in Condition B.

Considered together: the improvement in overall accuracy in each of the experimental conditions and the rate of improvement across experimental trials in Condition B (cl. sing., verb sing.), indicates that feedback seems to make a difference in how accurately learners interpret $O_{cl}$VS sentences in Spanish.

***Aggregated accuracy data from the No Feedback and Feedback groups***

A final step was taken in the statistical analysis to further compare the accuracy data in both experimental groups. A logistic regression was conducted on the aggregated accuracy data of both groups, with [+/- Feedback], Trial order and Condition as fixed factors. Feedback was set as the sum contrast, with effects in the direction of Feedback minus No Feedback. The results are shown in Table 23

**Table 23: Logistic regression for $O_{cl}$VS sentences in the No Feedback and Feedback groups (aggregated accuracy data)**

| | $\hat{\beta}$ | St. E | z value | Pr(>\|t\|) |
|---|---|---|---|---|
| Condition B | -0.1317 | 0.0325 | -4.06 | 0.0000 *** |
| **Trial order std.** | 0.1005 | 0.0333 | 3.01 | **0.0026 **** |

| | | | | |
|---|---|---|---|---|
| Condition A vs. Condition B | 0.0118 | 0.0457 | 0.26 | 0.7959 |
| Condition C vs. Condition B | -0.2388 | 0.0465 | -5.13 | **0.0000** *** |
| Condition D vs. Condition B | -0.2765 | 0.0467 | -5.92 | **0.0000** *** |
| **Feedback vs. No Feedback** | 0.7062 | 0.0325 | 21.76 | **0.0000** *** |
| **Trial Order Std. × (Condition A vs. Condition B)** | -0.1082 | 0.0464 | -2.33 | **0.0197 *** |
| Trial Order Std. × (Condition C vs. Condition B) | -0.0371 | 0.0473 | -0.78 | 0.4329 |
| Trial Order Std. × (Condition D vs. Condition B) | -0.0855 | 0.0467 | -1.83 | 0.0669 |
| **Trial Order × (Feedback vs. No Feedback)** | 0.0862 | 0.0333 | 2.59 | **0.0097**** |
| (Condition A vs. Condition B) × (Feedback vs. No Feedback) | -0.0542 | 0.0457 | -1.19 | 0.2351 |
| (Condition C vs. Condition B)× (Feedback vs. No Feedback) | 0.0126 | 0.0465 | 0.27 | 0.7868 |
| (Condition D vs. Condition B) × (Feedback vs. No Feedback) | 0.0710 | 0.0467 | 1.52 | 0.1290 |
| **Trial order std. × (Condition A vs. Condition B) × (Feedback vs. No Feedback)** | -0.1201 | 0.0464 | -2.59 | **0.0096**** |
| Trial order std. × (Condition C vs. Condition B) × (Feedback vs. No Feedback) | -0.1106 | 0.0473 | -2.33 | 0.0195* |
| Trial order std. × (Condition D vs. Condition B) × (Feedback vs. No Feedback) | -0.0787 | 0.0467 | -1.69 | 0.0918 |

*Signif. codes*: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

***Note*: Coefficients ($\beta$)'s operationalizations:**

- **Trial order:** accuracy improvement across experimental trial units (i.e., $O_{cl}VS$ sentences).
- **Condition A vs. Condition B:** $O_{cl}VS$ accuracy overall of Condition A against of the $O_{cl}VS$ accuracy overall of condition B (the baseline).
- **Condition C vs. Condition B:** $O_{cl}VS$ accuracy overall of Condition C against of the $O_{cl}VS$ accuracy overall of condition B (the baseline).
- **Condition D vs. Condition B:** $O_{cl}VS$ accuracy overall of Condition D against of the $O_{cl}VS$ accuracy overall of condition B (the baseline).
- **No Feedback vs. Feedback:** Differences in $O_{cl}VS$ accuracy in Condition B between Feedback vs. No Feedback, with effects in the direction of [+ Feedback] minus [- Feedback].
- **(Trial order) × (Conditions A vs. Condition B):** Experimental trial order effects in Condition A with respect to Condition B (the baseline) in the experiment *overall*.
- **(Trial order) × (Conditions C vs. Condition B):** Experimental trial order effects in Condition C with respect to Condition B (the baseline) in the experiment *overall*.
- **(Trial order) × (Conditions D vs. Condition B):** Experimental trial order effects in Condition D with respect to Condition B (the baseline) in the experiment *overall*.

- **(Trial order) × (Condition A vs. Condition B) × (No Feedback vs. Feedback):** Experimental trial order effects in Condition A with respect to Condition B, and in relation to the accuracy differences between No Feedback and the Feedback.
- **Trial order) × (Condition C vs. Condition B) × (No Feedback vs. Feedback):** Experimental trial order effects in Condition C with respect to Condition B, and in relation to the accuracy differences between No Feedback and the Feedback.
- **Trial order) × (Condition D vs. Condition B) × (No Feedback vs. Feedback):** Experimental trial order effects in Condition D with respect to Condition B, and in relation to the accuracy differences between No Feedback and the Feedback.

As seen in Table 23, the results of the logistic regression for the aggregated data confirmed the differences discussed in the previous section. First, the [+ Feedback] group significantly outperformed the [- Feedback] group in Condition B ($\hat{\beta}$= 0.7062, z= 21.76, $p$= 0.0000), as mentioned above. Whereas the probability of being accurate in Condition B (cl. sing., verb sing.) for the [- Feedback] group was 30%, in the [+ Feedback] group, it was 64%.

Second, Table 23 shows that in the aggregated data, the *speed* with which learners improved across experimental trials for Condition B in the [+ Feedback] group was significant; this means that learners improved faster in Condition B (the baseline, cl. sing., verb sing.) than they did in the [- Feedback] group ($\hat{\beta}$= 0.0862, z= 2.59, $p$= 0.0097).

Third, Condition A (cl. sing., verb pl.) was the only condition that was not significantly different from Condition B in the accuracy rate overall. In fact, Condition A was slightly better than Condition B overall when the accuracy data of both groups were aggregated; however, this difference was not significant ($\hat{\beta}$= 0.0118, z= 0.26, $p$= 0.7959).

Interestingly, a common pattern that emerged from the aggregated data is that for both groups, Condition B (cl. sing., verb sing.) and Condition A (cl. sing. verb pl.) are the two conditions in which participants improved the most overall. Accuracy in Condition C

(cl. pl., verb. pl.) and D (cl. pl., verb sing.) was significantly worse than in Condition B in both [- Feedback] and [+ Feedback] groups.

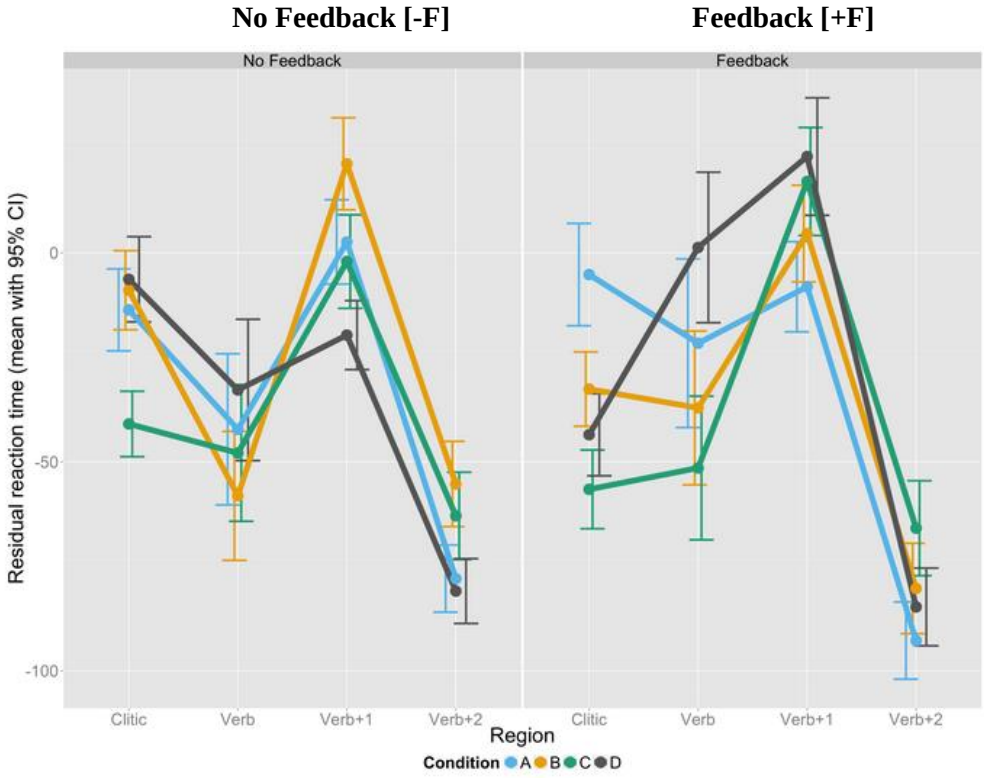
Lastly, as to the rate of improvement in Condition A (cl. sing, verb pl.), Table 23 showed that when participants received feedback they did not improve as quickly in this condition as they did in Condition B ($\hat{\beta}$= -0.1201, z= 2.59, $p$= 0.0096). This suggests that Condition A might have been more cognitively taxing that Condition B.

**Latency in $O_{cl}$VS sentences**

The Reaction Time (RT) data in the present experiment was analyzed in order to address differences in processing costs in each experimental group and agreement conditions.

Figure 15 illustrates the mean average for the residual Reaction Times (RTs) in milliseconds (ms) for each of the regions (words) in the $O_{cl}$VS sentences in all conditions and experimental groups

**Figure 15: Mean residual RTs overall for all conditions in the No Feedback and Feedback groups**




Visually, Figure 15 shows that learners in general tended to spend more time in the clitic region in the [-Feedback] group than in the [+ Feedback] group, which only spent more time in the clitic region of Condition A (cl. sing., verb pl.).

With regard to the verb region, Figure 15 shows that the [- Feedback] group exhibited a very similar pattern to that which was found in the group of learners in the pre-test of the SA study (Experiment 2). In the sentence around the verb region, both patterns resemble a V shape, with residual RTs lowering in the verb region itself. This suggests that learners who did not receive feedback on their responses tended to spend less time on the verb region. The [+ Feedback] group, on the other hand, showed a tendency to spend more time in the verb region than the [- Feedback] group, especially in Condition D (cl. pl., verb sing.) and Condition A (cl. sing, verb pl.), as predicted. This

partially replicates the pattern found in Experiment 1 for the verb region: the participants in Experiment 1 spent even more time in the verb region than learners in this study. To remind the reader, the processing pattern in the verb region for participants in t Experiment 1 resembled a ∧ shape. The slight differences suggests that verb type seems to also play a role in $O_{cl}$VS sentence processing; whereas in Experiment 1 the main verb tested was the auxiliary verb *estar*, "to be", the main verbs tested here were verbs in the simple present (e.g., *mirar*, "to look").

The residual RT data in each region of interest: clitic region, verb region and both spillover regions were submitted to further statistical analyses to detect potential patterns in learners' processing costs. A total of eight linear models with Condition and Trial order as fixed factors were conducted independently in each of the regions of interest in both experimental groups (- feedback / + feedback). Table 24 provides the results of these linear models.

Additionally, as with the accuracy data, the residual RT data from both groups were analyzed as aggregated data in order to establish further comparisons. Another eight linear models were conducted on the aggregated data for each of the regions of interest, this time with [+/- Feedback], Trial order and Condition as fixed factors. Feedback was set as the sum of contrast, with effects in the direction of Feedback minus No Feedback. The results of the linear models conducted on the combined residual RT data from both experimental groups are shown in Table 25.

**Table 24: Linear models for the clitic region, the verb region, and spillover region (verb + 1) & (verb + 2)**

| | CLITIC REGION NO FEEDBACK | | | | CLITIC REGION FEEDBACK | | | |
|---|---|---|---|---|---|---|---|---|
| | $\widehat{\beta}$ | St. E | *t* value | Pr(>\|t\|) | $\widehat{\beta}$ | St. E | *t* value | Pr(>\|t\|) |
| **Condition B** | -9.0594 | 4.7713 | -1.90 | 0.0576 | -32.2536 | 5.1998 | -6.20 | **0.0000*** ** |
| Trial Order Std | -6.8166 | 4.9276 | -1.38 | 0.1666 | -6.2085 | 5.3364 | -1.16 | 0.2447 |
| **Condition A** | -4.2110 | 6.7595 | -0.62 | 0.5333 | 27.0516 | 7.3472 | 3.68 | **0.0002*** ** |
| **Condition C** | -31.7828 | 6.7672 | -4.70 | **0.0000*** ** | -24.2598 | 7.3571 | -3.30 | **0.0010** ** |
| Condition D | 2.8095 | 6.7482 | 0.42 | 0.6772 | -11.5071 | 7.3622 | -1.56 | 0.1181 |
| **Trial Order Std. × Condition A** | 17.3549 | 6.8483 | 2.53 | **0.0113*** | 8.6396 | 7.5066 | 1.15 | 0.2498 |
| Trial Order Std. × Condition C | 1.3475 | 6.9541 | 0.19 | 0.8464 | 4.7707 | 7.3470 | 0.65 | 0.5161 |
| Trial Order Std. × Condition D | 1.0653 | 6.7286 | 0.16 | 0.8742 | 3.6750 | 7.4097 | 0.50 | 0.6199 |

| | VERB REGION NO FEEDBACK | | | | VERB REGION FEEDBACK | | | |
|---|---|---|---|---|---|---|---|---|
| | $\widehat{\beta}$ | St. E | *t* value | Pr(>\|t\|) | $\widehat{\beta}$ | St. E | *t* value | Pr(>\|t\|) |
| Condition B | -58.2536 | 8.5007 | -6.85 | 0.0000 | -35.9423 | 9.4766 | -3.79 | 0.0001 |
| **Trial Order Std** | -13.2877 | 8.7787 | -1.51 | 0.1301 | -19.7875 | 9.7194 | -2.04 | **0.0418*** |
| Condition A | 16.4797 | 12.0176 | 1.37 | 0.1703 | 14.0992 | 13.3622 | 1.06 | 0.2914 |
| Condition C | 10.5179 | 12.0566 | 0.87 | 0.3830 | -15.7947 | 13.3525 | -1.18 | 0.2369 |
| Condition D | 25.7319 | 12.0652 | 2.13 | **0.0330*** | 36.4600 | 13.3797 | 2.73 | **0.0064** ** |
| Trial Order Std. × Condition A | 25.4853 | 12.1747 | 2.09 | **0.0363*** | 7.8742 | 13.6486 | 0.58 | 0.5640 |
| Trial Order Std. × Condition C | 4.5720 | 12.3936 | 0.37 | 0.7122 | 24.3963 | 13.3344 | 1.83 | 0.0673 |
| Trial Order Std. × Condition D | -2.4611 | 12.0191 | -0.20 | 0.8378 | 11.9090 | 13.4627 | 0.88 | 0.3764 |

Continued (Table 24)

| | VERB + 1 REGION (determiner) NO FEEDBACK | | | | VERB + 1 REGION (determiner) FEEDBACK | | | |
|---|---|---|---|---|---|---|---|---|
| | $\widehat{\beta}$ | **St. E** | ***t* value** | **Pr(>\|t\|)** | $\widehat{\beta}$ | **St. E** | ***t* value** | **Pr(>\|t\|)** |
| Condition B | 21.1675 | 5.1974 | 4.07 | **0.0000** | 5.5209 | 6.3269 | 0.87 | 0.3829 |
| **Trial Order Std** | -8.8647 | 5.3666 | -1.65 | 0.0986 | -15.9787 | 6.4932 | -2.46 | **0.0139*** |
| Condition A | -18.1876 | 7.3529 | -2.47 | **0.0134*** | -13.5395 | 8.9462 | -1.51 | 0.1302 |
| Condition C | -23.1460 | 7.3611 | -3.14 | **0.0017*** | 11.6155 | 8.9519 | 1.30 | 0.1945 |
| Condition D | -40.8145 | 7.3456 | -5.56 | **0.0000***** | 16.4161 | 8.9581 | 1.83 | 0.0669 |
| Trial Order Std. × Condition A | 19.4224 | 7.4491 | 2.61 | **0.0091**** | 23.9017 | 9.1384 | 2.62 | **0.0089**** |
| Trial Order Std. × Condition C | 1.8524 | 7.5632 | 0.24 | 0.8065 | 14.0059 | 8.9406 | 1.57 | 0.1172 |
| Trial Order Std. × Condition D | 5.4421 | 7.3239 | 0.74 | 0.4575 | 4.6594 | 9.0159 | 0.52 | 0.6053 |

| | VERB + 2 REGION (noun) NO FEEDBACK | | | | VERB + 2 REGION (noun) FEEDBACK | | | |
|---|---|---|---|---|---|---|---|---|
| | $\widehat{\beta}$ | **St. E** | ***t* value** | **Pr(>\|t\|)** | $\widehat{\beta}$ | **St. E** | ***t* value** | **Pr(>\|t\|)** |
| Condition B | -55.3629 | 4.6709 | -11.85 | 0.0000 | -79.5875 | 5.2118 | -15.27 | 0.0000 |
| **Trial Order Std** | -5.4207 | 4.8238 | -1.12 | **0.0263*** | -11.8850 | 5.3488 | -2.22 | **0.0263*** |
| Condition A | -22.3426 | 6.6126 | -3.38 | 0.0753 | -13.1003 | 7.3643 | -1.78 | 0.0753 |
| Condition C | -7.2500 | 6.6292 | -1.09 | **0.0497*** | **14.4864** | **7.3793** | **1.96** | **0.0497** |
| Condition D | -25.5568 | **6.6060** | **-3.87** | 0.4813 | -5.1963 | 7.3793 | -0.70 | 0.4813 |
| Trial Order Std. × Condition A | 11.5846 | 6.6993 | 1.73 | **0.0321*** | 16.1260 | 7.5241 | 2.14 | **0.0321*** |
| Trial Order Std. × Condition C | -11.8421 | 6.8129 | -1.74 | 0.7464 | -2.3831 | 7.3684 | -0.32 | 0.7464 |
| Trial Order Std. × Condition D | 5.4308 | 6.5868 | 0.82 | 0.1359 | 11.0770 | 7.4269 | 1.49 | 0.1359 |

