6-2015

Clitic-doubled left dislocation in L2 Spanish: findings from learners' production data

Jose Sequeros-Valle
Depaul University, jseque3@uic.edu

Recommended Citation
http://via.library.depaul.edu/etd/203

This Thesis is brought to you for free and open access by the College of Liberal Arts and Social Sciences at Via Sapientiae. It has been accepted for inclusion in College of Liberal Arts & Social Sciences Theses and Dissertations by an authorized administrator of Via Sapientiae. For more information, please contact mbernal2@depaul.edu, MHESS8@depaul.edu.
CLITIC-DOUBLED LEFT DISLOCATION IN L2 SPANISH:
FINDINGS FROM LEARNERS’ PRODUCTION DATA

A Thesis
Presented in
Partial Fulfillment of the
Requirements for the Degree of
Master of Arts

June 2015

BY
Jose Sequeros-Valle
B.A., Universidad Complutense de Madrid, 2008

Department of Modern Languages
College of Liberal Arts and Social Sciences
DePaul University
Chicago, Illinois
ACKNOWLEDGMENTS

I would like to thank my thesis director Dr. Bradley Hoot and my thesis committee members Dr. Lourdes Torres and Dr. Nobuko Chikamatsu for all their work and time to help make this project possible. I am also grateful to my faculty advisor, Dr. Carolina Barrera-Tobón, for all her help and support during my Masters study, and Dr. Jennifer Cabrelli-Amaro for making her lab available for the experimental section of this project. I greatly appreciate the time I was able to borrow from the forty people who participated in the experiment for this thesis, as well as the support and assistance from my family and friends. Specially, I would like to thank both my Spain family and my Michigan family. Last but not least, I thank my girlfriend for her constant support since I began the adventure of graduate school.

AGRADECIMIENTOS

Me gustaría las gracias a mi director de tesis el Dr. Bradley Hoot y a los miembros de mi comité de tesis la Dr. Lourdes Torres y la Dr. Nobuko Chikamatsu por todo su tiempo y trabajo para hacer posible este proyecto. Estoy profundamente agradecido a mi tutora la Dr. Carolina Barrera-Tobón, por toda la ayuda y el apoyo que me ha ofrecido durante mis estudios de Maestría, y a la Dr. Jennifer Cabrelli-Amaro por permitirme usar su laboratorio para el experimento de esta tesis. Agradezco enormemente el tiempo que le pude tomar prestado a las cuarenta personas que participaron en el experimento de esta tesis, así como el apoyo y la ayuda de mis familiares y amigos. Me gustaría darle las gracias de manera especial tanto a mi familia de España como a mi familia de Michigan. Por último y no por ello menos importante, agradezco a mi novia el apoyo constante que me ha ofrecido desde que decidió comenzar la aventura de mis estudios de postgrado.
ABSTRACT

Although the Interface Hypothesis (IH) posits that structures involving an interface between syntax and discourse may be vulnerable to permanent residual optionality (Sorace, 2011, 2012), several studies have found that near-native learners are able to acquire discourse knowledge to a native-like level in relation to Spanish CLLD (Slabakova et al., 2012 and more) and similar structures in other languages. While the research on L2 acquisition of CLLD has used several methodologies such as acceptability judgment tasks, sentence selection tasks, and sentence completion tasks, the discussion is still missing data from production tasks. Three groups of learners (near-native, advanced, and intermediate) responded to questions that required either CLLD or contrastive answers (focus fronting and/or clefts) on a production task. The data collected for the present study shows a tendency toward convergence between Spanish native speakers and very advanced learners of Spanish as a second language. The data also found the complete convergence of one of the eighteen learners. While the data from the most advanced learners as a group could be seen as evidence for permanent optionality, the individual data from the mentioned participant challenges the claims made by the supporters of the IH.

Keywords: second language acquisition, syntax-discourse interface, Interface Hypothesis, Syntax-Before-Discourse Hypothesis, production, Spanish, clitic, left dislocation
TABLE OF CONTENTS

LIST OF TABLES 6
LIST OF FIGURES 7
LIST OF ABBREVIATIONS 8

PART I - INTRODUCTION AND BACKGROUND 9

1. Introduction 9
2. The Interface Hypothesis 12
   2.1. Syntax and Interfaces 13
       2.1.1. Narrow Syntax vs. Interfaces 13
       2.1.2. Internal vs. External Interfaces 15
   2.2. What causes optionality? 17
3. Criticisms to the Interface Hypothesis 19
4. Information Structure 22
   4.1. A Topic Construction: Clitic-Doubled Left Dislocation 22
   4.2. Contrastive Focus Constructions 25
       4.2.1. Focus Fronting 25
       4.2.2. Clefts 26
       4.2.3. Focal Adverbs 27
5. The L2 Acquisition of Spanish CLLD 28
   5.1. CLLD and FF in L2 Spanish (Slabakova et al., 2012) 31