**Table 25: Linear models for the Residual RT data in the No Feedback and Feedback groups (aggregated residual RT data)**

| **Clitic Region** | $\hat{\beta}$ | St. E | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| Condition B | -20.6565 | 3.5286 | -5.85 | 0.0000 |
| Trial order std. | -6.5125 | 3.6327 | -1.79 | 0.0730 |
| **Condition A** | 11.4203 | 4.9924 | 2.29 | **0.0222*** |
| Condition C | -28.0213 | 4.9985 | -5.61 | 0.0000*** |
| Condition D | -4.3488 | 4.9933 | -0.87 | 0.3838 |
| **No Feedback vs. Feedback** | -11.5971 | 3.5286 | -3.29 | **0.0010**** |
| Trial Order Std. × Condition A | 12.9972 | 5.0793 | 2.56 | 0.0105 |
| Trial Order Std. × Condition C | 3.0591 | 5.0647 | 0.60 | 0.5458 |
| Trial Order Std. × Condition D | 2.3702 | 5.0022 | 0.47 | 0.6356 |
| Feedback vs. No Feedback | 0.3041 | 3.6327 | 0.08 | 0.9333 |
| Condition A × **(**Feedback vs. No Feedback) | 15.6313 | 4.9924 | 3.13 | 0.0017 |
| Condition C × **(**No Feedback vs. Feedback) | 3.7615 | 4.9985 | 0.75 | 0.4517 |
| Condition D × **(**No Feedback vs. Feedback) | -7.1583 | 4.9933 | -1.43 | 0.1517 |
| Trial order std. Condition A × (No Feedback vs. Feedback) | -4.3577 | 5.0793 | -0.86 | 0.3909 |
| Trial order std. Condition C × (No Feedback vs. No Feedback) | 1.7116 | 5.0647 | 0.34 | 0.7354 |
| Trial order std. Condition D × (No Feedback vs. Feedback) | 1.3049 | 5.0022 | 0.26 | 0.7942 |

| **Verb Region** | $\hat{\beta}$ | St. E | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| Condition B | -47.0979 | 6.3633 | -7.40 | 0.0000 |
| **Trial order std.** | -16.5376 | 6.5488 | -2.53 | **0.0116 *** |
| Condition A | 15.2894 | 8.9842 | 1.70 | 0.0888 |
| Condition C | -2.6384 | 8.9954 | -0.29 | 0.7693 |
| **Condition D** | 31.0959 | 9.0078 | 3.45 | **0.0006 ***** |
| No Feedback vs. Feedback | 11.1556 | 6.3633 | 1.75 | 0.0796 |
| Trial Order Std. × Condition A | 16.6798 | 9.1394 | 1.83 | 0.0680 |
| Trial Order Std. × Condition C | 14.4841 | 9.1158 | 1.59 | 0.1121 |
| Trial Order Std. × Condition D | 4.7240 | 9.0187 | 0.52 | 0.6004 |
| Feedback vs. No Feedback | -3.2499 | 6.5488 | -0.50 | 0.6197 |
| Condition A × **(**Feedback vs. No Feedback) | -1.1903 | 8.9842 | -0.13 | 0.8946 |
| Condition C × **(**No Feedback vs. Feedback) | -13.1563 | 8.9954 | -1.46 | 0.1436 |
| Condition D × **(**No Feedback vs. Feedback) | 5.3640 | 9.0078 | 0.60 | 0.5515 |
| Trial order std. Condition A × (No Feedback vs. Feedback) | -8.8056 | 9.1394 | -0.96 | 0.3353 |
| Trial order std. Condition C × (No Feedback vs. No Feedback) | 9.9122 | 9.1158 | 1.09 | 0.2769 |
| Trial order std. Condition D × (No Feedback vs. Feedback) | 7.1851 | 9.0187 | 0.80 | 0.4256 |

(Continued Table 25)

| **Verb + 1 Region** | $\hat{\beta}$ | St. E | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| Condition B | 13.3442 | 4.0942 | 3.26 | 0.0011 |
| Trial order std. | -12.4217 | 4.2147 | -2.95 | **0.0032 **** |
| Condition A | -15.8635 | 5.7907 | -2.74 | **0.0062**** |
| Condition C | -5.7653 | 5.7958 | -0.99 | 0.3199 |
| Condition D | -12.1992 | 5.7917 | -2.11 | **0.0352*** |
| No Feedback vs. Feedback | -7.8233 | 4.0942 | -1.91 | 0.0560 |
| Trial Order Std. × Condition A | 21.6621 | 5.8909 | 3.68 | 0.0002 |
| Trial Order Std. × Condition C | 7.9292 | 5.8723 | 1.35 | 0.1769 |
| Trial Order Std. × Condition D | 5.0507 | 5.8019 | 0.87 | 0.3840 |
| Feedback vs. No Feedback | -3.5570 | 4.2147 | -0.84 | 0.3987 |
| Condition A × **(**Feedback vs. No Feedback) | 2.3241 | 5.7907 | 0.40 | 0.6882 |
| Condition C × **(**No Feedback vs. Feedback) | 17.3808 | 5.7958 | 3.00 | **0.0027**** |
| Condition D × **(**No Feedback vs. Feedback) | 28.6153 | 5.7917 | 4.94 | **0.0000 ***** |
| Trial order std. Condition A × (No Feedback vs. Feedback) | 2.2396 | 5.8909 | 0.38 | 0.7038 |
| Trial order std. Condition C × (No Feedback vs. No Feedback) | 6.0768 | 5.8723 | 1.03 | 0.3008 |
| Trial order std. Condition D × (No Feedback vs. Feedback) | -0.3913 | 5.8019 | -0.07 | 0.9462 |

| **Verb + 2 Region** | $\hat{\beta}$ | St. E | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| Condition B | -67.4752 | 3.4992 | -19.28 | 0.0000 |
| Trial order std. | -8.6528 | 3.6025 | -2.40 | 0.0163 |
| Condition A | -17.7214 | 4.9491 | -3.58 | 0.0003 |
| Condition C | 3.6182 | 4.9604 | 0.73 | 0.4657 |
| Condition D | -15.3765 | 4.9517 | -3.11 | 0.0019 |
| No Feedback vs. Feedback | -12.1123 | 3.4992 | -3.46 | 0.0005 |
| Trial Order Std. × Condition A | 13.8553 | 5.0353 | 2.75 | 0.0059 |
| Trial Order Std. × Condition C | -7.1126 | 5.0260 | -1.42 | 0.1570 |
| Trial Order Std. × Condition D | 8.2539 | 4.9606 | 1.66 | 0.0961 |
| Feedback vs. No Feedback | -3.2321 | 3.6025 | -0.90 | 0.3696 |
| Condition A × **(**Feedback vs. No Feedback) | 4.6211 | 4.9491 | 0.93 | 0.3505 |
| Condition C × **(**No Feedback vs. Feedback) | 10.8682 | 4.9604 | 2.19 | **0.0285*** |
| Condition D × **(**No Feedback vs. Feedback) | 10.1802 | 4.9517 | 2.06 | **0.0398*** |
| Trial order std. Condition A × (No Feedback vs. Feedback) | 2.2707 | 5.0353 | 0.45 | 0.6520 |
| Trial order std. Condition C × (No Feedback vs. No Feedback) | 4.7295 | 5.0260 | 0.94 | 0.3467 |
| Trial order std. Condition D × (No Feedback vs. Feedback) | 2.8231 | 4.9606 | 0.57 | 0.5693 |

In the next section, results from Tables 20 and 21 will be analyzed together in light of the findings obtained for the accuracy data that were presented in the previous section

***RT data analysis for the clitic region***

As seen in Table 20, the learners in the [- Feedback] group showed a clear tendency to read the clitic region in Condition B (the baseline, cl. sing., verb sing.) slower than participants in the [+ Feedback] group. Table 21 (in which residual RTs from both groups were aggregated) confirmed this: learners in the [+ Feedback] group read Condition B significantly faster ($\hat{\beta}$= -11.5971, *t*= -3.29, *p*= 0.0010, *d*= -0.0940).

As seen in Table 20, learners in the [+ Feedback] group spent significantly more time overall on the clitic region in Condition A (cl. sing., verb pl.,) than in the clitic region of Condition B ($\hat{\beta}$= 27.0516, *t*= -3.68, *p*= 0.0002, *d*= 0.0278). This difference can also be seen visually in Figure 14. Also, Table 20 shows that learners in the [+ Feedback] group spent even more time in the clitic region of Condition A (cl. sing., verb pl.) than the learners in the [- Feedback] group ($\hat{\beta}$= 15.6313, *t*= 3.13, *p*= 0.0017).

On the other hand, learners in the [- Feedback] group also showed different processing patterns for the clitic region in Condition A, but in this case, the slowdown was found across experimental trials instead of in the time spent overall in this clitic region ($\hat{\beta}$= 17.3549 *t*= -2.53, *p*= 0.0000).

Finally, as shown in Table 21, both experimental groups exhibited a speedup in the clitic region of Condition C (cl. pl., verb.pl) ($\hat{\beta}$= -28.0213, *t*=-5.61 , *p*=0.0000, *d*= -0.066). This means that when the clitic was plural, at least in Condition C, learners read this clitic region significantly faster in comparison to the clitic in singular in Condition B.

To remind the reader, learners in both experimental groups performed significantly worse in Condition C than they did in Condition B.

***RT data analysis for the Verb Region***

Table 20 shows that learners in the [+ Feedback] group showed an effect of trial order in the verb region in Condition B (the baseline, cl. sing., verb sing.) ($\hat{\beta}$= -19.7875 $t$= -2.04, $p$= 0.0418); this means that learners read the verb region in this condition more quickly as the experiment progressed. Additionally, Table 21 shows that learners in the [+ Feedback] group read the verb region in Condition B more quickly across experimental trials than learners in the [- Feedback] group ($\hat{\beta}$= -16.5376 $t$= -2.53, $p$= 0.0116). The speedup observed in the RT data in Condition B for the learners in the [+ Feedback] group corresponds with the speed of improvement observed in Condition B in this same group of learners. This suggests that feedback seemed to make learners both faster in processing and more accurate across experimental trials in Condition B (the baseline). Learners in the [+ Feedback] group spent more time in the verb region of Condition A (cl. sing., verb pl.), but this difference was not significant in comparison to the time spent in the verb region in Condition B.

On the other hand, as seen in Table 20, learners in the [-feedback] group spent significantly more time in the verb region in Condition A (cl. sing, verb pl.) as the experiment progressed ($\hat{\beta}$= 25.4853 $t$= 2.09, $p$= 0.0363). This means that across experimental trials, participants in the [- Feedback] group continued to show a slowdown in Condition A (cl. sing., verb pl.), a slowdown that began in the clitic region and was carried over to the verb region. To remind the reader, participants in the [-feedback] group exhibited a greater accuracy score in Condition A (cl. sing., verb pl.) overall than

in Condition B. This suggests that Condition A was more difficult to process (it demanded more processing costs) but also yielded higher accuracy scores for the [- Feedback] group.

Finally, as shown in Table 21, learners in both the [- Feedback] group and in the [+ Feedback] group spent significantly more time overall in the verb region in Condition D (cl. pl., verb sing.), compared to the time they spent in the verb region in Condition B ($\hat{\beta}$= 31.0959 $t$= 3.47, $p$= 0.0006). This suggests that for the both experimental groups, the verb region in Condition D was the most taxing. The accuracy rate in Condition D for both groups, however, was lower than the accuracy obtained in Condition B.

***RT data analysis for the spillover regions: determiner and noun (post-verbal subject)***

As shown in Table 20, the analysis conducted on the spillover regions corresponding to the post-verbal subject, verb + 1 (the determiner) and verb+ 2 (the noun), shows the same processing patterns observed in previous regions: the learners in the [+ Feedback] group exhibited a speedup in the post-verbal subject in Condition B as the experiment progressed since an effect of order was found in both spillover regions ($\hat{\beta}$= -15.9787 $t$= -2.46, $p$= 0.0139, for the determiner region in Condition B for the [+ Feedback] group, and $\hat{\beta}$= -11.8850 $t$= -2.22, $p$= 0.0263, for the noun region in Condition B for the [+ Feedback] group). Condition A (cl. sing, verb pl.), on the other hand, exhibited a slowdown across experimental trials in the [+ Feedback] group ($\hat{\beta}$= 23.9017 $t$= 2.62, $p$= 0.0089 for the determiner region in Condition A for the [+ Feedback] group; and $\hat{\beta}$= 16.1260 $t$= 2.14, $p$= 0.0321 for the noun region in Condition A for the [+ Feedback] group). In addition, Table 21 shows that the participants in the [+ Feedback]

group spent more time in the spillover regions in Condition D (cl. pl., verb sing.) than leaners in the [- Feedback] group ($\hat{\beta}$= 28.6153 $t$= 5.79, $p$= 0.0000 for the determiner in Condition D for the [+ Feedback] group; and $\hat{\beta}$=10.1802 $t$= 2.06, $p$= 0.0398 for the noun region in Condition D for the [+ Feedback] group).

With respect to the [- Feedback] group, Table 20 shows that learners in this group significantly spent more time in the determiner region in Condition B (the baseline, cl. sing, verb pl.) ($\hat{\beta}$= 21.1675 $t$= 4.07, $p$= 0.0000); learners spent even more time in this region than participants in the [+ Feedback] group, as shown by Table 21 ($\hat{\beta}$= -20.6565 $t$= -5.85, $p$= 0.0000). This may indicate that learners needed more time to decide who the agent of the sentence was in sentences in which the clitic and verb were both singular (the baseline condition).

Finally, in comparison to Condition B, learners in the [- Feedback] group spent more time in the spillover regions of Condition A (cl. sing., verb pl.) as the experiment progressed, since an order effect was found in both regions ($\hat{\beta}$= 19.4224 $t$= 2.61, $p$= 0.0091 for the determiner in Condition A for the [- Feedback] group; and $\hat{\beta}$= 11.5846, $t$= 1.73, $p$= 0.0321, for the noun region in Condition A for the [- Feedback] group). This slowdown continued from the previous regions, suggesting once again that as the experiment progressed participants in this group seemed to paid more attention to this particular condition (Condition A, cl. sing, verb pl.).

**Discussion**

This study investigated how English-speaking advanced learners of Spanish processed $O_{cl}$VS sentences conveying different agreement manipulations in the presence or absence of computer-delivered feedback. Accuracy data was analyzed overall and over

the course of the experiment to assess whether the agreement conditions conveying contrastive agreement cues between the clitic and the verb (i.e. mismatching conditions) were more beneficial in triggering a change in the current cue strength of the advanced English-speaking learners, and whether the provision of feedback enhanced such readjustment of cue weights in the direction of the L2 conflict validity, as suggested by CM research (MacWhinney, 2011; McDonald, 1987, 1989). A Self-Paced Reading (SPR) word-by-word was employed to establish processing costs comparisons in each of the match/mismatching conditions created in both [+/- Feedback] experimental groups.

*RQ1: Do accuracy rates in the experiment overall differ between a group which received computer-delivered feedback on accuracy responses and a group which did not receive feedback?*

In relation to learners' overall accuracy rates in the interpretation of $O_{cl}$VS sentences, overall accuracy scores showed once again that reliance on word order when interpreting $O_{cl}$VS sentences is still prevalent among learners in their third year of Spanish coursework at the college level. These results are consistent with the overall results obtained in Experiment 1 and in the pre-test of the SA study (Experiment 2).

Importantly, overall accuracy scores in the present study showed, as predicted, that the provision of feedback made a difference in how accurate learners were in coming to the correct interpretation of $O_{cl}$VS sentences. The participants in the [+ Feedback] group significantly outperformed the participants in the [- Feedback] group, and showed improvement of 100% or greater in all experimental conditions. By contrast, the accuracy rates overall for $O_{cl}$VS sentences in the [- Feedback] group were below chance. These

results suggests, as proposed by CM scholars (MacWhinney, 2011 and McDonald, 1987, 1989), that feedback plays an active role in the adjustment of cue strength when learners interpret conflict sentences, such as $O_{cl}$VS structures. However, it is worth noting that since the overall accuracy rates in the [+ Feedback] group were still at 63%, continued exposure to $O_{cl}$VS sentences seem to be needed in order for advanced English-speaking learners to continue readjusting the cue strength and to establish correct form-to-function mapping of $O_{cl}$VS sentences. As McDonald (1987, 1989) argued, the readjustment of L2 cue strengths can only increase *progressively* with exposure to new input sentences.

*RQ2: Do learners in the [+Feedback] and the [- Feedback] groups exhibit similar accuracy rates, overall and across experimental trials, in the matching/mismatching conditions between the clitic and the verb?*

*Overall scores:*

The participants in the [- Feedback] group showed the highest overall scores in Condition A (cl. sing., verb pl.), as expected. Condition A is the condition conveying contrastive agreement that in Experiment 1 exhibited higher accuracy scores across experimental trials. In the present study, even though the overall accuracy in Condition A and Condition B was very similar (32 % and 30 % respectively) and, therefore, this difference was not significant, the probability (effect size) of being accurate in Condition A for the learners in the [- Feedback] group was 51%, whereas the probability of being accurate in Condition B was 30%.

On the other hand, the participants in the [+ Feedback] group obtained maximum overall accuracy scores in the baseline condition, Condition B (cl sing., verb sing.).

Condition A (cl. sing., verb pl.) was second but very close in overall means. There were not statistical differences between Condition A and Condition B. The probabilities of a learner being accurate in Condition B and Condition A were 49% and 64% respectively.