PART II - WHAT THIS THESIS INTENDS TO ADD TO THE DISCUSSION 34

6. Data From Production Tasks 34
7. Research Questions 35
8. Predictions 36
   8.1. Hypothesis A: Learners’ convergent behavior 37
   8.2. Hypothesis B: Learners’ divergent behavior 38

PART III - METHODOLOGY AND SCOPE 40

9. Participants 40
10. Methodology 42
11. Considerations in the Design 44
PART IV - RESULTS

12. Results by Level of Proficiency
   12.1. Dislocation and Clitic-Doubling
   12.2. Codes at the Data Set
   12.3. Topic and Focus Conditions

13. Results by Tendency
   13.1. Topic and Focus Conditions by Tendency
   13.2. Equivalence and Set-Subset Topic Conditions

PART V - DISCUSSION AND CONCLUSIONS

14. Revisiting the first research question
   14.1. A Note on the Contrastive Focus Constructions
15. Revisiting the second research question
16. Revisiting the Thesis’ Predictions
17. Implications to the Interface Hypothesis
18. Limitations
19. Conclusions

APPENDICES

Appendix A: Stimuli
Appendix B: DELE Test
Appendix C: Clitic Test
Appendix D: Data Set

REFERENCES
LIST OF TABLES

Table 1. Score ranges in the DELE and clitic tests used by this thesis and by Slabakova et al. (2012) to create groups of proficiency. 41

Table 2. Samples responses by one native and three non-native participants to the same pair of questions, one of them topic condition and the other one focus condition. 50

Table 3. Sample responses for a concrete focus condition stimulus, by type/number of participant. 55
LIST OF FIGURES

Figure 1. Dislocations and clitic-doubling performed under topic condition, by level of proficiency. 47

Figure 2. Type of construction performed under topic and focus conditions, by level of proficiency. 49

Figure 3. Percentage of participants from each level of proficiency (near-native, advanced, intermediate) who performed with a certain tendency (native-like speakers, overextenders of the clitic-doubling, or underextenders of the clitic-doubling). 52

Figure 4. Type of construction performed in topic and focus conditions, by tendency. 54

Figure 5. Type of construction performed under topic and focus conditions, by each one of the three native-like L2ers, compared to the average of native speakers. 55

Figure 6. Type of structure performed under the equivalence and the set-subset topic conditions, by tendency. 56
LIST OF ABBREVIATIONS

# Ungrammatical sentence or unfelicitous discourse.

ACC[+hum] Indicator of a human element with accusative case. Accusative means that the element has a direct object function.

Antec Antecedent, the element to which the CLLD is referring.

CAN+ Canonical sentence plus a left dislocation.

CLLD Clitic-doubling Left Dislocation.

Cl.acc Accusative clitic.

CL.dat Dative clitic.

DAT Dative case indicator. Dative means that the element has an indirect object function.

FF Focus fronting.

GASLA Generative Approaches to Second Language Acquisition conference.

L2 Second language.

L2ers Second language learners.

NEG Negation.

PAST Past tense particle.

put.fut.1st.plu Verb “to put” in the first person plural conjugation in the future tense.

see.1st Verb “to see” in the first person singular conjugation in the present tense.

see.inf Verb “to see” in the infinitive form.

send.1st.fut Verb “to send” in the first person singular conjugation in the future tense.

SP Subject Pronoun.
PART I
INTRODUCTION AND BACKGROUND

1. Introduction

The Interface Hypothesis (IH) (see Sorace 2011 and Sorace 2012 for an updated version) claims that constructions that involve an external interface between syntax and other cognitive domains—including the syntax-discourse interface—are especially vulnerable to residual optionality or instability in bilingual grammars. However, the evidence is inconsistent (see Montrul 2011; White 2011a for an overview), and more research is necessary to elucidate which interface properties are vulnerable and in what way. The present work contributes to this debate by testing the IH using experimental evidence of clitic-doubled left dislocation (CLLD), a syntax-discourse interface phenomenon, in adult learners of Spanish as a foreign language.