To summarize, with regard to how participants in the two experimental groups processed $O_{cl}VS$ sentences in the matching/mismatching conditions *overall*, the results showed that Condition A (cl. sing., verb pl.) and Condition B (the baseline, cl. sing., verb sing.) were the two conditions that elicited the higher accuracy scores overall in participants in *both groups*. This means that Condition A (cl. sing., verb pl.) was the only condition that was not significantly different from B in overall accuracy, in both [-Feedback] and [+ Feedback] groups. This same pattern in the accuracy overall was observed in Experiment 1 and in the SA study (Experiment 2). The other two conditions, Condition C (cl. pl., verb pl.) and Condition D (cl. pl., verb sing.) were significantly *worse* than Condition B, in both groups. This shared pattern between groups suggests that learners may have found easier to interpret $O_{cl}VS$ sentences conveying singular object clitics (Condition A and Condition B). It is worth noting here that this study hypothesized that both mismatching conditions would elicit higher accuracy rates, but again, as in the two previous studies, Condition A (cl., sing., Verb pl.) was qualitatively different from Condition D (cl. pl., Verb sing.), suggesting that plural number markers attached to the verb (rather than the clitic) seem to be more helpful or qualitatively more beneficial in L2 cue hierarchy activation.

*Accuracy rates across experimental trials: Trial order effects*

However, the rate of improvement across experimental trials in both experimental groups differed substantially:

No trial order effects were found in any of the agreement conditions in the [- Feedback] group, indicating that participants in this group did not show significant improvements as the experiment progressed. As mentioned earlier, this is one of the main differences between the participants in the [- Feedback] group in the present study and the participants in the SA study (Experiment 2). In the pre-test of the SA study, where participants did not receive feedback, they did show improvements as the experiment progressed, even though their accuracy scores overall in the pre-test were below 50%. As mentioned earlier, future research regarding the role that individual differences (e.g., motivation, aptitude) may play in the processing of $O_{cl}$VS sentences needs to be conducted.

This study predicted that the [+ Feedback] group would exhibit an effect of order (improvement across the experiment) for Condition A (cl. sing., verb pl.) (based on the results of Experiment 1). Surprisingly, an order effect was found for the [+ Feedback] group in Condition B (the baseline), the condition in which this group's participants also showed the highest overall scores: this indicated that in Condition B (cl. sing., verb sing.) participants in the [+ Feedback] group not only exhibited the highest overall accuracy scores but also improved faster as the experiment progressed. The reason for this seems to be related to the high processing costs demanded by Condition A (cl., sing. verb pl.), which also exhibited similar overall accuracy scores (as discussed above). It was expected that Condition A (cl., sing., verb pl.) would require higher processing costs than Condition B (the baseline) since it was predicted that if learners misinterpreted the clitic pronoun as the agent of the sentence, this would create an expectation that the clitic and verb should agree in number, which would result in processing difficulty (i.e., higher

RTs) when verb and clitic mismatch as opposed to when they match in number. As will be further discussed in the next (RQ3), participants in Experiment 1 (the other study in this dissertation in which feedback was provided) did not show a significant slowdown in any region of Condition A as participants in this experiment did, and in Experiment 1, Condition B, the baseline condition, did not exhibit a speedup either. This suggests that verb types (auxiliary verbs vs. non-auxiliary verbs) are not processed in the same way, potentially because the auxiliary is more frequent. Moreover, in Experiment 1, the verb was presented to learners in repeated opportunities, presumably enhancing processability.

*RQ3: When interpreting $O_{cl}$VS sentences, do learners who receive computer-delivered feedback exhibit the same processing costs as learners who do not receive feedback? And if so, do processing costs between the two experimental groups differ in the $O_{cl}$VS sentences where clitic and verb match/mismatch in number?*

When analyzing the residual RT data as aggregated data, a pattern emerged showing that both experimental groups [- Feedback] and [+ Feedback] spent significantly more time in the verb region in Condition D (cl. pl., verb sing.) in comparison to Condition B (the baseline). This suggests that all participants in this experiment found this condition more difficult to process (and also harder to interpret correctly, due to the low accuracy scores exhibited). This difficulty observed in the verb region of Condition D was even more significantly pronounced in the [+ Feedback] group, which exhibited the highest mean RTs in the verb region in Condition D (cl. pl., verb sing.).

As discussed in both previous chapters, following McCarthy (2008), learners seem to rely on morphological defaults in comprehension, so it is likely that they are more predisposed to establish agreement between a plural element in sentence-initial

position (in this case, a clitic, *los*-them) and a singular unmarked verb, which, in turn, would lead to a greater number of interpretation errors. Or, from a CM perspective (Haskell & MacDonald, 2003), this combination (cl. pl., verb sing) seems to be more 'plausible' for the learners. For instance, it may be the case that learners in the [+ Feedback] group were surprised to receive error messages when they made mistakes in this condition, and that triggered their focus on the verb region in this condition.

With regard to how participants in the experiment processed the different matching/mismatching condition *across experimental trials*, the residual RT data showed that the [+ Feedback] group presented a significant speedup across experimental trials in Condition B (the baseline, cl. sing., verb sing.), which was the condition in which learners also exhibited the highest scores in accuracy and a trial order effect in rate of improvement. The speedup was observed in all regions, from the clitic region to the spillover regions. This means that in Condition B, learners in the [+ Feedback] group became more accurate overall, improved faster across experimental trials, and read faster as the experiment progressed (this condition demanded significantly less processing costs). This speedup found in Condition B supports Sasaki's (1998) results which showed that learners' response latency shortened throughout feedback training sessions. Regarding Condition A (cl. sing., verb pl.), the participants in the [+ Feedback] group showed a slowdown in this condition in both spillover regions, suggesting that this condition was cognitively taxing for participants in the [+ Feedback] group.

Participants in the [- Feedback] group also showed a slowdown across experimental trials in Condition A (cl. sing, verb pl.), but from earlier regions. In fact, the slowdown was observed in all regions, from the clitic region to the spillover regions,

suggesting that participants paid more attention to this condition in particular as the experiment progressed. This slowdown correlated with the higher accuracy scores this condition elicited from participants in the [- Feedback] group (although this difference was not significant).

*Differences in the {+] feedback group with the results in Experiment 1:*

Coming back to the question of why learners in the [+ Feedback] group exhibited a significant improvement across experimental trials in Condition B (the baseline, cl. sing., verb sing.)– a result which was not expected based on the finding from Experiment 1, which showed an order effect in Condition A (cl–sing., verb pl.), this difference may be attributed to different factors: (1) there were differences in the experimental design between Experiment 1 and the present study; whereas the main verb in Experiment 1 was the auxiliary verb *estar*, "to be," in this chapter's study, $O_{cl}$VS sentences conveyed verbs in the simple present (e.g., *mirar*, "to look") and fillers items included contextual sentences; (2) the accuracy overall in Condition A (cl. sing., verb pl.), although numerically lower, was not significantly different from Condition B, as were the other two conditions, Condition C (cl. pl., verb pl.) and Condition D (cl. pl., verb sing.), which yielded overall accuracy scores that were significantly lower than the ones yielded by Condition B, (3) Condition A (cl sing., verb pl.) was significantly more difficult to process in this experiment than it was in Experiment 1. Participants Experiment 1 did not show a significant slowdown in any region of Condition A, as participants in this experiment did, and Condition B in Experiment 1 did not exhibit a speedup either. This suggests that verb types (auxiliary verbs vs. non-auxiliary verbs) have not been processed

in the same way. It seems that, whereas in Experiment 1, the auxiliary verb – which is highly frequent in the classroom (e.g., Guntermann 1992) and which was also repeatedly presented to learners in the target items over the course of the experiment – did not involve the same degree of processing difficulty in the mismatching conditions as verbs in the present simple –which were presented once-did in the present experiment.

Since the accuracy performance overall in Condition A and Condition B are very tight in this study in both experimental groups, one possible explanation that may account from the differences found between Experiment 1 and the present study regarding the rate of improvement across experimental trials, where the main difference resides, is that the participants in the [+ Feedback] group in this study may have started making improvements across experimental trials in Condition A (cl. sing., verb pl.) at the beginning of the experiment; however, due to this condition's difficulties for processing— as shown by the significant slowdown Condition A exhibited as the experiment progressed— it might follow that participants were not able to keep up the rate of improvement for this condition because this condition was becoming more and more cognitive taxing over time. However, it is worth noting here that its difficulty did not prevent participants perform well in this condition. In other words, participants in the [+ Feedback] group, and to some extent participant in the [- Feedback] group as well, were able to be accurate in Condition A *in spite of* its difficulties for processing. And this is *not* the case for Condition D (for none of the experimental groups). Therefore, the internal dynamics of this study may be conceived of as having given ground to a positive 'washback effect': learners in the [+ Feedback] group may have started to reset their cue weights towards L2 cue validity in Condition A, and used this altered cue strength, as

McDonald (1987, 1989) suggested, to better interpret less cognitively taxing input sentences such as those in Condition B (the baseline, cl. sing., verb pl.), which also conveys singular object clitics. To prove this hypothesis, further studies manipulating different verb type in $O_{cl}$VS sentences (e.g., auxiliary vs. no auxiliary), as well as studies comparing different agreement conditions order in the experiment (e.g., mismatching sentences in the first half, and matching sentences in the second half, and viceversa) would need to be conducted.

Finally, since the [- Feedback] group did show a move towards more efficient form-to-function mappings in Condition A (cl. sing., verb pl.), this condition exhibited higher probabilities to yield better accuracy scores, and participants showed sensitivity to agreement mismatches since condition A exhibited a slowdown across experimental trials, suggesting that manipulated input is not ignored in more implicit conditions not receiving feedback, as suggested by Sanz and Morgan-Short (2004). However, it seems that learners who only have access to manipulated input need more time, more practice, and a greater amount of exposure to other sources of input – for example, Sanz and Morgan-Short's (2004) study employed written, aural, and textual activities – in order to exhibit a significant alteration in their current cue strength when processing non-canonical sentences.

**Conclusions**

The results in the present study show that computer-delivered feedback makes a difference when English-speaking advanced learners process $O_{cl}$VS sentences in Spanish, and more specifically that advanced learners benefit from the implicit feedback provided by means of a computer when interpreting complex structures that are meaningful for

efficient communication. Learners who received implicit feedback in this study were better able to come to correct form-to-function mappings when deciding who the agent of the sentence was, giving support to the work of CM scholars such as MacWhinney (2011) and McDonald (1987, 1989) who have indicated that feedback can progressively alter current cue strengths in the direction of the L2 cue conflict validity.

In spite of the encouraging results found in this study, limiting factors need to be considered. The computer-delivered feedback effects which were found only addressed immediate effects on the target items, and so no evidence was provided as to the lasting effects that implicit computer-delivered feedback may have over time.

With regards to the role that contrastive agreement cues between the clitic and the verb plays in the processing of $O_{cl}$VS sentences in advanced learners of Spanish, results show that Condition A (cl. sing., verb pl.) continues to exhibit qualitative differences, but its facilitative effects on $O_{cl}$VS sentence interpretation may be mediated by verb type. For this reason more research is needed in this direction.

Finally, the low accuracy scores observed in the [- Feedback] group do not reflect learners' potential of benefiting from manipulated L2 input, since participants in this group showed higher probabilities to get better in Condition A and showed sensitivity to agreement mismatches between the clitic and the verb in Condition A as well. Therefore, more research needs to be conducted to test how advanced learners process $O_{cl}$VS sentences in different agreement conditions between the clitic and the verb in studies that may include more tokens, more practice sessions, different activities at the sentential and suprasentential level, and different type of outcome measures (e.g., visual agent-choice

tasks) so as to provide further empirical evidence as to how different exposure conditions to structured input alone may benefit cue strengths readjustments.

## CHAPTER FIVE: Conclusions

As outlined in the introduction, this dissertation has been motivated by the need to further investigate both offline and online how advanced learners of Spanish process non-canonical sentences, specifically $O_{cl}$VS sentences in the written modality, and whether learners' processing patterns are influenced by different external factors. This dissertation focused on the role of: (a) agreement morphology, (b) immersion experience, and (c) computer-delivered feedback.

The present chapter looks at the three experiments conducted and provides a discussion of the similarities and differences found, taking into account changes in the research design. First, a brief review of the similarities and differences between the three experiments will be outlined along with the reasons for introducing changes in the research design, and a concise evaluation of how those changes worked; then, the results of the three experiments will be addressed. Finally, a section on implications, limitations, and areas for future research will follow, along with overall conclusions that have been drawn from the results found.

### Similarities and differences in the experiments conducted

Experiments 1, 2, and 3 tested English-speaking learners in their third year of Spanish coursework at the same college institution, and shared several research design features: (1) the methodology, which was word-by-word SPR, (2) the type of object clitics examined: direct object clitics in the 3rd person, which referred to animate referents, (3) the manipulation of *agreement* (Matching/Mismatching) and *Number* (singular/ plural) as factors so as to create different agreement conditions, (4) the outcome measure, True/False questions targeting the agent of the $O_{cl}$VS sentences.

Experiments 1, 2, and 3 also had several differences, which responded mainly to improvements or readjustment in the research design based on the specific research questions of each experiment. Table 26 shows the main differences between the three experiments conducted:

**Table 26: Main differences in research design in the three experiments conducted**

| | **Experiment 1 (Pilot study)** | **Experiment 2 (SA study)** | **Experiment 3 (Feedback Study)** |
|---|---|---|---|
| *N* | 38 | 20 | 90 |
| Yes-No Feedback | Yes | No | Yes /No [+/-Feedback] |
| Number of Tokens $O_{cl}VS$ sentences | 24 (6) | 48 (12) | 32 (8) |
| SV sentences | Yes | Yes | No |
| GJTs | No | Yes | No |
| Design | 2 x 2 within-subjects design (Number and Agreement as factors) | 2 x 2 within-subjects design (Number and Agreement as factors) & Pre-post test design (Week 1 and Week 5) | 2 x 2 within-subjects design (Number and Agreement as factors) & [+/- Feedback] between-subjects design |
| Verb/tense | Auxiliary '*estar*' ("to be") + gerund | Simple present | Simple present |
| Fillers | No contextual sentence provided | No contextual sentence provided | Contextual sentence provided/ specific pattern followed |
| Length of SVO sentences | Not strictly controlled | Not strictly controlled | Strictly controlled |
| Vocabulary quiz | No quiz | Short quiz (focused mainly on professions) | Comprehensive quiz: target sentences and fillers |

As seen in Table 26, the number of tokens varied in the three studies. Whereas in Experiment 1, the number of tokens was 24 – 6 per condition – in Experiment 2, which had a smaller *N* size, the number of tokens was doubled in order to increase power. However, in Experiment 3, this number was reduced so as to make the whole experiment more manageable.

Experiment 1 and Experiment 2 included SV sentences as a control sub-

experiment to check whether participants were able to use morphological markers online. The rational was that, if learners used agreement morphology to process agreement mismatches in both ungrammatical SV sentences (**El perro duermen*, "The dog sleep") and $O_{cl}$VS sentences conveying contrastive agreement (*Lo*-cl. sing *están*-v. pl. *buscando las chicas*, "The girls are looking for him), then learners should present a slowdown in the verb region (or spillover region) in both canonical and non-canonical sentences.

Both Experiment 1 and Experiment 2 showed that the Residual RT data for SV sentences exhibited slowdowns in the regions of interest, suggesting that learners were able to use agreement markers in real-time in canonical sentences. However, the differences in speed between grammatical and ungrammatical sentence processing were not significant. A power analysis revealed that these sentences needed a high number of observations. Moreover, the GJT included in Experiment 2 showed that these learners were extremely accurate at judging grammaticality offline. The differences in outcome observed in the SPR and the GJT in SV sentences might be explained by the differences in processing costs that online tasks, as opposed to offline tasks, entail, since in real-time processing, cognitive resources must be recruited to enable sensitivity to agreement violations between the subject and the verb (Hoshino et al., 2009). Also, in SV sentences, agreement computations were made on ungrammatical sentences in 50 % of the sentences; therefore, for reasons already explained – due to the ambiguity in the ungrammatical sentences – the comprehension questions for these sentences could not be directed to the meaning of the full sentence. By contrast, in the 50 % of the $O_{cl}$VS sentences where clitic and verb exhibited contrastive morphological markers, agreement computations were made on grammatical sentences; therefore, comprehension questions

could be directed to learners' interpretations of the entire sentence (form-to-function mappings). Finally, the element in sentence-initial position in the $O_{cl}$VS sentences consisted of a clitic, which is not perceptually salient. By contrast, the SV sentences consisted of a full NP, which are longer in duration and length than clitics, more reliable, and easier to parse because they are not psychologically bound to another word. Therefore, since it may be the case that the manner in which advanced learners use agreement markers in canonical and non-canonical sentences was not completely comparable, neither SV sentences nor GJTs were included in Experiment 3.

Regarding tense, Experiment 2 and Experiment 3 investigated $O_{cl}$VS sentences conveying a verb in the simple present instead of an auxiliary verb in the present continuous (auxiliary *estar* 'to be' + gerund), as in Experiment 1, to address how learners processed agreement markers attached to frequent verbs in Spanish rather than on an auxiliary verb such *estar*,"to be." This also precluded the repetition of the same verb in all target sentences. Using verbs in the simple present in all target items followed previous SA studies which employed the present simple in their stimuli (e.g., Grey et al., 2015).

Even though, in Latin square designs, sentence length should not affect learners' processing of target sentences, since all learners are exposed to the same items, with the sole difference that the target items vary in the agreement condition they are read, further steps were taken in Experiment 3 to control for the length of the sentence so as to give more unity to the experiment as a whole.