According to Lopez (2009), CLLD in Spanish and other Romance languages involves moving the direct or indirect object to the left periphery of the sentence, and adding a resumptive clitic pronoun that does not add any meaning to the sentence. CLLD in Spanish involves the interface of syntax and discourse context. In other words, while a canonical sentence can be used in any discursive context, a sentence containing a CLLD can only be used if the CLLD expands on a discursive topic that has been introduced in previous statements. Several empirical studies have researched the acquisition of CLLD by adult learners of Spanish (Valenzuela, 2005, 2006; Slabakova et al., 2012; Leal Mendez et al., 2015) and similar structures in other languages, such as Bulgarian (Ivanov, 2009, 2012) and French (Donaldson, 2011, 2012).
Some studies in language acquisition have compared CLLD to focus fronting (FF) (Slabakova et al., 2012; Leal Mendez et al., 2015; among others), as a syntactic structure that also involves an interface between syntax and discourse; however, they have different discursive meanings. While CLLD expands on the topic, FF contradicts it and offers an alternative focus to the discourse. Other constructions that have the same contrastive function are clefts (Rodríguez-Ramalle, 2005). Slabakova et al. (2012) and Leal Mendez (2015) showed evidence of stable acquisition in developing grammars at the interface between the syntax and the discursive restrictions of those two constructions, CLLD and FF, unlike what is predicted by the IH. Other studies showed similar results challenging the IH with different constructions in different languages (Ivanov, 2009, 2012; Donaldson, 2011, 2012). The closest study to the focus and design of this thesis is the study by Slabakova et al. (2012) due to its specific focus on L2 Spanish, Spanish CLLD and its discursive restrictions, and the IH. The experimental section of this thesis is in fact based on Slabakova et al.'s with a different methodology.

While the research on L2 acquisition of CLLD has used several methodologies such as acceptability judgment tasks (Slabakova et al., 2012; Leal Mendez et al., 2015, among several others), sentence selection tasks (Valenzuela, 2005, 2006), and sentence completion tasks (Valenzuela, 2005, 2006), the discussion is still missing data from production tasks. Data from production tasks have been used successfully in the study of several other syntactic structures that involve interfaces (Tsimpli & Sorace, 2006; Belletti et al., 2007; Montrul & Rodríguez-Louro, 2006), even in left dislocations similar to Spanish CLLD (Donaldson, 2011, 2012), but never with Spanish CLLD.
In the present study, three groups of L2 learners (near-native speakers, advanced learners, and intermediate learners), plus a native speaker control group, responded to twenty-four questions that required either CLLD or contrastive answers (FF, clefts, or focal adverbs). The main goal of the study is to discover if adult learners of Spanish as a foreign language are able to use complex syntactic structures such as CLLD, FF, etc., and if they are able to do so in the correct discursive context. The design of this experiment was made as close as possible to Slabakova et al.’s (2012) so that both data sets are comparable.

Although the Interface Hypothesis posits that structures involving an interface between syntax and discourse may be vulnerable to permanent residual optionality (Sorace, 2011, 2012), several studies have found that near-native learners are able to acquire discourse knowledge to a native-like level in relation to Spanish CLLD (Valenzuela, 2005, 2006; Slabakova et al., 2012; Leal Mendez et al., 2015) and similar structures in other languages (Ivanov, 2009, 2012; Donaldson, 2011, 2012).

The data collected for the present study shows a tendency toward convergence between the grammars of Spanish native speakers and non-native speakers. Even though one could argue that the most advanced group's results seem to show instability when comparing them to those from the native speakers, a closer look at the individual results show that at least one participant performed completely native like. This individual result challenges the claims posited by the IH.

The rest of Part I reviews relevant literature on the IH, criticisms to the theory, information structure, and L2 acquisition of Spanish CLLD. Part II justifies the use of a production task and states the research questions and some predictions. Part III presents the individuals who participated in the empirical task, as well as explaining the task itself and some
considerations in its design. Part IV presents the results, which are discussed in Part V. This last part also includes a review of the research questions and predictions in light of the results, implications of the results to the field of generative second language acquisition, limitations in the design of the study, and conclusions.

2. The Interface Hypothesis

One of the main debates in the field of generative acquisition today is whether or not all syntactic structures can be acquired without residual instability. Sorace (2011), as the main advocate for the Interface Hypothesis, argues that structures in which syntax interfaces with non-linguistic cognitive domains, such as discourse or pragmatics, are impossible to acquire without residual optionality for L2 learners.

Sorace (2011) offers some examples of optionality of the syntax-discourse interface in adult learners of a foreign or second language, which are the type of interface and the type of language students that will be analyzed in this thesis. For example, the author explains that near-native learners of Italian present a higher degree of acceptance of overt subject pronouns with an antecedent than do Italian monolingual speakers.

(1) Perché [Giovanna]_{Antec} non è venuta?  
Perché [_{SP} (lei)] non ha trovato un taxi.

Why didn’t [Giovanna]_{Antec} come?  
Because [_{SP} she] couldn’t find a taxi.