Fillers were also improved throughout the three experiments. In Experiment 1, filler sentences were usually short in length, and many of the T/F questions asked

participants to recognize the mentioning of a specific word (e.g., In this phrase the word 'dog' is mentioned). In Experiment 2, fillers were more informative and engaging, and the T/ F questions that followed each filler were more challengig than those in Experiment 1: for example, true-false comprehension questions required learners to compare certain information given in the filler sentence with information provided in the comprehension question (e.g., the place where the action took place). In Experiment 3, fillers consisted of sentences related to learners' lives or to every-day topics, and they conveyed expressions with the verbs *ser* (e.g., "*ser trabajador"*, "to be hardworking'), *estar* ( e.g., "to be sad"); and *tener* (e.g., "*tener hambre*","to be hungry," "*tener sueño*," "to be sleepy"). Moreover, fillers included a contextual sentence, as in the $O_{cl}VS$ sentences. Fillers not only followed a specific pattern but also served as true distractors. False comprehension questions for *ser* and *estar* ('to be') sentences conveyed an antonym or indicated the opposite of what was conveyed in the sentence.

Whereas in Experiment 1, no vocabulary quiz was included since target sentences consisted of highly frequent words extracted from textbooks used in low-level courses, in Experiment 2 and Experiment 3, a vocabulary quiz was administered in order to ensure that learners knew the vocabulary conveyed in the sentences; this was done so as to prevent potential latencies in the reading times due to unfamiliarity with the experimental stimuli (Sagarra, 2007). The vocabulary quiz in Experiment 2 comprised words from the target sentences and fillers: mainly of the professions ($k$=36). The vocabulary quiz in Experiment 3 included verbs, professions, nationalities, and a few prepositional phrases in the target sentences and fillers ($k$= 123). Additionally, in Experiment 3 learners completed a questionnaire in which they were asked to self-assess the difficulty of the

vocabulary and write down the words they did not know or did not remember well on an answer sheet. This information allowed the researcher to address how learners perceived the vocabulary and also to check which words presented more difficulties. Overall, the inclusion of a vocabulary quiz in the research design was beneficial for learners in activating the meaning of some words that they did not remember. Moreover, learners actively engaged with the vocabulary quiz since it was administered with free online software (Quizlet) that most participants had previously used to learn vocabulary on their own.

In sum, because of the differences outlined in Table 26, no direct comparisons can be established between the three experiments; however, these experiments are interrelated due to the many features they do share, as well as the population of learners they tested. In addition, inferential statistics have been standardized so as to allow for better comparability and replicability. Taking into account these similarities and differences, the next section discusses the findings obtained in the three experiments in terms of accuracy and latency in the target sentences.

## Discussion

### Accuracy outcomes in $O_{cl}VS$ sentences

Accuracy was measured in the three experiments overall and across experimental trials. By conducting both analyses, it was possible to examine not only how well learners did in Condition A (cl. sing., verb pl.), C (cl. pl., verb pl.) and D (cl. sing., verb pl.) in comparison to Condition B (the baseline; cl. sing., verb sing.), but also whether learners made major improvements in each of those conditions as the experiment progressed.

In Experiment 1 (Chapter 1), it was hypothesized that if advanced learners

continued to rely on word order as the most valid cue for agenthood, but, at the same time, they were also able to benefit from contrastive agreement in order to correctly assign sentence-initial object clitics the role of patient, then the two conditions conveying contrastive agreement between the clitic and the verb would present higher accuracy rates than the two conditions not exhibiting contrastive agreement.

The overall scores in the pilot study (*mean accuracy*: min. 58% - max. 61%) were the first to confirm that even L2 learners at advanced levels still exhibit a tendency to rely on word order when processing $O_{cl}$VS sentences in the written modality. The mismatching plural verb condition (cl. sing., verb pl.) exhibited the highest overall accuracy scores, followed by the baseline ( cl. sing., verb sing.) in second place, and the mismatching plural clitic condition (cl. pl., verb sing.) and the condition where clitic and verb were both in plural (cl. pl., verb pl.) in third and fourth place respectively. Whereas no significant differences were found in the overall accuracy scores, interestingly, an interaction was found in the mismatching plural verb condition (cl sing., verb pl.) when the rate of improvement across experimental trials was analyzed. This meant that learners were able to improve significantly faster in the mismatching condition in which the morphological cue indexing agreement mismatch was on the verb. One possible explanation for this outcome was that the detectability of agreement cues may be enhanced when plural markers are attached to content words such as verbs. Likewise, the use of the auxiliary *estar*, "be," may also have played a role since it is a highly frequent verb in the L2 classroom (e.g., Guntermann 1992), and this may have enhanced the saliency of the agreement cue, favoring processability. A final potential explanation was that, since learners seem to rely on morphological defaults in comprehension, they would

be less predisposed to establish agreement between a singular noun in sentence-initial position and a plural marked verb (McCarthy, 2008). From a CM perspective, this tendency may be explained in terms of "plausibility," following Haskell and MacDonald (2005); that is, it would be less plausible for learners to establish agreement between a singular clitic and a plural verb. The findings in the pilot study served as a starting point for the predictions made in Experiments 2 and 3.

Unlike Experiment 1, Experiment 2 (Chapter Three) did not provide learners with feedback on their responses since the immersion context was characterized as implicit. In addition, this study was not conceived as a training experiment. The research design in Experiment 2 was more complex than that of Experiment 1 since, in Experiment 2, learners performed an SPR at Week 1 and also at Week 5 (pre-post test). The overall accuracy in the pre-test (Week 1) was lower than in the pilot study, potentially due to the lack of feedback, and showed, in line with the pilot study, that learners still preferred word order as a cue rather than morphology for assigning semantic functions to NPs in $O_{cl}$VS sentences (*mean accuracy*: min. 35% - max. 45%). Interestingly, two patterns were found in the pre-test: first, the order in which learners exhibited better and worse overall accuracy scores was the same as in the pilot study: The mismatching plural verb condition (cl. sing., verb pl.) was followed in terms of overall accuracy by the baseline (cl. sing., verb sing.), the mismatching plural clitic condition (cl. pl., verb pl.), and the condition in which the clitic and verb were both in plural (cl. pl., verb pl.). Additionally, as in the pilot study, learners' accuracy significantly improved across experimental trials in the mismatching plural verb condition, which indicated that the mismatching plural verb condition was also more helpful for learners to overcome word order bias, even in

the absence of feedback.

In the post-test of Experiment 2, as predicted, learners significantly improved in all agreement conditions, showing that learners were better at assigning semantic roles to NPs in $O_{cl}$VS sentences after five weeks of immersion abroad. This significant improvement in $O_{cl}$VS sentence accuracy was accompanied by a significant improvement in the participants' ability to judge grammaticality in canonical sentences at Week 5. Moreover, the order in which learners exhibited better and worse overall accuracy scores in the $O_{cl}$VS sentences in the post-test was similar to that seen in the pilot study and the pre-test; the only difference was that learners performed equally in the mismatching plural clitic condition (cl. pl., verb pl.) and the condition in which the clitic and verb were both in plural (cl. pl., verb pl.), which were tied for third place. No significant differences were found in overall performance, and no interaction was observed between Condition × Trial Order in the post-test. However, when the accuracy scores from the pre-test and the post-test were analyzed together, an interaction also emerged for the mismatching plural verb condition (cl. sing., verb pl.), as in the pre-test, suggesting once again that the mismatching plural verb condition exhibited qualitative differences.

Experiment 3 (Chapter Four) compared, in a between-subjects design, how learners who received and did not receive computer-delivered feedback processed $O_{cl}$VS sentences in the same four agreement conditions tested in Experiments 1 and 2. In this experiment, which incorporated a larger number of changes in comparison to the pilot study (e.g., filler characteristics, length of $O_{cl}$VS sentences, length of the vocabulary quiz) (see Table 22), it was hypothesized that learners would perform better when feedback was provided on their responses. In addition, it was predicted for both

experimental groups that the mismatching plural verb condition (cl. sing., verb pl.) would be the condition exhibiting the highest overall accuracy scores, and that the [+ Feedback] group would be more likely to show a significant improvement across experimental trials in the mismatching plural verb condition than the [- Feedback] group. Consistent with the overall results obtained in Experiments 1 and 2, the overall accuracy in both [+/ - Feedback] groups showed that learners still have difficulties processing $O_{cl}$VS sentences at this level of proficiency. However, as predicted, learners in the [+ Feedback} group (*mean accuracy*: min. 59%, max. 63%) significantly outperformed learners in the [- Feedback] group (*mean accuracy*: min. 23%, max. 32%), who performed below chance. These results showed that learners benefited from computer-delivered feedback to come to the right interpretation of $O_{cl}$VS sentences, even when agreement information was not contrastive; the improvement was observed in all agreement conditions. The baseline (cl. sing., verb sing.) was the condition that exhibited the highest overall accuracy scores. In overall performance, this condition was closely followed by the mismatching plural verb condition (cl. sing., verb pl.): the difference between these two conditions was 1%. That is, the mismatching plural verb condition was not significantly different from the baseline. The [- Feedback] group, however, followed the same pattern as learners in Experiments 1 and 2 in that this group of learners showed the highest overall accuracy scores in the mismatching plural verb condition (cl. sing., verb pl.), and secondly, in the baseline condition (cl. sing., verb sing.). Furthermore, an interesting pattern emerged in both experimental groups, distinguishing this experiment from Experiments 1 and 2: learners in both groups performed significantly worse in the condition where clitic and verb were both in plural (cl. pl., verb pl.) and the mismatching plural clitic condition (cl.

pl., verb sing.) overall. This suggested that learners may have found sentences conveying plural object clitics harder to process. In this respect, Lee (1987) and Malovrh (2006) have argued (for aural stimuli), following Slobin (1979), that clitic plural forms lead learners to more interpretation errors than singular clitic forms since plural markers (when attached to clitics) would be morphologically more complex and cognitively more demanding. In addition, it might have been the case that since the experimental fillers were also more complex in Experiment 3 than in the other two experiments – fillers not only were preceded by a contextual sentence so as to better fulfill their function as distractors, but they also required learners to pay attention to expressions conveying *ser* and *estar* ("to be") –the task was therefore more cognitively demanding in Experiment 3 and learners' cognitive resources might have been taxed by the cognitive load imposed in processing both $O_{cl}$VS sentences and sentences conveying *ser* and *estar* ("to be") expressions.. Also it worth noting here that in Experiment 1, the target sentences conveyed the auxiliary '*estar*' ("be") instead of verbs in the simple present. As mentioned earlier, the auxiliary '*estar*' used for target items might have favored general processability in Experiment 1.

In relation to the rate of improvement across experimental trials, the patterns observed in the [+ Feedback] and [- Feedback] groups differed substantially. Whereas no trial order effects were found in the [- Feedback] group, an order effect was found for the [+ Feedback] group in the baseline condition (cl. sing., verb sing.), instead of in the mismatching plural verb condition (cl. sing., verb pl.) as predicted; therefore, the baseline condition (cl. sing., verb sing.) was the condition in which participants in the [+ Feedback] group exhibited the highest overall accuracy scores and also the greatest

improvement as the experiment progressed. This outcome seems to be related to the processing costs which the mismatching plural verb condition (cl., sing. verb pl.) seemed to have demanded from learners, as will be discussed in the next section.

**$O_{cl}$VS sentence latency**

Regarding $O_{cl}$VS sentences processing costs, in Experiment 1 (Chapter Two) it was hypothesized that if learners misinterpreted the clitic pronoun as the agent of the sentence, and if this created an expectation that the clitic and verb should agree in number, then, when the verb and clitic mismatched, the verb region, or spillover regions, would present higher RTs in comparison with the conditions in which the clitic and verb matched in number. The RT data in the pilot study showed that participants exhibited a tendency to spend more time in the verb region (the auxiliary '*estar*', "be") when the $O_{cl}$VS sentences mismatched in number, and that the mismatching plural verb condition (cl., sing. verb pl.) was the condition that presented the higher RTs in the verb region, although these differences were not significant.

Visually, the RT data in Experiment 2 (Chapter Three) showed that when learners did not receive feedback, they tended to spend less time in the verb region than when they received feedback: the processing patterns observed in the verb region in Experiment 1 resembled a ∧ shape, whereas in Experiment 2 the learners' processing pattern in the verb region resembled a V shape. With regard to the speed with which learners processed $O_{cl}$VS sentences as the experiment progressed at Week 1 and at Week 5, a speedup was observed across experimental trials in the verb region and spillover regions in the baseline condition (cl. sing., verb sing.) in the pre-test. By contrast, the post-test latency data (Trial Order × Pre vs. Post) showed a slowdown across

experimental trials in the verb region and, especially, in the post-verbal subject in this same condition (the baseline), which suggested that learners at Week 5 might have found these regions more relevant for assigning agenthood in sentences in which agreement was not contrastive, since this slowdown was accompanied by a general improvement in accuracy, that is, improvement including the baseline condition.

In Experiment 3 (Chapter Four), the [+ Feedback] group presented a significant speedup across experimental trials in the clitic, verb, and spillover regions in the baseline condition (cl. sing., verb sing.), which was the condition in which learners also exhibited the highest accuracy scores overall and also major improvements as the experiment progressed (i.e., the baseline condition exhibited order effects in terms of accuracy and latency). The speedup found in the baseline condition in the feedback study supports Sasaki's (1998) results, which showed that learners' response latency shortened under the influence of feedback (i.e., learners who received feedback got faster).

Regarding the mismatching plural verb condition (cl. sing., verb pl.), the participants in the [+ Feedback] group showed a slowdown in this condition in both spillover regions, suggesting that this condition was somehow taxing for participants in the [+ Feedback] group. The higher processing costs exhibited in this condition were not accompanied by a trial order effect in accuracy; therefore, it was hypothesized that, since learners got significantly faster in the baseline condition (the baseline; cl. sing., verb sing), the higher processing costs required by the mismatching plural verb condition prevented learners from keeping up the rate of improvement in this condition, since the mismatching plural verb condition was becoming more and more cognitively taxing over time. However, such difficulty in processing did not prevent participants from performing

well in the mismatching plural verb condition, which was the only condition that was not significantly different from the baseline condition. Based on McDonald's (1987, 1989) learning on error mechanism, learners in the [+ feedback] group may have started to reset their cue weights towards L2 cue validity in the mismatching plural verb condition and used this altered cue strength to better interpret less cognitively taxing input sentences such as those in the baseline (cl. sing., verb pl.), which also conveys singular object clitics.

Participants in the [- Feedback] group also showed a slowdown across trials in the mismatching plural verb condition (cl. sing, verb pl.), but in earlier regions, from the clitic region to the spillover regions, suggesting that participants paid more attention to these regions in this condition in particular as the experiment progressed. This slowdown correlated with the higher accuracy scores this condition elicited from participants in the [- Feedback] group (although this difference was not significant).

Finally, the aggregated RT data showed that both experimental groups [+/- Feedback] spent significantly more time in the verb region overall in the mismatching plural clitic condition (cl. pl., verb sing.). This significant slowdown in the verb region was in line with the low performance learners in both groups presented in this condition. As discussed in the accuracy section, learners seemed to have had difficulties correctly interpreting sentences conveying plural clitics, and this difficulty in interpretation reflected on the RT data as a significant slowdown in the verb region of the mismatching plural clitic condition, which was the only condition which not only conveyed plural object clitics, but also conveyed contrastive agreement. Therefore, it is not surprising that the mismatching plural clitic condition (cl. pl., verb sing.) was more taxing than the

condition where clitic and verb were both in plural (cl. pl., verb pl.).

**Methodological, theoretical, and pedagogical implications**

The results of the studies in this dissertation have several methodological, theoretical, and pedagogical implications. Methodologically, the experiments conducted here confirm the relevance of addressing sentences processing both offline and online in an effort to better explain the effects that the variables under investigation have on accuracy and on processing costs. As discussed above, results in these experiments have shown that differences in accuracy are generally accompanied by changes in latency patterns. For example, slowdowns in a verb region or spillover region across experimental trials may reflect participants' increasing attention to regions that are highly informative for categorization, as shown in the post-test of Experiment 2. By contrast, a speedup across experimental trials may indicate that a condition is more efficiently processed as the experiment progressed, as shown for learners in the [+ Feedback] group in Experiment 3, specifically, for the baseline condition (cl. sing., verb sing.). In other words, latency data needs to be analyzed in relation to accuracy data in order to determine whether slowdowns in processing are related to learners' increasing attention or processing difficulties, or whether speedups are related to increasing effectiveness in processing or to learners' tendency to skip a specific region. In addition, offline and online data need to also be analyzed in relation to the research design and the characteristics of the input to which learners were exposed. It is also informative for the researcher to analyze both accuracy data and RT data overall and across experimental trials, as this dissertation has done, in order to detect order effects that reflect learners' improvements or changes in online processing. It is worth noting here that online

methodologies such as SPR can be used in a pre post-test design so as to tap into the double nature of processing – accuracy and latency. This was the approach adopted in Experiment 2 on the effects of immersion, for which the pre-test and post-test SPRs were conducted *in situ*.

Theoretically, the experiments in this dissertation have shown that, following the Competition Model (Bates & MacWhinney, 1981, 1982, 1987, 1989; MacWhinney, 2012), contrastive agreement cues can help L2 learners overcome L1 processing biases. In other words, as suggested by CM scholars (e.g., Hernandez et al., 1994), English-speaking advanced learners of Spanish can exhibit different processing patterns depending on the quality of the input, as was also the case for early bilinguals. This suggests that intra-subject variation can emerge amongst advanced learners of Spanish. Experiment 3 has also shown that feedback can progressively alter L2 cue strengths in the direction of the L2 cue conflict validity, supporting McDonald's (1987, 1989) learning on-error model. However, since in Experiment 2 (the SA study) learners were also able to improve across experimental trials in the pre-test, and overall after immersion, therefore, the role of manipulated input (and the richness of the input) on L2 morphosyntactic processing cannot be disregarded. It seems that the role of feedback on L2 cue strength realignment is related to the speed with which L2 data is processed.

Finally, the pedagogical implications of this dissertation point to the fact that even advanced learners of Spanish exhibit difficulties in processing $O_{cl}$VS sentences. The results of the three experiments conducted here consistently pointed in this same direction. In addition, as shown in the Experiments 1, 2 and 3, learners seem to benefit from (a) sentences conveying agreement mismatches between the clitic and the verb,

most specifically when the clitic is singular and the verb is plural, (b) continuous exposure to these sentences in immersion contexts, and (c) the provision of simple, yes/no feedback on learners' performance.

**Limitation and future research**

In spite of the encouraging results found in this dissertation, limiting factors need to be considered. The first limitation of this dissertation, especially with regards to Experiments 1 and 2, is the number of subjects involved in the analyses. This was especially problematic in the sub-experiment conveying subject-verb agreement violations in SV sentences, the purpose of which was to determine whether advanced learners were sensitive to subject-verb agreement violations online.

A second important limitation is that this dissertation did not address the lasting effects of the variables under investigation; discussions were based on immediate effects of either input manipulations, or of provision of feedback, or both. Additionally, in the SA study no delayed-post-test was administered in order to examine whether the learning gains found were retained after a considerable amount of time after immersion.

A third limitation concerns the experiments' use of True-False comprehension questions as the only outcome measure. Even though True-False questions undoubtedly fulfill the purpose of addressing whether learners have interpreted the sentence-initial object clitic as the agent or the patient, they do not inform us as to how learners' responses might change if they were engaged in more communicative tasks.

The findings in this dissertation serve as a fruitful arena for future research. For example, the examination of different populations, such as heritage speakers, could be beneficial to better understand the potential differential effects of individual differences

such as language proficiency and experience on the rate at which cue realignment takes place when processing $O_{cl}$VS sentences in Spanish. Also the role of individual differences (e.g., motivation) needs to be further explored in immersive contexts testing performance on non-canonical sentences since learners who take part in SA programs are self-selected, and usually more independent and motivated than learners in regular courses at home (Grey et. al., 2015). Furthermore, investigating whether the nature of the verb itself plays a role in cue agreement detectability is worth investigating: for example, comparisons involving different types of verbs (auxiliary vs. non-auxiliary) would shed light on the issue of salience and processability. In the same way, studies comparing whether specific agreement conditions (e.g., the mismatching plural verb condition, cl. sing., verb pl., and The mismatching plural clitic condition, cl. pl., verb sing.) have differential effects on a baseline condition (cl. sing., verb sing.) could be conducted employing blocks with a fixed order in order to determine whether $O_{cl}$VS sentences conveying contrastive-agreement (the mismatching plural verb condition, cl. sing., verb pl) impact the way a baseline condition is processed. Finally, more research is needed in relation to how learners process aural stimuli when agreement cues are manipulated: for this purpose eye-tracking would be a useful methodology to replicate the results found here with a visual-world study.

**Overall conclusions**

The results of overall accuracy in this dissertation have consistently shown that learners in their third year of Spanish coursework at the college level continue to rely on word order as the most valid cue for agenthood when interpreting $O_{cl}$VS sentences. This means that after years of exposure to the Spanish language, these learners still exhibit

difficulties interpreting $O_{cl}VS$ sentences correctly. These results extended Lee and Malovrh's (2009, 2010) findings in the aural modality. Yet, the tendency for learners to rely on word order does not seem to be entirely fixed. The experiments conducted in this dissertation have shown that the mismatching plural verb condition (cl sing., verb pl.), is qualitatively more beneficial for leaners than its counterpart, the mismatching plural clitic condition(cl. pl., verb sing.). The mismatching plural verb condition helped learners to make improvements as the experiment progressed. Improvements have also been observed in cases in which participants did not receive feedback. These findings support the idea that not all number contrasts are the same. Specifically, only plural number markers attached to the verb (a content word) rather than to the clitic proved to be more helpful in L2 cue strength readjustments. RT data have also showed that the mismatching plural verb condition (cl. sing., verb pl.) tended to trigger higher processing costs. Learners either spent more time in the verb region, or showed significant slowdowns across experimental trials in this condition. When the mismatching plural verb condition (cl sing., verb pl.) did not exhibit significant improvements across experimental trials, it did continue to exhibit good performance; that is, it was not significantly different from the baseline (cl. sing., verb sing.). Since learners have also been shown to perform well in the baseline condition (cl. sing., verb sing.) – for example, accuracy in this condition has shown improvements as a result of immersion experience and even more particularly under the influence of yes-no feedback – these results suggest that the baseline condition (cl. sing., verb sing.), which conveys a clitic in the singular form, may benefit from the mismatching plural verb condition (cl. sing. verb pl.). In other words, it may be the case that once learners start realigning their L2 cue strength due to the progressive exposure to

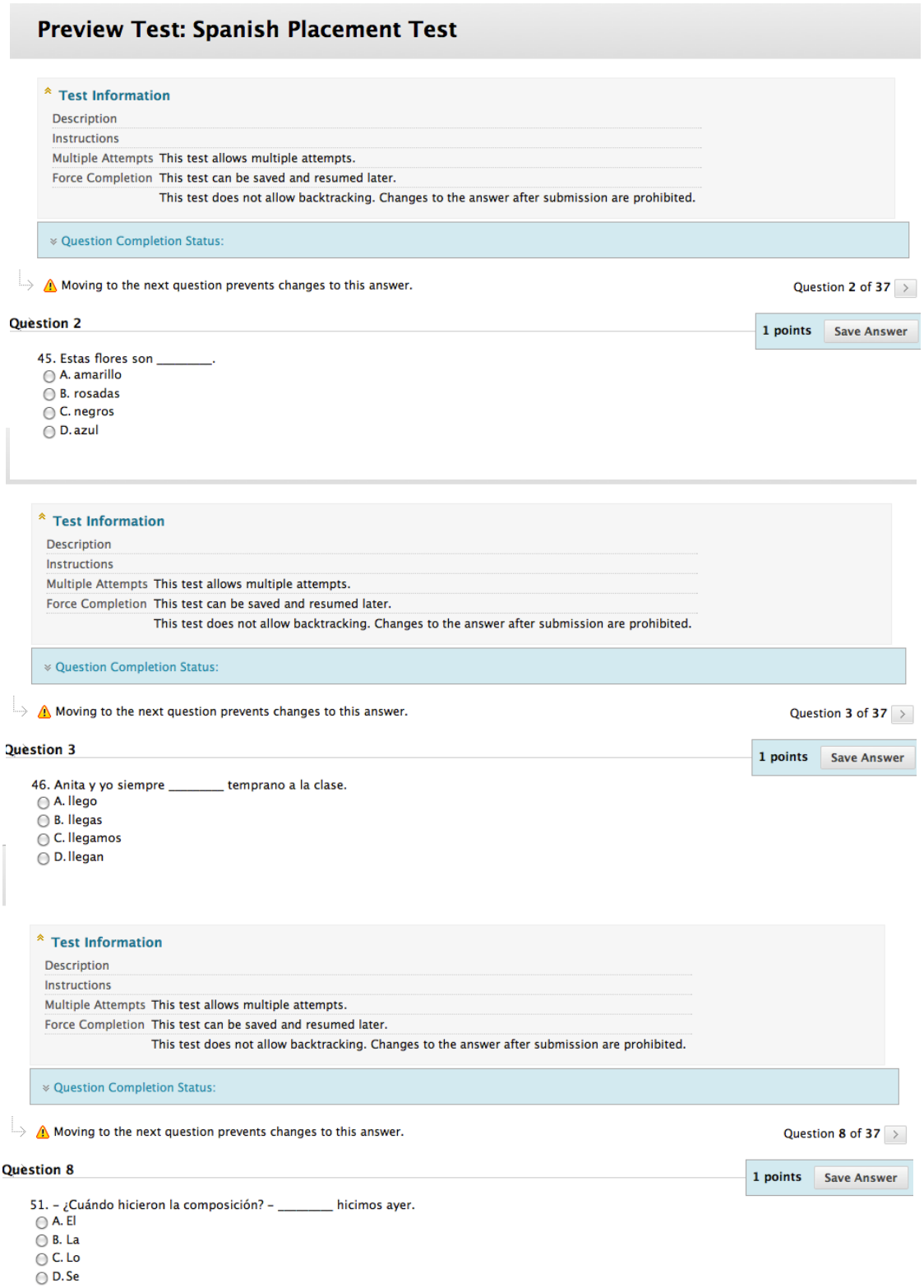
$O_{cl}$VS sentences in the mismatching plural verb condition (cl. sing., verb pl.), they may start using their altered L2 strategies to better categorize $O_{cl}$VS sentences in the baseline condition. The baseline condition (cl. sing., verb sing.) has been shown to be significantly less taxing for learners; that is, speedups have been observed for sentences in this condition.

The findings in this dissertation have also contributed to the debate as to whether short-term studies overseas influence L2 grammar skills. Results from learners who took part in a 5-week study program overseas have shown that immersion experience promotes a general improvement in $O_{cl}$VS sentence processing, that is, in all agreement conditions. In term of latency, even though learners did not get significantly faster when processing $O_{cl}$VS sentences after immersion, they showed qualitative differences in their processing patterns: mainly, RT data revealed that, in the baseline condition (cl. sing., verb sing.), learners tended to pay more attention to the verb and the post-verbal subject, highly informative regions for determining 'who does what to whom' in $O_{cl}$VS sentences.

Finally, this dissertation has contributed to shedding light on the extent to which exposure to manipulated input alone helps learners reconfigure their L2 processing strategies. Even though the provision of feedback significantly influences the *speed* with which learners readjust the L2 cue strength towards the L2 cue validity, participants who only had practice decoding manipulated input (matching/mismatching conditions), seemed to also benefit from such manipulation, specifically from the mismatching verb condition (cl. sing., verb pl.). One main difference for learners who do not receive feedback on their responses is that $O_{cl}$VS sentence processing for these learners takes

place more slowly. Potentially, in order to see bigger improvements over time, they require more exposure to the target sentences.

## APPENDIX A: Examples of grammar section of Spanish Placement Test



**Preview Test: Spanish Placement Test**

**Test Information**

Description
Instructions
Multiple Attempts This test allows multiple attempts.
Force Completion This test can be saved and resumed later.
This test does not allow backtracking. Changes to the answer after submission are prohibited.

Question Completion Status:

Moving to the next question prevents changes to this answer.

Question 2 of 37

**Question 2** 1 points Save Answer

45. Estas flores son _____.
- A. amarillo
- B. rosadas
- C. negros
- D. azul

**Test Information**

Description
Instructions
Multiple Attempts This test allows multiple attempts.
Force Completion This test can be saved and resumed later.
This test does not allow backtracking. Changes to the answer after submission are prohibited.

Question Completion Status:

Moving to the next question prevents changes to this answer.

Question 3 of 37

**Question 3** 1 points Save Answer

46. Anita y yo siempre _____ temprano a la clase.
- A. llego
- B. llegas
- C. llegamos
- D. llegan

**Test Information**

Description
Instructions
Multiple Attempts This test allows multiple attempts.
Force Completion This test can be saved and resumed later.
This test does not allow backtracking. Changes to the answer after submission are prohibited.

Question Completion Status:

Moving to the next question prevents changes to this answer.

Question 8 of 37

**Question 8** 1 points Save Answer

51. – ¿Cuándo hicieron la composición? – _____ hicimos ayer.
- A. El
- B. La
- C. Lo
- D. Se

## APPENDIX B: Bio questionnaire

### Background Questionnaire

***Please answer the following questions to the best of your knowledge. Please contact the researcher if any question is not clear. Thank you for your help!***

1. Which language do you consider your native language?
   Spanish English Both Other(s) _____
2. Which language do you consider your dominant language?
   Spanish English Both Other(s) _____

3. List all languages you know and rate your ability on the following aspects in each language. Please rate according to the following scale (write down the number in the table):

Very poor 1 Poor 2 Fair 3 Functional 4 Good 5 Very good 6 Native-like 7

| **Language** | **Reading ability** | **Writing ability** | **Speaking ability** | **Listening ability** |
|---|---|---|---|---|
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |

4. Provide the age at which you were first exposed to each language in terms of speaking, reading, and writing, and the number of years you have spent learning each language.

| **Language** | **Age first exposed to the language** | | | **Number of years learning (cumulative)** |
|---|---|---|---|---|
| | **Speaking** | **Reading** | **Writing** | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |

7. Estimate, in terms of percentages, how often you use English, Spanish and other languages per day (in all daily activities combined):

| | | | | | |
|---|---|---|---|---|---|
| Spanish | <25% | 25% | 50% | 75% | 100% |
| English | <25% | 25% | 50% | 75% | 100% |

Additional language: _____ <25% 25% 50% 75% 100%

Additional language: _____ <25% 25% 50% 75% 100%

Additional language: _____ <25% 25% 50% 75% 100%

8. If there is anything else that you feel is interesting or important about your language background or language use, please comment below.

_____
___

9. Do you have additional questions that you feel are not included above? If yes, please write down your questions and answers.

(6) How many years have you studied Spanish? _____

(7) How old were you when you started studying Spanish?
_____

| (8) Where have you studied Spanish? (check as many as needed) | How long? (years) | With a Native Spanish speaker? (yes / no) |
|---|---|---|
| ___ Kindergarten | _____ | _____ |
| ___ Elementary school | _____ | _____ |
| ___ Middle school _ | _____ | _____ |
| ___ High school | _____ | _____ |
| ___ Language School | _____ | _____ |
| ___ Private Tutoring | _____ | _____ |

(9) What Spanish classes are you taking now? (Class numbers and names)
_____

(10) What Spanish classes will you be taking next semester? (Class numbers and names)
_____

(11) Are you studying Spanish anywhere else now? Where? What are you studying?

_____

(12) How many hours do you spend per week using Spanish outside class to …

| | | | | |
|---|---|---|---|---|
| Do homework | 0 | 1-2 | 3-4 | 5-6 |

| | | | | |
|---|---|---|---|---|
| Prepare for quizzes & exams | 0 | 1-2 | 3-4 | 5-6 |
| Listen to language tapes | 0 | 1-2 | 3-4 | 5-6 |
| Read for fun | 0 | 1-2 | 3-4 | 5-6 |
| Listen to music | 0 | 1-2 | 3-4 | 5-6 |
| Watch TV, videos & movies | 0 | 1-2 | 3-4 | 5-6 |
| Talk to friends | 0 | 1-2 | 3-4 | 5-6 |
| Talk to family members | 0 | 1-2 | 3-4 | 5-6 |

(13) Have you ever been to a Spanish-speaking country (Spain, Mexico, Colombia, Chile, Argentina, another Latin-American country, etc.)?

Yes _____ No_____
If yes, how long were you there? _____ What did you do there? _____

_____

(14) Was Spanish ever spoken in your home when you were a child?

Yes _____ No_____

If yes, how much contact did you have with the language?

___Very little ___Some ___ A great deal ___ I consider myself bilingual

Age (in years): _____

Sex (circle one): Male / Female

Participant ID: _____

Today's Date: _____

**APPENDIX C: Exit questionnaire**

1) **What do you think the goal of this task was?**

2) **Did you find anything in this task interesting?**

3) **Did you find anything in this task difficult?**

4) **Did you learn anything by doing this task?**

## Appendix D: Elicited Imitation Task stimuli and scoring protocol

Appendices I and J from Ortega, L. (2000). *Understanding syntactic complexity: The measurement of change in the syntax of instructed L2 Spanish learners*. Unpublished doctoral dissertation. Department of Second Language Studies, University of Hawaii. Honolulu: HI.

### ELICITED IMITATION TASK STIMULI

Note: The test was developed by Ortega, Iwashita, Rabie, and Norris (in preparation).

1. Quiero cortarme el pelo (7 syllables)
2. El libro está en la mesa (7 syllables)
3. El carro lo tiene Pedro (8 syllables)
4 Él se ducha cada mañana (9 syllables)
5. ¿Qué dice usted que va a hacer hoy? (9 syllables)
6. Dudo que sepa manejar muy bien (10 syllables)
7. Las calles de esta ciudad son muy anchas (11 syllables)
8. Puede que llueva mañana todo el día (12 syllables)
9. Las casas son muy bonitas pero caras (12 syllables)
10. Me gustan las películas que acaban bien (12 syllables)
11. Después de cenar me fui a dormir tranquilo (13 syllables)
12. El chico con el que yo salgo es español (13 syllables)
13. Quiero una casa en la que vivan mis animales (14 syllables)
14. A vosotros os fascinan las fiestas grandiosas (14 syllables)
15. Ella ha terminado de pintar su apartamento (14 syllables)
16. El niño al que se le murió el gato está triste (14 syllables)
17. Ella sólo bebe cerveza y no come nada (15 syllables)
18. Me gustaría que el precio de las casas bajara (15 syllables)
19. Cruza a la izquierda y después sigue todo derecho (15 syllables)
20. Me gustaría que empezara a hacer más calor pronto (15 syllables)
21. Una amiga mía cuida a los niños de mi vecino (16 syllables)
22. El gato que era negro fue perseguido por el perro (16 syllables)
23. Antes de poder salir él tiene que limpiar su cuarto (16 syllables)
24. La cantidad de personas que fuman ha disminuido (17 syllables)
25. Después de llegar a casa del trabajo tomé la cena (17 syllables)
26. El ladrón al que cogió la policía era famoso (17 syllables)
27. Le pedí a un amigo que me ayudara con la tarea (16 syllables)
28. El examen no fue tan difícil como me habían dicho (17 syllables)
29. ¿Serías tan amable de darme el libro que está en la mesa? (17 syllables)
30. Me pregunto si el tren de las ocho habrá llegado ya o no (17 syllables)

**SCORING PROTOCOL FOR ELICITED IMITATION TASK**

Note: The scoring protocol was adapted from the scoring system developed by Ortega, Iwashita, Rabie, and Norris (in preparation).