Native speakers of Romance languages can omit subject pronouns, but subject expression or omission is contextually constrained. Native speakers normally add overt subjects to sentences only when there is a change of referent, when the antecedent happens far earlier in the discourse, or when it is not clear to whom the sentence is referring. According to Sorace (2011), near-natives accept overt pronouns in a higher degree than native speakers due to the complexity
of processing structures with this interface between syntax and discourse. The acceptance of overt vs. null subject pronouns is only one of the different syntactic phenomena that has been studied in the interface between syntax and other linguistic and nonlinguistic domains, which has been found to be unstable for bilinguals.

After this brief introduction, and before explaining the main empirical task of this thesis, the rest of Part I reviews the relevant literature as follows: first, the Interface Hypothesis (IH) will be discussed in Section 2. Section 3 presents some of the main criticisms that have been raised to some of the claims of the IH. The syntactic theory of Spanish CLLD and other relevant constructions are defined in Section 4. Section 5 reviews the latest studies in L2 acquisition of Spanish CLLD in relation to the IH.

2.1 Syntax and Interfaces

2.1.1. Narrow Syntax vs. Interfaces

The IH as originally proposed stated that bilingual L1 speakers, native speakers under L1 attrition due to extended exposure to a second language, and L2 learners are subject to permanent optionality at the interface between syntax and other cognitive domains. The phenomena was presented as a dichotomy between interfaces and narrow syntax (Sorace, 2006). According to this first version of the IH, syntactic structures at interfaces are subject to permanent instability while pure syntax is perfectly acquirable.

According to Sorace (2011), the term “Interface Hypothesis” appeared for the first time in Sorace and Filiaci (2006). The authors asked their L1 English speakers to read sentences with anaphora in L2 Italian. In the sentences provided, some included overt pronouns and some included null pronouns. The participants were then asked to decide which of the given pictures
represented the action explained in the sentence, in which the agent was either the subject or the complement of the main clause. Near-native speakers interpreted null-pronoun anaphora similarly to native speakers, but differed in interpreting overt pronoun anaphora. Example (2) shows how a native speaker would understand the sentence with a null pronoun (shown in 2a) and with an overt pronoun (shown in 2b). Results in the study showed that this was not the case for near-native speakers.

(2) a. La vecchietta, saluta la ragazza, quando \( \text{pro}_{ij} \) attraversa la strada.

b. La vecchietta, saluta la ragazza, quando lei attraversa la strada.

*The old woman greets the girl when \( \varnothing /s\)he crosses the road.*

According to Sorace and Filiaci (2006), null-subjects in anaphora are acquired without indeterminacy because they follow the basic pattern of reference and because they are part of narrow syntax. However, in languages that allow null subjects such as Italian, overt pronouns are used to indicate a topic shift from the basic pattern of reference. \([+\text{Topic Shift}]\) is considered a discursive feature which interfaces with syntax. According to the authors, overt pronoun anaphora are subject to permanent optionality in L2 learners of Italian due to this non-syntactic feature.

Similar results were found by Belletti et al. (2007). The authors studied the interpretation and production of null and overt preverbal subjects and postverbal subjects by L2 learners of Italian by using both acceptability and production tasks. Like Sorace and Filiaci (2006), Belletti et al. (2007) also found an over-extension in the use of overt pronouns. Additionally, the authors found an overextension in the use of preverbal subjects by near-natives in situations in which native speakers tended to use a postverbal subject to express new information. For example, to
answer the question (3a), near-natives tended to use an answer like (3b) while native speakers normally use a sentence like (3c):

(3)  
   a. Chi parlerà? Who will-talk  
   b. #Gianni parlerà. Gianni will-talk  
   c. Parlerà Gianni. Will-talk Gianni (Gianni will talk.)  

The overuse of preverbal subjects by near-native learners of L2 Italian was understood by Belletti et al. (2007) as a result of the instability in the discursive use of the Italian focus, the most common manifestation of which is the movement of the subject to a postverbal position. However, the same learners did not over-extend the scope of preverbal subjects which are, according to the authors, related to narrow syntax and perfectly acquirable.

The differentiation between narrow syntax vs. interfaces was eventually abandoned by Sorace and colleagues (Sorace, 2012). Sorace (2011) recognizes that it was difficult to identify syntactic structures subject only to syntax itself and not conditioned by any other linguistic or non-linguistic domains. This realization made the IH evolve into a new distinction between internal and external interfaces.