Table J.1. EIT score 0 descriptor

| Criteria | Examples |
|---|---|
| • Nothing (Silence) | |
| • Garbled (unintelligible, usually transcribed as XXX) | |
| • Minimal repetition, then item abandoned:<br>- Only 1 word repeated<br>- Only 1 content word plus function word(s)<br>- Only function word(s) repeated<br>- Only 1 or 2 content words out of order plus extraneous words that weren't in the original stimulus | - Manana (10- item 4)<br>- El examen que [gibberish] (09- item 28)<br>- Despues mue- XX tranquilo (01-item 11)<br>- Tu que sepa a- m- muy bien (12-item 6)<br>- Me gustaria las se se se el XXX (16-item 18) |

Table J.2. EIT score 1 descriptor

| Criteria | Examples |
|---|---|
| • When only about half of idea units are represented in the string but a lot of important information in the original stimulus is left out; sometimes the resulting meaning is unrelated (or opposed) to stimulus | -Antes de poder seguir (3 sec.) perdio su cuarto (02-item 23)<br>-Dudo que sepa ma- tambien (04-item 6)<br>-Seria en que el libro esta en la mesa (11-item 29)<br>-El gato que eran pelo negro dan something el perro (14-item 22) |

| | |
|---|---|
| • Or when string doesn't in itself constitute a self-standing sentence with some (related or not to stimulus) meaning (This may happen when only 2 of 3 content words are repeated and no grammatical relation between them is attempted) | -El ladron que XX la policia famoso (11-item 26)<br>-Despues de cenar fue en- tranquilo (03-item 11)<br>-Le pendu una amiga que XXX la tarea (01-item 27)<br>-La cantidad de personas fumar alguno, alguno (03-item 24) |



Table J.3. EIT score 2 descriptor

| Criteria | Examples |
|---|---|
| • When content of string preserves at least more than half of the idea units in the original stimulus; string is meaningful, and the meaning is close or related to original, but it departs from it in some slight changes in content, which makes content inexact, incomplete, or ambiguous | -Despues de cenar me fui a X tranquilo (11-item 12)<br>-Ella sola cerveza y no come nada (05-item 17)<br>-Quieres una casa que viven los alemanes an-animales (07-item 13)<br>-El chico con lo que es algos es espanol (08-item 12)<br>-El chico con yo salgo es muy bien (15-item 12)<br>-Despues a trabajo tome la cena (16-item 25) |

Table J.4. EIT score 3 descriptor

| Criteria | Examples |
|---|---|

| | |
|---|---|
| • Original, complete meaning is preserved as in the stimulus. Strings which are ungrammatical can get a 3 score, as long as exact meaning is preserved. Some synonymous substitutions are acceptable:<br><br>-Anything with or without ‘’muy’ (very’) should be considered synonymous.<br>-Substitutions of ‘y’/’pero’ (and & but) are acceptable.<br><br>• Changes in grammar that don’t affect meaning should be scored as 3. (Ambiguous changes in grammar that could be interpreted as meaning changes from a NS perspective should be scored as 2. That is, as a general principle in case of doubt about whether meaning has changed or not, score 2.) | -Me gustaria el precio de las casas baraja (2 sec.) baja (15-item 18)<br>-El nino que se m- murio cato esta triste (02-item 16)<br>-[gibberish] se ducha cada manana (18-item 4)<br>-Quiero cortar mi pelo (05-item 1)<br>-Las calles de esta ciudad son anchas (13-item 7)<br>-El chico que yo salgo es espanol (06-item 11)<br>-El chico con el salgo es espanol (05-item 11)<br>-El examen no fuen tan dificil come han di- como me han dicho (12-item 28)<br>-Las casa son muy bonitas pero caras (07-item 9)<br>-Quiero una casa en que viven mis animales (12-item 13)<br>-Dudo que saba a ma- manejar muy bien (11-item 6)<br>-Ella he terminado X pintar sus apartamiento (11-item 15) |

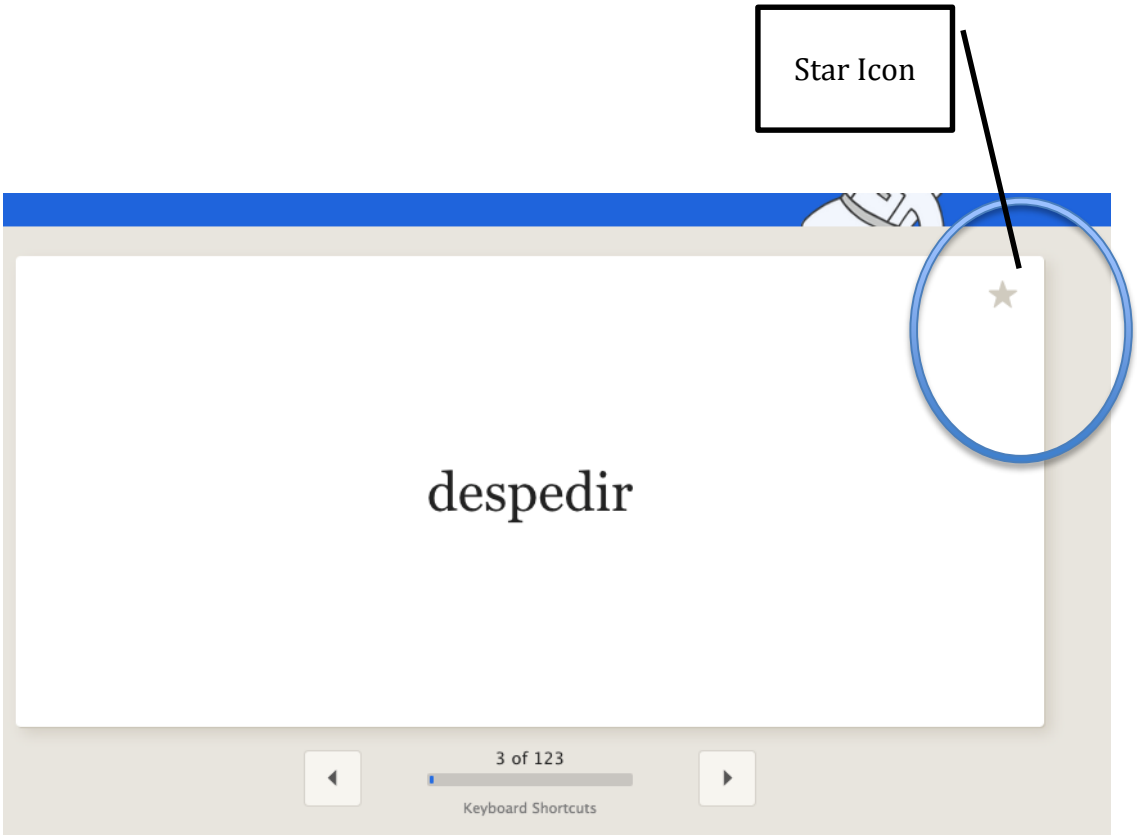


Table J.5. EIT score 4 descriptor

| Criteria | Examples |
|---|---|
| • Exact repetition: String matches stimulus exactly. Both form and meaning are correct without exception or doubt. | |

**Appendix E: Vocabulary test in Quizlet (example)**

**Appendix F: Sample of words in Quizlet**



**Appendix G: Vocabulary answer sheet**

Participants' research number: _____

*Instructions*: After completing the vocabulary task in Quizlet, please rate the difficulty of that vocabulary. Please, **mark your answer with an X**.

*(1) How difficult was this vocabulary for you?*

- Very easy:
- Easy:
- Moderate:
- Difficult:
- Very difficult:

(2) *Which words did you not know beforehand?*

## Bibliography


Alarcón, I. (2009). The processing of gender agreement in L1 and L2 Spanish: Evidence from reaction time data. *Hispania*, 92, 814-828.

Anton-Mendez, I. (1996). *Clitics and attraction errors: An experimental study on language production.(*Unpublished Master Thesis). University of Arizona, Arizona.

Avello, P., & Lara, A. R. (2014). Phonological development in L2 speech production during study abroad programmes differing in length of stay. In C. Pérez-Vidal (Ed), *Second Language Acquisition in Study Abroad and Formal Instruction Contexts* (pp. 137–166). Philadelphia: John Benjamins.

Bangs, P. (2002, August). *Why is feedback dying of starvation? Let's try to revive it.* Paper presented at the EuroCall Conference, Jyväskylä, Finland.

Bañón, J. A., Fiorentino, R., & Gabriele, A. (2014). Morphosyntactic processing in advanced second language (L2) learners: An event-related potential investigation of the effects of L1–L2 similarity and structural distance. *Second Language Research*, 30(3), 275-306.

Barber, H., & Carreiras, M. (2005). Grammatical gender and number agreement in Spanish: An ERP comparison. *Journal of Cognitive Neuroscience*, 17(1), 137-153.

Barlow, M., & Ferguson, C. A. (1988). *Agreement in natural language: Approaches, theories, descriptions Center for the Study of Language* (CSLI).

Bates, E., & McWhinney, B. (1979). A functionalist approach to the acquisition of grammar. In E. Ochs, & B. Schieffelin (Ed.), *Developmental pragmatics*. New York: Academic Press.

Bates, E., & Elman, J. L. (1993). Connectionism and the study of change. In Mark Johnson (Ed.), *Brain development and cognition: A reader* (pp. 623-642). Oxford: Blackwell Publishers.

Bates, E., & MacWhinney, B. (1981). Second- language acquisition from a functionalist perspective: Pragmatic, semantic, and perceptual strategies. *Annals of the New York Academy of Sciences*, 379(1), 190-214.

Bates, E., & MacWhinney, B. (1982). Functionalist approaches to grammar. In E. Wanner, & L. Gleitman (Ed.), *Language acquisition: The state of the art* (pp. 173-218). New York: Cambridge University Press.

Bates, E., & MacWhinney, B. (1987). Competition, variation, and language learning. In B.MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 157-193). Hillsdale, New Jersey: Erlbaum,

Bates, E., & MacWhinney, B. (1989). Functionalism and the competition model. In B. MacWhinney & E. Bates (Eds.), *The crosslinguistic study of sentence processing*. New York: Cambridge University Press.

Bates, E., MacWhinney, B., Caselli, C., Devescovi, A., Natale, F., & Venza, V. (1984). A cross-linguistic study of the development of sentence interpretation strategies. *Child Development*, 55, 341-354.

Bates, E., McNew, S., MacWhinney, B., Devescovi, A., & Smith, S. (1982). Functional constraints on sentence processing: A cross-linguistic study. *Cognition*, 11(3), 245-299.

Bever, T. G. (1970). The cognitive basis for linguistic structures. In R. Hayes (Ed.), *Cognition and language development* (pp. 279-362). New York: Wiley & Sons.

Bock, K., & Miller, C. A. (1991). Broken agreement. *Cognitive Psychology*, 23(1), 45-93.

Bowles, M. A. (2005). *Effects of Verbalization Condition and Type of Feedback on L2 Development in a CALL Task*. (Unpublished doctoral dissertation). Georgetown University, Washington, DC.

Cadierno, T. (1995). Formal instruction from a processing perspective: An investigation into the Spanish past tense. *The Modern Language Journal*, 79(2), 179-193.

Calenge, C. (2006). The package "adehabitat" for the R software: A tool for the analysis of space and habitat use by animals. *Ecological Modeling*, 197(3), 516-519.

Camblor-Portilla, M. T. (2006). *Type of Written Feedback, Awareness and L2 Development: A Computer-Based Study*. (Unpublished doctoral dissertation). Georgetown University, Washington, DC.

Carpenter, H., Jeon, K. S., MacGregor, D., & Mackey, A. (2006). Learners' interpretations of recasts. *Studies in Second Language Acquisition*, 28(02), 209-236.

Carroll, S., & Swain, M. (1993). Explicit and implicit negative feedback. *Studies in Second Language Acquisition*, 15(03), 357-386.

Carroll, S., Swain, M., & Roberge, Y. (1992). The role of feedback in adult second language acquisition: Error correction and morphological generalizations. *Applied Psycholinguistics*, 13(02), 173-198.

Cazden, C., & Brown, R. (1975). The early development of the mother tongue. *Child Language: A Book of Readings*. Englewood Cliffs, New Jersey: Prentice-Hall.

Chapelle, C. (2001). *Computer applications in second language acquisition*. New York: Cambridge University Press.

Chapelle, C. (2009). The relationship between second language acquisition theory and Computer- Assisted language learning. *The Modern Language Journal, 93*(1), 741-753.

Christiansen, M. H., Chater, N., & Seidenberg, M. S. (1999). Special issue-connectionist models of human language processing: Progress and prospects. *Cognitive Science, 23*(4), 415-415.

Clahsen, H. (1986). Verb inflections in German child language: Acquisition of agreement markings and the functions they encode. *Linguistics, 24*(1), 79-122.

Clahsen, H., & Penke, M. (1992). The acquisition of agreement morphology and its syntactic consequences: New evidence on German child language from the Simone-Corpus. *The acquisition of verb placement* (pp. 181-223). New York: Springer.

Clifton, C., & Staub, A. (2011). Syntactic influences on eye movements during reading. *Eye, 3*, 2.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd ed.).* Hillsdale, New Jersey: Lawrence Erlbaum Associates

Collentine, J. (2004). The effects of learning contexts on morphosyntactic and lexical development. *Studies in Second Language Acquisition, 26*(02), 227-248.

Corbett, G. (1991). Gender. *Cambridge Textbooks in Linguistics*. Cambridge: Cambridge University Press.

Corbett, G. G. (1988). Gender in Slavonic from the standpoint of a general typology of gender systems. *The Slavonic and East European Review, 66*, 1-20.

Corbett, G. G. (1989). An approach to the description of gender systems. In D. Arnold, M. Atkinson, J. Durand, C. Grover, and L. Sadler (Eds.), *Essays on Grammatical Theory and Universal Grammar* (pp. 53 - 89). Oxford: Clarendon Press.

De Garavito, J. (2003). *Learners' competence may be more accurate than we think: Spanish L2 and agreement morphology*. In Liceras J. M. et al., Proceedings of the 6th Generative Approaches to Second Language Acquisition Conference (pp.17-23). Somerville, MA: Cascadilla Press.

DeKeyser, R. (2014). Methodological considerations about research on language development during study abroad. In C. Pérez-Vidal (Ed), *Second Language Acquisition in Study Abroad and Formal Instruction Contexts* (pp. 313-326). Philadelphia: John Benjamins.

DeKeyser, R. (1991). Foreign language development during a semester abroad. In B. Freed (Ed.), *Foreign Language Acquisition Research and the Classroom* (pp. 104-119). Lexington, M.A.: DC Heath.

DeKeyser, R. (2007). Study abroad as foreign language practice. In. R. DeKeyser (Ed.), *Practice in a Second Language: Perspectives from Applied Linguistics and Cognitive Psychology* (pp. 208-226). Cambridge: Cambridge University Press.

DeKeyser, R. (2010). Monitoring processes in Spanish as a second language during a study abroad program. *Foreign Language Annals*, *43*(1), 80-92.

Despain, S. (1997). *The effects of two delivery systems for listening comprehension exercises on the language performance and attitude of beginning Spanish students.* (Unpublished doctoral dissertation). Indiana University, Bloomington, Indiana.

Dewey, D. P. (2008). Japanese vocabulary acquisition by learners in three contexts. *Frontiers: The Interdisciplinary Journal of Study Abroad, 15*, 127-148.

Díaz-Campos, M. (2004). Context of learning in the acquisition of Spanish second language phonology. *Studies in Second Language Acquisition, 26*(02), 249-273.

Dillon, B., Levy, J., Staub, A., & Clifton, C. (2014, March). *Linear order effects in agreement: Evidence from English wh-questions*. Paper presented at the 27th annual CUNY Human Sentence Processing Conference, Columbus, Ohio.

Domínguez, L. (2014). Subject inversion in non-native Spanish. *Lingua*, *145*, 243; 243-265; 265.

Döpke, S. (2000). Generation of and retraction from cross-linguistically motivated structures in bilingual first language acquisition. Bilingualism: Language and *Cognition, 3*(03), 209-226.

Doughty, C. (1987). Relating second-language acquisition theory to CALL research and application. In W.F. Smith (Ed.), *Modern media in foreign language education: Theory and implementation* (pp. 133-167). Lincolnwood, IL: National Textbook Company.

Dowens, M. G., Vergara, M., Barber, H. A., & Carreiras, M. (2010). Morphosyntactic processing in late second-language learners. *Journal of Cognitive Neuroscience*, *22*(8), 1870-1887.

Duffield, N., & White, L. (1999). Assessing L2 knowledge of Spanish clitic placement: Converging methodologies. *Second Language Research, 15*(2), 133-160.

Duperron, L. (2006). Study abroad and the second language acquisition of tense and aspect in French: Is longer better? In S. Wilkinson (Ed.), *Insights from Study Abroad for Language Programs* (pp. 47-51). Boston: Heinle.

Dussias, P. E., Valdés Kroff, J. R., Guzzardo Tamargo, R. E., & Gerfen, C. (2013). When gender and looking go hand in hand. *Studies in Second Language Acquisition, 35*(02), 353-387.

Dussias, P. E. (2010). Uses of eye-tracking data in second language sentence processing research. *Annual Review of Applied Linguistics, 30,* 149-166.

Ellis, N. C. (1998). Emergentism, connectionism and language learning. *Language Learning, 48(*4), 631-664.

Ellis, N. C. (2003). Constructions, chunking, and connectionism: The emergence of second language structure. In C. Doughty and M.H. Long (Eds.), *The Handbook of Second Language Acquisition* (pp. 63-103). Oxford: Blackwell Publishing.

Ellis, N. C., & Sagarra, N. (2010). Learned attention effects in L2 temporal reference: The first hour and the next eight semesters. *Language Learning, 60*(2), 85-108.

Ellis, R., & Sheen, Y. (2006). Reexamining the role of recasts in second language acquisition. *Studies in Second Language Acquisition, 28*(4), 575-600.

Elman, J. L. (2001). Connectionism and language acquisition. In M. Tomasello and E. Bates (Eds.), *Language Development: The Essential Readings* (pp. 295-306). Oxford: Blackwell.

Elman, J. (2003). Development: It's about time. *Developmental Science, 6*(4), 430-433.

Elman, J. L. (2005). Connectionist models of cognitive development: Where next? *Trends in Cognitive Sciences, 9*(3), 111-117.

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking Innateness*. Cambridge, MA: MIT Press.

Farley, A. P. (2001). The effects of processing instruction and meaning-based output instruction. *Spanish Applied Linguistics, 5*(1), 57-94.

Feingold, A. (2013). A Regression Framework for Effect Size Assessments in Longitudinal Modeling of Group Differences. *Review of General Psychology : Journal of Division 1, of the American Psychological Association, 17*(1), 111–121.

Felix, U. (2002). The web as a vehicle for constructivist approaches in language teaching. *Recall, 14*(1), 2-15.

Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language,* 25(3), 348-368.

Fleiss, J., & Berlin, J. (2009). Effect sizes for dichotomous data. In H. Cooper & L.V. Hedges (Eds.), *The Handbook of Research Synthesis and Meta-Analysis* (pp. 237-253). New York: Russell Sage Foundation.

Foote, R. (2011). Integrated knowledge of agreement in early and late English–Spanish bilinguals. *Applied Psycholinguistics,* 32(1), 187-220.

Foucart, A. (2008). *Grammatical gender processing in French as a first and as a second language.* (Unpublished doctoral dissertation). Université Aix-Marseille I, France and University of Edinburgh, Scotland.

Freed, B. F. (1995). *Second language acquisition in a study abroad context*. Philadelphia: John Benjamins Publishing.

Freed, B. F., Segalowitz, N., & Dewey, D. P. (2004). Context of learning and second language fluency in French: Comparing regular classroom, study abroad, and intensive domestic immersion programs. *Studies in Second Language Acquisition*, 26(2), 275-301.

Garnsey, S. M., Pearlmutter, N. J., Myers, E., & Lotocky, M. A. (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language*, 37(1), 58-93.

Grey, S., Cox, J. G., Serafini, E. J., & Sanz, C. (2015). The role of individual differences in the study abroad context: Cognitive capacity and language development during Short- Term intensive language exposure. *The Modern Language Journal*, 99(1), 137-157.

Grüter, T., Lew-Williams, C., & Fernald, A. (2012). Grammatical gender in L2: A production or a real-time processing problem? *Second Language Research*, 28(2), 191-215.

Guntermann, G. (1992). An analysis of interlanguage development over time: Part II, ser and estar. *Hispania*, 75, 1294-1303.

Guntermann, G. (1995). The Peace Corps experience: Language learning in training and in the field. In B. Freed (Ed.), *Second Language Acquisition in a Study Abroad Context, (Studies in Bilingualism 9*) (pp. 149-169). Amsterdam: John Benjamins.

Haddock, C. K., Rindskopf, D., & Shadish, W. R. (1998). Using odds ratios as effect sizes for meta-analysis of dichotomous data: A primer on methods and issues. *Psychological Methods*, 3(3), 339.

Han, Z. (2002). A study of the impact of recasts on tense consistency in L2 output. *TESOL Quarterly*, 36(4), 543-572.

Han, Z. (2008). On the role of meaning in focus on form. In Z. Han (Ed.) *Understanding Second Language Process* (pp. 45-79). New York: Multilingual Matters.

Harrington, M. (1987). Processing transfer: Language-specific processing strategies as a source of interlanguage variation. *Applied Psycholinguistics*, 8(4), 351-377.

Hartsuiker, R. J., Antón-Méndez, I., & van Zee, M. (2001). Object attraction in subject-verb agreement construction. *Journal of Memory and Language*, 45(4), 546-572.

Haskell, T. R., & MacDonald, M. C. (2003). Conflicting cues and competition in subject–verb agreement. *Journal of Memory and Language*, 48(4), 760-778.

Haskell, T. R., & MacDonald, M. C. (2005). Constituent structure and linear order in language production: Evidence from subject-verb agreement. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(5), 891.

Heift, T. (2001). Error-specific and individualised feedback in a web-based language tutoring system: Do they read it? *Recall, 13*(1), 99-109.

Heift, T. (2004). Corrective feedback and learner uptake in CALL. *Recall, 16*(2), 416-431.

Heift, T. (2010). Prompting in CALL: A longitudinal study of learner uptake. *The Modern Language Journal, 94*(2), 198-216.

Heilenman, L. K. and McDonald, J. L. (1993), Processing Strategies in L2 Learners of French: The Role of Transfer. *Language Learning*, *43*, 507–557.

Hernandez, A. E., Bates, E. A., & Avila, L. X. (1994). On-line sentence interpretation in Spanish–English bilinguals: What does it mean to be "in between"? *Applied Psycholinguistics*, *15*(4), 417-446.

Hernandez, A. E., Sierra, I., & Bates, E. A. (2000). Sentence interpretation in bilingual and monolingual Spanish speakers: Grammatical processing in a monolingual mode. *Spanish Applied Linguistics, 4*(2), 179-213.

Hernández, T. A. (2010). The relationship among motivation, interaction, and the development of second language oral proficiency in a study- abroad context. *The Modern Language Journal, 94*(4), 600-617.

Hofmeister, Philip. 2011. "Representational Complexity and Memory Retrieval in Language Comprehension." *Language and Cognitive Processes 26*(3): 376-405.

Hoshino, N., Dussias, P. E., & Kroll, J. F. (2010). Processing subject–verb agreement in a second language depends on proficiency. *Bilingualism: Language and Cognition, 13*(2), 87-98.

Houston, T. (1997). Sentence processing in Spanish as a second language: A study of word order and background knowledge. In W.R. Glass & A.T. Pérez-Leroux (Eds.), *Contemporary Perspectives on the acquisition of Spanish volume 2: Production, processing and Comprehension* (pp. 123-134). Somerville: Cascadilla Press.

Howard, M. (2001). The effects of study abroad on the L2 learner's structural skills: Evidence from advanced learners of French. *Eurosla Yearbook, 1*(1), 123-141.

Hughes, A., Wilkens, T., Wildemuth, B. M., & Marchionini, G. (2003). Text or pictures? an eyetracking study of how people view digital video surrogates. *Image and Video Retrieval* (pp. 271-280) New York: Springer.

Humphreys, K. R., & Bock, K. (2005). Notional number agreement in English. *Psychonomic Bulletin & Review, 12*(4), 689-695.

Ife, A., Vives Boix, G., & Meara, P. (2000). The impact of study abroad on the vocabulary development of different proficiency groups. *Spanish Applied Linguistics, 4*(1), 55-84.

Ionin, T. (2013). Review article: Recent publications on research methods in second language acquisition. *Second Language Research, 29*(1), 119.

Isabelli, C. A., & Nishida, C. (2005). Development of the Spanish subjunctive in a nine-month study-abroad setting. In D. Eddignton, (Ed.), *Selected Proceedings of the 6th Conference on the Acquisition of Spanish and Portuguese as First and Second Languages* (pp.78-91). Somerville MA: Cascadilla Press.

Isabelli, C. (2003). Development of oral communication skills abroad. *Frontiers: The Interdisciplinary Journal of Study Abroad,* 149-173.

Isabelli- García, C. (2010). Acquisition of Spanish gender agreement in two learning contexts: Study abroad and at home. *Foreign Language Annals, 43*(2), 289-303.

Jackson, J. (2013). The transformation of 'a frog in the well': A path to a more intercultural, global mindset. In C. Kinginger (Ed.), *The social turn in study abroad research* (pp. 179-204). Amsterdam: John Benjamins.

Jegerski, J. (2014). Self-paced reading. In J. Jegerski & B. VanPatten, B. (Eds.), *Research methods in second language psycholinguistics* (pp. 20-49). New York: Routledge.

Jiang, N. (2004). Morphological insensitivity in second language processing. *Applied Psycholinguistics, 25*(4), 603-634.

Jiang, N. (2007). Selective integration of linguistic knowledge in adult second language learning. *Language Learning, 57*(1), 1-33.

Jiang, N. (2013). *Conducting reaction time research in second language studies*. New York: Routledge.

Juan-Garau, M., Salazar-Noguera, J., & Prieto-Arranz, J. I. (2014). English L2 learners' lexico-grammatical and motivational development at home and abroad. In C. Pérez-Vidal (Ed.), *Language Acquisition in Study Abroad and Formal Instruction Contexts* (pp. 235-258). Amsterdam: John Benjamins.

Juffs, A. (2004). Representation, processing and working memory in a second language. *Transactions of the Philological Society, 102*(2), 199-225.

Juffs, A., & Harrington, M. (1995). Parsing effects in second language sentence processing. *Studies in Second Language Acquisition, 17*(04), 483-516.

Juffs, A., & Harrington, M. (1996). Garden path sentences and error data in second language sentence processing. *Language Learning, 46*(2), 283-323.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87*(4), 329.

Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General, 111*(2), 228.

Kail, M. (1989). Cue validity, cue cost, and processing types in sentence comprehension in French and Spanish. In B. MacWhinney & E. Bates (Eds.), *The Cross-Linguistic Study of Sentence Processing* (pp. 77 – 117). New York: Cambridge University Press.

Kail, M., & Charvillat, A. (1988). Local and topological processing in sentence comprehension by French and Spanish children. *Journal of Child Language, 15*(3), 637-662.

Kan, I. P., & Thompson-Schill, S. L. (2004). Effect of name agreement on prefrontal activity during overt and covert picture naming. *Cognitive, Affective, & Behavioral Neuroscience, 4*(1), 43-57.

Keating, G. D. (2009). Sensitivity to violations of gender agreement in native and nonnative Spanish: An Eye- Movement investigation. *Language Learning, 59*(3), 503-535.

Kennedy, A., & Murray, W. S. (1984). Inspection times for words in syntactically ambiguous sentences under three presentation conditions. *Journal of Experimental Psychology: Human Perception and Performance, 10*(6), 833.

Kilborn, K., & Cooreman, A. (1987). Sentence interpretation strategies in adult Dutch–English bilinguals. *Applied Psycholinguistics, 8*(4), 415-431.

Kilborn, K., & Ito, T. (1989). Sentence processing strategies in adult bilinguals. In B. MacWhinney & E. Bates (Eds.), *The crosslinguistic study of sentence processing* (pp. 257-291). New York: Cambridge University Press.

Kinginger, C. (2010). American students abroad: Negotiation of difference? *Language Teaching, 43*(2), 216-227.

Kouider, S., Halberda, J., Wood, J., & Carey, S. (2006). Acquisition of English number marking: The singular-plural distinction. *Language Learning and Development, 2*(1), 1-25.

Krashen, S. D. (1985). *The Input Hypothesis: Issues and implications*. London: Longman.

Kroll, J. F., Dussias, P. E., Bogulski, C. A., & Valdes Kroff, J. R. (2012). Juggling two languages in one mind: What bilinguals tell us about language processing and its consequences for cognition. In B. Ross (Ed.), *The Psychology of Learning and Motivation, Volume 56* (pp. 229-262). San Diego: Academic Press.

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language, 33*(2), 149-174.

LaBrozzi, R. M. (2009). *Processing of Lexical and Morphological Cues in a Study Abroad Context.* (Unpublished doctoral dissertation). Pennsylvania State University, Pennsylvania.

LaBrozzi, R. (2012). The role of study abroad and inhibitory control on processing redundant cues. In K. Geeslin & M. Díaz-Campos (Eds.), *Selected proceedings of the 14th Hispanic Linguistics Symposium* (pp. 228-241). Somerville, MA: Cascadilla Proceedings Project.

Lachman, R., Lachman, J. L., & Butterfield, E. C. (1979). *Cognitive psychology and information processing: An introduction.* Hillsdale, New Jersey: Erlbaum

Lafford, B. A. (2006). The effects of study abroad vs. classroom contexts on Spanish SLA: Old assumptions, new insights and future research directions. In Klee, C. A. & Face T.L (Eds), *Selected Proceedings of the 7th Conference on the Acquisition of Spanish and Portuguese as First and Second Languages* (pp. 1-25). Somerville, MA: Cascadilla Press.

Lago, S., Shalom, D. E., Sigman, M., Lau, E. F., & Phillips, C. (2015). Agreement attraction in Spanish comprehension. *Journal of Memory and Language, 82*, 133-149.

Lee, J. F. (1987). Morphological Factors Influencing Pronominal Reference Assignment by Learners of Spanish. In T. Morgan, J.F. Lee & B. VanPatten (Eds.), *Language and Language Use: Studies in Spanish* (pp. 221-232). Landham: University Press of America.

Lee, J. F. (2000). Comprehending subject-object relations while processing object pronouns in written input. In R.P. Leow & C. Sanz.(Eds.), *Spanish Applied Linguistics at the Turn of the Millennium: Papers from the 1999 conference on the L1 & L2 acquisition of Spanish and Portuguese* (pp. 119-140). Somerville, MA: Cascadilla Press.

Lee, J. F. (2003). Cognitive and linguistic perspectives on the acquisition of object pronouns in Spanish as a second language. In B. Lafford & R. Salaberry (Eds.), *Spanish and Second Language Acquisition: State of the Science* (pp. 98-129). Washington, DC: Georgetown University Press.

Lee, J. F., & Malovrh, P. A. (2009). Linguistic and non-linguistic factors affecting OVS processing of accusative and dative case pronouns by advanced L2 learners of Spanish. Paper presented at the Selected Proceedings of the 11th Hispanic Linguistics Symposium. Ed. Joseph Collentine, Maryellen García, Barbara Lafford, and Francisco Marcos Marín. Somerville, MA: Cascadilla, 105-116.

Leeman, J. (2003). Recasts and second language development. *Studies in Second Language Acquisition, 25*(1), 37.

Lehmann, C. (1988). On the function of agreement. In M. Barlow & C.A. Ferguson (Eds.), *Agreement in natural language: Approaches, theories, descriptions* (pp. 55-65). Stanford: Center for the Study of Language and Information.

Levy, J. P., Bairaktaris, D., Bullinaria, J. A., & Cairns, P. (2014). *Connectionist models of memory and language.* New York: Psychology Press.

Lew-Williams, C., & Fernald, A. (2010). Real-time processing of gender-marked articles by native and non-native Spanish speakers. *Journal of Memory and Language, 63(*4), 447-464.

Leow, R. P. (1993). To simplify or not to simplify: A look at intake. *Studies in Second Language Acquisition, 15*(3), 333-355.

Leow, R. P. (1995). Modality and intake in second language acquisition. *Studies in Second Language Acquisition, 17*(1), 79-89.

Leow, R. P., Grey, S., Marijuan, S., & Moorman, C. (2014). Concurrent data elicitation procedures, processes, and the early stages of L2 learning: A critical overview. *Second Language Research, 30*(2), 111-127.

Li, P., Zhang, F., Tsai, E., & Puls, B. (2014). Language history questionnaire (LHQ 2.0): A new dynamic web-based research tool. *Bilingualism: Language and Cognition, 17*(3), 673-680.

Li, P., & Zhao, X. (2012). *Connectionism*. New Jersey: Wiley Online Library.

Li, S. (2010). The effectiveness of corrective feedback in SLA: A meta-analysis. *Language Learning, 60*(2), 309-365.

Liceras, J. M. (1985). The value of clitics in non-native Spanish. *Second Language Research, 1*(2), 151-168.

Liceras, J. M., Maxwell, D., Laguardia, B., Fernández, Z., Fernández, R., and Diaz, L. (1997). A longitudinal study of Spanish non-native grammars: Beyond parameters.

In A.T. Pérez-Leroux & W.R. Glass (Eds.), *Contemporary perspectives on the acquisition of Spanish, vol. 1, Developing grammars* (pp. 99-132). Somerville, MA: Cascadilla Press.

Linck, J. A., Kroll, J. F., & Sunderman, G. (2009). Losing access to the native language while immersed in a second language: Evidence for the role of inhibition in second-language learning. *Psychological Science, 20*(12).

Liversedge, S. P., Paterson, K. B., & Pickering, M. J. (1998). Eye movements and measures of reading time. In Underwood, G. (Ed.), *Eye guidance in reading and scene perception* (pp. 55-75). New York: Elsevier.

Llanes, À. (2011). The many faces of study abroad: An update on the research on L2 gains emerged during a study abroad experience. *International Journal of Multilingualism, 8*(3), 189-215.

LoCoco, V. (1982). The role of word order as a cue in L2 comprehension. Paper presented at the 11th Annual Linguistics Symposium. University of Wisconsin-Milwaukee, March.

LoCoco, V. (1987).Learner comprehension of oral and written sentences in German and Spanish: The importance of word order. In B. VanPatten, T. Dvorak, & J.F. Lee (Eds.), *Foreign language learning: A research perspective* (pp. 119-129). Rowley, MA: Newbury House.

Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In Ritchie, W. C., & Bhatia, T. K. (Eds.), *Handbook of Second Language Acquisition* (pp. 413-468). London: Academic Press.

Long, M. H. (1998). Focus on form theory, research, and practice. In M. H. Long & P Robinson (Eds.), *Focus on Form in Classroom Second Language Acquisition* (pp. 15-41). Stuttgart: Ernst Klett Sprachen.

Long, M. H. (2007). Recasts in SLA: The story so far. In M.H. Long (Ed.), *Problems in SLA* (pp. 75-116). Mahwah: Erlbaum.

Long, M. H., Inagaki, S., & Ortega, L. (1998). The role of implicit negative feedback in SLA: Models and recasts in Japanese and Spanish. *The Modern Language Journal, 82*(3), 357-371.

Lyster, R. (2004). Differential effects of prompts and recasts in form-focused instruction. *Studies in Second Language Acquisition, 26*(3), 399-432.

Lyster, R., & Saito, K. (2010). Oral feedback in classroom SLA. *Studies in Second Language Acquisition, 32*(2), 265-302.