2.1.2. Internal vs. External Interfaces

Parallel to the differentiation between narrow syntax vs. interfaces, a second version of the IH was being developed (Tsimpli & Sorace, 2006; Sorace & Serratrice, 2009). White (2009) explains the way Sorace and colleagues understand interface phenomena up to today. Internal interfaces, which are the interfaces between different domains within language, would be unproblematic and perfectly acquirable for L2 learners. Some examples of internal interfaces would be the phonology-morphology or syntax-semantics interfaces. External interfaces, which
are the interfaces between a linguistic and a non-linguistic domain, would be problematic and subject to permanent optionality for L2 learners. Some instances of external interfaces would be the syntax-pragmatics or syntax-semantics-pragmatics interfaces. Data supporting these claims can be found in Tsimpli and Sorace (2006) and Sorace and Serratrice (2009).

Tsimpli and Sorace (2006) found divergence between L2 Greek speakers whose L1 is Russian and native speakers of Greek at the syntax-discourse interface (an external interface), while they found convergence at the syntax-semantics interface (an internal interface). The external interface was studied through the use of overt vs. null subject pronouns. The authors found an over-extension of the overt pronouns. That was interpreted as a discursive difficulty given that in Greek the appearance of a subject pronoun depends on the discourse (similarly to example (1) in Italian). The internal interface was studied by a comparison of the FF and CLLD. According to the authors, the fact that even the intermediate learners were able to differentiate between FF and CLLD by the use or absence of clitic-doubling is understood as successful acquisition at the syntax-semantics interface. See example (4):

(4) a. TON PETRO sinandise i adhelfi mu. (FF)
the-acc Petro met-3s the-nom sister my

_It was Petro that my sister met._

b. Ton Petro ton-sinadise i adhelfi mu. (CLLD)
the-acc Petro clitic-met-3s the-nom sister my

_Petro, my sister met him._

---

1 In Tsimpli & Sorace (2006), the difference between FF and CLLD was understood as semantic, instead of discursive. Sentences were used out of context, and their meanings were differentiated only by the use of absence of a clitic-doubling.
Sorace and Serratrice (2006) reviewed data from two studies to support the difference between internal and external interfaces for L2 learners. The first study shows that learners differ from native speakers in the use of overt and null subject pronouns in Italian. As mentioned in section 2.1.1., native speakers interpret the discursive information of a sentence depending on whether or not the sentence has a subject pronoun (see example (2)). However, L2 learners appear to be unable to acquire this information without instability. The second study compares monolingual English-speaking children and bilingual Italian-Spanish children in their Italian language. While the former acquire the generic vs. specific distinction (syntax-semantics interface, internal) and anaphora (syntax-discourse interface, external) without major problems, the latter acquire the internal interface but present permanent optionality in the external interface.²

The assumption that external interfaces are subject to permanent instability will be important for this study. Spanish CLLD requires discursive knowledge in order to be felicitously used, so it can be considered to be at the syntax-discourse interface. As part of the syntax-pragmatics interface, one of the external interfaces, the IH would predict that Spanish CLLD will not be acquirable without optionality and will create permanent issues for L2 learners.

2.2 What causes optionality?

Similarly to the debate about narrow syntax, internal interfaces, and external interfaces, different positions about the cause of optionality have developed in parallel based on more than ten years of research (see Sorace, 2011 for a review). While some studies state that optionality is

² See Sorace and Serratrice (2009, p. 204) for examples of the different structures studied and the interfaces that they imply in both English and Italian. The examples will not be displaced in this paper due to lack of space.
present at the level of “representation” and part of the grammatical system, others explained that the problems are caused by “processing” factors, the real-time use of language.

Problems at the representation level are understood as divergences in learners’ vs. native speakers’ mental grammars. Specifically, Tsimpli and Sorace (2006) and Belletti et al. (2007) believe that, while syntactic parameters can be reset, discursive formal features are not consistently reset in the L2, and L1 transfer takes place even at the end-state of L2 acquisition. For example, a learner of Spanish whose first language is English may be able to reset the parameter that allows Spanish sentences to have a flexible word order, and move the direct object of a sentence to the left periphery. However, and borrowing Lopez’s (2009) terminology, that learner would not be able to perfect discursive knowledge about the topic condition. Therefore, he or she would never be able to use a Spanish CLLD in the right context consistently.

The second way by which the causes of permanent optionality are explained take processing factors into the account. In Chomskian terms, we could say that, while the first explanation claims that errors happen at the competence level, this second opinion claims that non-target-like linguistic behaviors at interfaces are just mistakes at the performance level. According to Sorace and Serratrice (2009), individuals who speak more than one language are less efficient than monolinguals on processing in real time all the different moving parts of a syntactic structure at an external interface. Coming back to the previous example, the learner will be able to acquire the syntactic parameter that allows flexible word order and the discursive knowledge about the topic condition. The problem, according to this view, would be that he or
she would not be able to always perform like a native speaker due to the real time use of language in conversation.