Mackey, A., Gass, S., & McDonough, K. (2000). How do learners perceive interactional feedback? *Studies in Second Language Acquisition, 22*(4), 471-497.

Mackey, A., & Goo, J. (2007). Interaction research in SLA: A meta-analysis and research synthesis. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A collection of empirical studies* (pp. 407-452). Oxford: Oxford University Press.

Mackey, A., & Philp, J. (1998). Conversational interaction and second language development: Recasts, responses, and red herrings? *The Modern Language Journal, 82*(3), 338-356.

MacWhinney, B. (1978). The acquisition of morphophonology. *Monographs of the Society for Research in Child Development, 43*, 1-123.

MacWhinney, B. (1992). Transfer and competition in second language learning. *Advances in Psychology, 83*, 371-390.

MacWhinney, B. (1997).Second language acquisition and the Competition Model. In J. Kroll & A. De Groot (Eds.), *Tutorials in Bilingualism: Psycholinguistic Perspectives* (pp.113-142). Mahwah, NJ: Lawrence Erlbaum.

MacWhinney, B. (1999). *The Emergence of Language.* New Jersey: Taylor & Francis.

MacWhinney, B. (2002). Extending the competition model. *Advances in Psychology, 134*, 31-57.

MacWhinney, B. (2005). Extending the competition model. *International Journal of Bilingualism*, *9*(1), 69-84.

MacWhinney, B. (2008). A unified model. In P. Robinson & N. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition*. Mahwah, NJ: Lawrence Erlbaum Associates

MacWhinney, B. (2012). The logic of the Unified Model. In S. M. Gass & A. Mackey (Eds.), The Routledge handbook of second language acquisition. New York: Routledge.

MacWhinney, B., Bates, E., & Kliegl, R. (1984). Cue validity and sentence interpretation in English, German, and Italian. *Journal of Verbal Learning and Verbal Behavior, 23*(2), 127-150.

MacWhinney, B., Pleh, C., & Bates, E. (1985). The development of sentence interpretation in Hungarian. *Cognitive Psychology*, *17*(2), 178-209.

Magnan, S. S. (1986). Assessing speaking proficiency in the undergraduate curriculum: Data from French. *Foreign Language Annals, 19*(5), 429-438.

Malovrh, P. A. (2006). L2 sentence processing of Spanish OVS word order and direct object pronouns: An analysis of contextual constraints. Paper presented at the Selected Proceedings of the 9th Hispanic Linguistics Symposium, 169-179.

Malovrh, P. A., & Lee, J. F. (2010). Connections between processing, production and placement. *Research in Second Language Processing and Parsing, 53*, 231.

Marinis, T., Roberts, L., Felser, C., & Clahsen, H. (2005). Gaps in second language sentence processing. *Studies in Second Language Acquisition, 27*(1), 53-78.

McCarthy, C. (2008). Morphological variability in the comprehension of agreement: An argument for representation over computation. *Second Language Research, 24*(4), 459-486.

McClelland, J. L., Rumelhart, D. E., & PDP Research Group. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition. Volume II.* Cambridge: MIT Press.

McDonald, J. L. (1987). Sentence interpretation in bilingual speakers of English and Dutch. *Applied Psycholinguistics, 8*(4), 379-413.

McDonald, J. (1989). The acquisition of cue-category mappings. In B. MacWhinney & E. Bates (Eds), *The crosslinguistic study of sentence processing* (pp. 375-396). New York: Cambridge University Press.

McDonald, J. L., & MacWhinney, B. (1989). Maximum likelihood models for sentence processing research. In B. MacWhinney & E. Bates (Eds), The crosslinguistic study of sentence processing (pp. 397-421). New York: Cambridge University Press.

Medler, D. A. (1998). *A brief history of connectionism. Neural Computing Surveys, 1,* 18-72.

Milton, J., & Meara, P. (1995). How periods abroad affect vocabulary growth in a foreign language. *ITL.Institut Voor Togepaste Linguistik*,17-34.

Montrul, S. (2010). How similar are adult second language learners and Spanish heritage speakers? Spanish clitics and word order. *Applied Psycholinguistics, 31*(01), 167-207.

Montrul, S., Foote, R., & Perpiñán, S. (2008). Gender agreement in adult second language learners and spanish heritage speakers: The effects of age and context of acquisition. *Language Learning, 58*(3), 503-553.

Montrul, S., Foote, R., Perpiñán, S., Thornhill, D., & Vidal, S. (2006). Full access and age effects in adult bilingualism: An investigation of spanish accusative clitics and word order. Paper presented at the Selected Proceedings of the 9th Hispanic Linguistics Symposium, 217-228.

Mora, J. C., & Valls- Ferrer, M. (2012). Oral fluency, accuracy, and complexity in formal instruction and study abroad learning contexts. *TESOL Quarterly, 46*(4), 610-641.

Moravcsik, E. A. (1988). Agreement and Markedness. In Michael Barlow and Charles A. Ferguson (Eds), *Agreement in Natural Language: Approaches, Theories, Descriptions* (pp. 89-106). Stanford: CSLI Publications.

Moreno, N. (2007). *The effects of type of task and type of feedback on L2 development in CALL*.(Unpublished doctoral dissertation). Georgetown University, Washington, DC.

Morett, L.M., & MacWhinney, B. (2013). Syntactic transfer in English-speaking Spanish learners. *Bilingualism: Language and Cognition, 16*(1), 132-151.

Mozgalina, A. (2015). *Applying an argument-based approach for validating language proficiency assessments in second language acquisition research: The elicited imitation test for Russian.*(Unpublished doctoral dissertation). Georgetown University, Washington, DC.

Nagata, N. (1993). Intelligent computer feedback for second language instruction. *The Modern Language Journal, 77*(3), 330-339.

Nagata, N., & Swisher, M. (1995). A study of consciousness- raising by computer: The effect of metalinguistic feedback on second language learning. *Foreign Language Annals, 28(*3), 337-347.

Nicholas, H., Lightbown, P. M., & Spada, N. (2001). Recasts as feedback to language learners. *Language Learning, 51*(4), 719-758.

Nicol, J., & Greth, D. (2003). Production of subject-verb agreement in Spanish as a second language. *Experimental Psychology, 50*(3), 196.

Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta- analysis. *Language Learning*, 50(3), 417-528.

Norris, J. M., & Ortega, L. (2006). *Synthesizing research on language learning and teaching.* Amsterdam: John Benjamins Publishing.

O'brien, I., Segalowitz, N., Freed, B., & Collentine, J. (2007). Phonological memory predicts second language oral fluency gains in adults. *Studies in Second Language Acquisition, 29(*4), 557-581.

Ortega, L., Iwashita, N., Norris, J., & Rabie, S. (2002). An investigation of elicited imitation tasks in crosslinguistic SLA research. Paper presented at the Second Language Research Forum, Toronto,

Ortega, L. (2013). SLA for the 21st century: Disciplinary progress, transdisciplinary relevance, and the bi/multilingual turn. *Language Learning, 63*(1), 1-24.

Ortega, L. (2009). *Understanding second language acquisition*, New York: Routledge.

Ortega, L., & Byrnes, H. (2008). *The longitudinal study of advanced L2 capacities*. New York: Routledge.

Pérez-Vidal, C. (2014). *Language acquisition in study abroad and formal instruction contexts.* Amsterdam: John Benjamins Publishing.

Pinker, S., & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition, 28*(1), 73-193.

Plunkett, K. (1998). Language acquisition and connectionism. *Language and Cognitive Processes, 13*(2-3), 97-104.

Potter, J. (2000). Post-cognitive psychology. *Theory & Psychology,10*(1), 31-37.

Quirk, R., & Greenbaum, S. (1973). *A Concise Grammar of Contemporary English*. New York: Harcourt Brace Jovanovich.

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin,* 114(3), 510.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*(3), 372.

Reyes, I., & Hernandez, A. E. (2006). Sentence interpretation strategies in emergent bilingual children and adults. *Bilingualism: Language and Cognition, 9*(1), 51-69.

Rosa, E. M. (1999). *A Cognitive Approach to Task-Based Research: Explicitness, Awareness and L2 Development* (Unpublished doctoral dissertation). Georgetown University, Washington, DC.

Rosa, E. M., & Leow, R. P. (2004). Awareness, different learning conditions, and second language development. *Applied Psycholinguistics, 25*(2), 269-292.

Rothweiler, M., Chilla, S., & Clahsen, H. (2012). Subject–verb agreement in specific language impairment: A study of monolingual and bilingual German-speaking children. *Bilingualism: Language and Cognition, 15*(1), 39-57.

Russell, J., & Spada, N. (2006). The effectiveness of corrective feedback for the acquisition of L2 grammar: A meta-analysis of the research. In J. M. Norris & L. Ortega (Eds.), *Synthesizing research on language learning and teaching* (pp. 133-164). Amsterdam: John Benjamin Publishing.

Sachs, R. R. (2011). *Individual differences and the effectiveness of visual feedback on reflexive binding in L2 Japanese.* (Unpublished doctoral dissertation). Georgetown University, Washington, DC.

Sagarra, N. (2000). *The longitudinal role of working memory on adult acquisition of L2 grammar.* (Doctoral dissertation). The University of Illinois at Urbana Champaign.

Sagarra, N. (2007). From CALL to face-to-face interaction: The effect of computer-delivered recasts and working memory on L2 development. In A. Mackey (Ed.), *Conversational Interaction in Second Language Acquisition* (pp. 229-248). Oxford: Oxford University Press.

Sagarra, N. (2014). 5 Absence of morphological transfer in beginners: evidence from eye tracking. In Z. Han and R. Rast (Eds.), *First Exposure to a Second Language: Learners' Initial Input Processing* (pp. 139-170). Oxford: Oxford University Press.

Sagarra, N., & Abbuhl, R. (2013). Optimizing the noticing of recasts via Computer-Delivered feedback: Evidence that oral input enhancement and working memory help second language learning. *The Modern Language Journal,* 97(1), 196-216.

Sagarra, N., & Herschensohn, J. (2010). The role of proficiency and working memory in gender and number agreement processing in L1 and L2 Spanish. *Lingua, 120*(8), 2022-2039.

Sagarra, N., & Herschensohn, J. (2011). Proficiency and animacy effects on L2 gender agreement processes during comprehension. *Language Learning, 61*(1), 80-116.

Sagarra, N., & Herschensohn, J. (2012). Processing of gender and number agreement in late Spanish bilinguals. *International Journal of Bilingualism, 17*(4), 657-664.

Sanchez, L., & Al-Kasey, T. (1999). L2 acquisition of Spanish direct objects. *Spanish Applied Linguistics, 3*, 1-32.

Sanz, C. (1997). Experimental tasks in SLA research: Amount of production, modality, memory, and production processes. *Contemporary Perspectives on the Acquisition of Spanish, 2*, 41-56.

Sanz, C. (2003). Computer delivered implicit vs. explicit feedback in processing instruction. In B. VanPatten (Ed.), *Processing Instruction: Theory, Research, and Commentary* (pp. 241-256). Mahwah, NJ: Erlbaum.

Sanz, C. (2014). Contributions of study abroad research to our understanding of SLA processes and outcomes: The SALA project, an appraisal. In C. Pérez-Vidal (Ed), *Second Language Acquisition in Study Abroad and Formal Instruction Contexts* (pp. 1-14). Philadelphia: John Benjamins.

Sanz, C., & Morgan-Short, K. (2004). Positive evidence versus explicit rule presentation and explicit negative feedback: A computer-assisted study. *Language Learning,* 54(1), 35-78.

Sanz, C., Park, H. I., & Lado, B. (2015). A functional approach to cross-linguistic influence in ab-initio L3 acquisition. *Bilingualism: Language and Cognition, 18*(2), 236-251.

Sasaki, Y. (1998). Processing and learning of Japanese double-object active and causative sentences: An error-feedback paradigm. *Journal of Psycholinguistic Research, 27*(4), 453-479.

Schneider, D., & Phillips, C. (2001). Grammatical search and reanalysis. *Journal of Memory and Language, 45*(2), 308-336.

Sebastián Gallés, N., Antonín, M., & MA, C. V.MF, & Cuetos Vega, F.(2000). Lexesp. Léxico Informatizado Del Espanol. CD-ROM. Edicions De La Universitat De Barcelona.

Segalowitz, N., Freed, B., Collentine, J., Lafford, B., Lazar, N., & Díaz-Campos, M. (2004). A comparison of Spanish second language acquisition in two different learning contexts: Study abroad and the domestic classroom. *Frontiers: The Interdisciplinary Journal of Study Abroad, 10*, 1-18.

Sekerina, I. A., Fernández, E. M., & Clahsen, H. (2008). *Developmental psycholinguistics: On-line methods in children's language processing.* Amsterdam: John Benjamins Publishing.

Serrano, R., Llanes, A., & Tragant, E. (2011). Analyzing the effect of context of second language learning: Domestic intensive and semi-intensive courses vs. study abroad in Europe. *System, 39*(2), 133-143.

Slobin, D. I. (1966). Grammatical transformations and sentence comprehension in childhood and adulthood. *Journal of Verbal Learning and Verbal Behavior, 5*(3), 219-227.

Slobin, D. I. (1973). Cognitive prerequisites for the development of grammar. *Studies of Child Language Development, 1*, 75-208.

Slobin, D. I. (1979). *Psycholinguistics*, Glenview, IL: Scott, Foresman.

Slobin, D. I. (1989). Foreword. In E. Bates & B. MacWhinney (Eds.), The crosslinguistic studies of sentence processing (pp. vii-ix). New York: Cambridge University Press.

Slobin, D. I., & Bever, T. G. (1982). Children use canonical sentence schemas: A crosslinguistic study of word order and inflections. *Cognition, 12*(3), 229-265.

Sokolov, J. L. (1989). The development of role assignment in Hebrew. In B. MacWhinney & E. Bates (Eds.), *The crosslinguistic study of sentence processing* (pp. 158-184). New York: Cambridge University Press.

Stafford, C. A., Bowden, H. W., & Sanz, C. (2012). Optimizing language instruction: Matters of explicitness, practice, and cue learning. *Language Learning, 62*(3), 741-768.

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. *Input in Second Language Acquisition, 15*, 165-179.

Tokowicz, N., & Warren, T. (2010). Beginning adult L2 learners' sensitivity to morphosyntactic violations: A self-paced reading study. *European Journal of Cognitive Psychology, 22*(7), 1092-1106.

Tremblay, A. (2011). Proficiency assessment standards in second language acquisition research. *Studies in Second Language Acquisition, 33*(3), 339-372.

Trueswell, J. C., Tanenhaus, M. K., & Kello, C. (1993). Verb-specific constraints in sentence processing: separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*(3), 528.

Valls-Ferrer, M., & Mora, J. (2014). L2 fluency development in formal instruction and study abroad: The role of initial fluency level and language contact. In C. Pérez-Vidal (Ed), *Second Language Acquisition in Study Abroad and Formal Instruction Contexts* (pp. 111–136). Philadelphia: John Benjamins.

Van Berkum, J. J., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(3), 443.

VanPatten, B. (1984). Learners' comprehension of clitic pronouns: More evidence for a word order strategy. *Hispanic Linguistics, 1*(1), 57-68.

VanPatten, B. (2004). Input processing in second language acquisition. In B. VanPatten (Ed.), *Processing Instruction: Theory, Research and Commentary* (pp. 5-31). Mahwah: Erlbaum

VanPatten, B. (2007). Input processing in adult SLA. In B. VanPatten & J. Williams (Eds.), *Theories in Second Language Acquisition* (pp. 115-135). Mahwah: Lawrence Erlbaum.

VanPatten, B., & Borst, S. (2012). The roles of explicit information and grammatical sensitivity in processing instruction: Nominative- Accusative case marking and word order in German L2. *Foreign Language Annals, 45*(1), 92-109.

VanPatten, B., & Cadierno, T. (1993). Explicit instruction and input processing. *Studies in Second Language Acquisition, 15*(2), 225-243.

VanPatten, B., Dvorak, T. R., & Lee, J. F. (1987). *Foreign language learning: A research perspective.* New York: Harper Collins College Division.

VanPatten, B., & Houston, T. (1998). Contextual effects in processing L2 input sentences. *Spanish Applied Linguistics, 2*(1), 53-70.

VanPatten, B., Keating, G. D., & Leeser, M. J. (2012). Missing verbal inflections as a representational problem: Evidence from self-paced reading. *Linguistic Approaches to Bilingualism, 2*(2), 109-140.

VanPatten, B., & Oikkenon, S. (1996). Explanation versus structured input in processing instruction. *Studies in Second Language Acquisition, 18*, 495-510.

VanPatten, B., & Sanz, C. (1995). From input to output: Processing instruction and communicative tasks. In F. Eckman, De. Highland, P. Lee, J. Milehan, & R. Rutkowski Weber (Eds.), *Second language acquisition theory and pedagogy* (pp. 169-185). Mahwah: Lawrence Erlbaum.

Verhagen, J., & Blom, E. (2014). Asymmetries in the acquisition of subject–verb agreement in dutch: Evidence from comprehension and production. *First Language, 34*(4), 315-335.

Wagers, M. W., Lau, E. F., & Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language, 61*(2), 206-237.

Weber-Fox, C. M., & Neville, H. J. (1996). Maturational constraints on functional specializations for language processing: ERP and behavioral evidence in bilingual speakers. *Journal of Cognitive Neuroscience 8*(3), 231-256.

Wechsler, S., & Zlatić, L. (2003). *The many faces of agreement.* Stanford: Center for the Study of Language and Information.

Yager, K. (1998). Learning Spanish in Mexico: The effect of informal contact and student attitudes on language gain. *Hispania, 81*(4), 898–913

Zandvoort, R. W. (1961). Varia syntactica. In *Language and Society* (pp. 193-203). Copenhagen, Denmark: Det.

Zwicky, A. M. (1985). Clitics and particles. *Language 6*1(2), 283-305.