Sorace (2006) also argues that learners display non-target performance due to processing factors. The author is the first one targeting the need for on-line methodologies in the research of interfaces and the inclusion of psycholinguistic data and methods. If problems at the interface for L2 learners are due to processing limitations, research should use methodologies that can study processing itself rather than the result of it. This type of “interdisciplinary work” (Sorace, 2011, p. 25) was later supported by Sorace (2011, 2012) and colleagues (Sorace & Serratrice, 2009), and many other researchers in the field of generative acquisition (Montrul, 2011; Montrul & Rodriguez-Louro, 2006; Rothman & Slabakova, 2011; Slabakova & Ivanov, 2012; Donaldson, 2011, 2012; Slabakova et al., 2012; Leal Mendez et al., 2015).

3. Criticisms to the Interface Hypothesis

The theoretical basis of the IH has been strongly criticized based on newer data. Slabakova (2009, 2014), Rothman and Slabakova (2011), and Lardiere (2011) state that the origin of problems at very high levels of proficiency for L2 learners is functional morphology, instead of the interfaces. Montrul (2011), White (2009, 2011a), and Slabakova (2009, 2014) believe that external interfaces are not necessarily more difficult than the external ones, and criticize this division. Pires and Rothman (2011) explain that, given the contradictory data that has been found on interface phenomena, the problem may not be in the interfaces per se, and suggest exploring beyond the interfaces. White (2011b) proposes that the IH could and should be tested beyond the methods and the individuals for whom it was originally proposed, such as testing the IH on heritage speakers (Montrul, 2011; Montrul and Polinsky, 2006; Leal Mendez et
However, for the purpose of this paper, the focus will be on three main criticisms: the predictive power of the IH, the linguistic theory on which the IH is based, and the Syntax-Before-Discourse Hypothesis.

The IH has been criticized for the use of the word “may” in its formulation. For example, Sorace (2011) summarizes the IH in a few words like this:

“narrow syntactic properties are completely acquirable in a second language, even though they may exhibit significant developmental delays, whereas interface properties involving syntax and another cognitives domains may not be fully acquirable.”

The use of the word “may” has been criticized by Ivanov (2009, 2012). The author believes that the IH is unfalsifiable due to the way it is currently proposed. The IH could virtually accommodate any result, because as much as interfaces “may” cause problems for L2 learners, they “may not,” and this result would still not challenge the IH. The author states that any scientific hypothesis needs to offer clear predictions that can be tested and, based on the results of the tests, the hypothesis can be supported or challenged. For the purpose of this paper, the assumption will be that the IH predicts that external interfaces ARE NOT fully acquirable, rather than external interfaces MAY NOT BE fully acquirable. This way, if this study finds convergent data, this evidence will be used as support to the IH. However, if this study finds divergent data, this evidence will be a challenge against the IH.

The IH has also been criticized due to the lack of connection to linguistic theory. For example, Sorace has been criticized for not taking any position about the access to Universal Grammar by adult learners (Pires & Rothman, 2011, and others), which is an important debate within the field of generative acquisition. However, due to the topic of this thesis, the focus will be on the lack of basis in a theory of information structure. Sorace (2011) argues that topic and
focus constructions\(^3\) are different in nature. The author believes that, while the former involves external conditions, the latter offers new information in relation to a given topic. Sorace also states that the topic condition is at the syntax-pragmatics interface while the focus condition is assumed to be part of the core syntax (Tsimpli & Sorace, 2006). However, Slabakova (2011) reviews the work of several linguistic theorists including Lopez (2009), arguing that both topic and focus conditions are actually two sides of the same discursive phenomena. Both a topic structure such a CLLD and a focus structure such a FF are restricted by the discourse in order to be felicitous. Making this point is important for this study since all the structures that the participants will potentially use in the empirical task of this thesis (CLLD, FF, and clefts) are at the syntax-discourse interface level.

The third criticism to the IH is about the prediction about L2 endstate. Instead of assuming that syntactic structures at the syntax-discourse interface are subject to permanent optionality, the Syntax-Before-Discourse Hypothesis (Montrul & Rodriguez-Louro, 2006; Rothman, 2009) agrees that the acquisition of structures at the syntax-discourse interface is the last or one of the latest stages in L2 acquisition. However, the hypothesis believes that discursive knowledge can be acquired without instability once syntax has been perfected. Montrul and Rodriguez-Louro (2006) asked their intermediate, advanced, near-native participants to narrate a story. These natural production data were analyzed for the three different groups. Intermediate learners showed several cases of syntactic ungrammaticality and discursive non-felicitousness, advanced learners showed almost target-like syntactic behavior but instability at the discursive level, and near-native learners showed target-like behavior both at the syntactic and discursive

\(^3\) Topic and focus conditions are further explained in section 4.2.
levels. Rothman (2009) found similar results by using interpretation, felicitousness, and translation tasks; discursive knowledge can be acquired without optionality once syntax has been perfected. The Syntax-Before-Discourse Hypothesis is important for this paper because the data set shows that L2 learners can behave in a target-like manner at the syntax-discourse interface. These and other data (offered in section 5) will be important when formulating predictions for the results of this thesis.

4. Information Structure

First defined by Halliday (1967), information structure can be described as the group of mechanisms that speakers use to organize units of information in a sentence or a discourse. Topic and focus are two of these mechanisms of the information structure. For the purpose of this study, the topic construction CLLD is explained in section 4.1. Additionally, some focus constructions are defined in section 4.2. CLLD and one focus construction, FF, are the main focus of Slabakova et al. (2012). As mentioned in section 1, the experiment for this thesis is based on Slabakova et al.’s study but using a different methodology, so the relevant syntactic theory must be properly explained.

The way information is formulated according to the discursive context and its acquisition has been studied in languages such as French (Donaldson, 2011, 2012), Bulgarian (Ivanov, 2009, 2012), and Spanish (Valenzuela, 2005, 2006; Slabakova et al., 2012; Leal Mendez et al., 2015). To master this realm, the speaker needs to take into account syntactic elements, such as word order or clitic-doubling, but also discursive elements, such as the relationships between antecedents and topics and several extra-linguistic aspects.

4.1 A Topic Construction: Clitic-Doubled Left Dislocation
One way to organize the information in a sentence is to use CLLD, which involves the extralinguistic elements mentioned. This structure is also highly complex syntactically, which will allow this study to analyze the stages in the acquisition of several elements.

Lopez (2009) analyzes CLLD and compares it with other ways to organize a discourse. He says that, in Romance languages, all arguments can be dislocated. He offers several examples from Spanish and Catalan to demonstrate that both the direct and indirect objects can be dislocated to the left periphery of the sentence. The dislocated element must be duplicated by a clitic pronoun.

(5) Hace tiempo que no veo a María. (Canonical sentence)

Does time that NEG see.1st ACC[+hum] Maria

I haven’t seen Maria in a while

(6) A María hace tiempo que no la veo. (CLLD)

ACC[+hum] Maria does time that NEG Cl.acc see.1st

Maria, I haven’t seen her in a while

Sentence (5) represents the basic form of the sentence (6) offered by Lopez (2009). This example shows that the element “a María” can be dislocated to the left periphery of the sentence, but it requires the addition of a clitic.

Lopez (2009) explains that the different ways to organize the information in a sentence are only felicitous in certain discursive contexts, and not in others. He explains that CLLD is felicitous only in the case of coordination or subordination to a given topic. This means that the element that is dislocated to the left periphery of the sentence is either the topic in the discourse or a reference to a sub-category of the topic. Lopez offers several examples in Catalan:
El Pere va veure [la Maria] Antec.

The Pere PAST see.inf the Maria

Pere saw Maria.

[CLLD] La Maria li agrada molt, a ell.

The Maria Cl.dat likes much, DAT him

Maria, he likes her a lot.

El Pere va veure [la Maria] Antec.

The Pere PAST see.inf the Maria

Pere saw Maria.

# Després [CLLD] la Susanna també la va veure, a ella.

Later the Susanna also Cl.acc PAST see.inf ACC[+hum] her

# Later, Susana saw her too.

¿Qué farem amb [els mobles] Antec?

what are we doing with [the furniture] Antec?

- [CLLD] Les taules de fòrmica], les posarem a la cuina.

the tables of formica Cl.acc put.fut.1st.plu in the kitchen

The tables of formica, we will put them in the kitchen.

Discourse (7) is felicitous because the dislocated element CLLD refers to an antecedent. Discourse (8) is not felicitous because the dislocated element CLLD does not refer to the antecedent that the conversation is discussing. The discourse (9) is felicitous because the dislocated element CLLD refers to a sub-category of the antecedent. In this last example, the

---

4 It is common in Catalan to add a definite article before a person’s name.
element “les taules de fòrmica” is a sub-category within “els mobles.” It can be dislocated to the left periphery since it expands on the given topic. Simply put, “les taules de fòrmica” explains what is going to happen to some of the elements of a larger category, “els mobles,” that are the antecedent and the topic of the discourse.

4.2 Contrastive Focus Constructions

Slabakova et al. (2012) compare Spanish CLLD, a topicalization, to Focus Fronting (FF), a focalization. These two structures can be easily compared because both of them involve a dislocation to the left periphery, and CLLD uses a clitic-doubling while FF does not. However, they are appropriate in different discursive contexts. Since this thesis will collect data by using a production task, there is not a way to make sure what type of sentence the participants will use to answer. Due to the way the experiment is designed (explained in section 11), topicalizations are controlled, such that CLLD is the only possible answer. However, the focus cases are open to FF (like in Slabakova et al. 2012), as well as to clefts and focal adverbs, as shown in example (10). These three contrastive focus constructions are explained in the following sub-sections.

(10)a. ¿Viste a Pedro en el parque? Did you see Pedro at the park?

b. - ¡No! ¡A Luis vi! No! Luis I saw! (FF)

c. - ¡No! ¡A Luis es al que vi! No! Luis is the one I saw! (Cleft)

d. - ¡No! ¡A Luis solamente vi! No! Luis I only saw! (Focal Adv)

4.2.1. Focus Fronting

Rodriguez-Ramalle (2005) explains that FF is another way to organize information in a sentence, also based on the movement of an element from its natural position to the left periphery. However, FF does not allow clitic-doubling, and the dislocated element adds new
information that contradicts the question. It requires the postposition of the subject after the verb, and it is normally produced with a strong prosodic emphasis. Slabakova et al. (2012) offer the following example, in which the FF “la carne” contradicts the antecedent “la sopa:”

(11)a. María felicitó a Juan por [la sopa]Antec que había hecho.

Maria complimented Juan on [the soup]Antec he cooked.

b. [FF LA CARNE] preparó Juan, no la sopa.

THE MEAT prepared Juan, not the soup

4.2.2. Clefts

Rodriguez-Ramalle (2005) compares FF and clefts due to their similarities in discursive use, because clefts can also be used to contradict the antecedent. However, the syntactic structure is very different. In Spanish, clefts are formed by a preverbal subject, a copula with the verb ser “to be” and a relative clause starting with a relative pronoun such as quien ”who,” el que ”the one (person) that,” lo que “the one (thing) that,” and many more. Sentence (12) is an example of cleft in Spanish.

(12) Aránzazu fue la que ganó ayer el concurso (y no Inmaculada).

Aranzazu was her that won yesterday the competition (and not Inmaculada)

Aranzazu was the one who won the competition yesterday (and not Inmaculada).

The word order is extremely flexible in these types of sentences. However, Rodriguez-Ramalle (2005) explains that there are some discursive differences between them. For example, while (13a) is just an informative sentence, (13b) and (13c) imply a contrast with an
antecedent. In fact, the word order of (13c) is the one that may be used by the participants of this study, as explained in section 11.

(13)a. La que ganó el concurso fue Aránzazu.

Her that won the competition was Aranzazu

*The one who won the competition was Aranzazu.*

b. Fue Aránzazu la que ganó el concurso.

Was Aranzazu her that won the competition

*It was Aranzazu the one who won the competition.*

c. Aránzazu fue la que ganó el concurso.

Aranzazu was her that won the competition

*Aranzazu was the one who won the competition.*

4.2.3. Focal Adverbs

Rodriguez-Ramalle (2005) identifies four types of adverbs that can create a discursive contrast: Exclusive (e.g. sólo (only)), inclusive (e.g. también (also)), identificative (e.g. exactamente (exactly)), and specifics (e.g. especialmente (especially)). Focal adverbs can be located in different parts of a sentence, and discourse is needed in order to know which element the adverb is focalizing. Discourse (14) shows how a focal adverb can be used to create contrastive focus:

(14)a. ¿Por qué visitaste a Pablo en el hospital?

Why visit.past.1s ACC[+hum] Pablo in the hospital

*Why did you visit Pablo at the hospital?*

b. A Mario solamente visité.
As explained in following sections, the data used in this paper show the use of the focus constructions FF and clefts only. However, none of the participants, native or not, used focal adverbs. This section included a description of focal adverbs because they are grammatical and felicitously possible answers as shown in example (10) at the beginning of section 4.2. Therefore, focal adverbs are taken into account in this thesis at a theoretical level.

5. The L2 acquisition of Spanish CLLD

The acquisition of information structure in developing grammars has been widely debated in the field of generative acquisition, especially since the appearance of the IH (Sorace & Filiaci, 2006). One syntactic structure that is studied in many papers is CLLD and its acquisition. This section will summarize the findings on the acquisition of this structure in Spanish (Valenzuela, 2005, 2006; Slabakova et al., 2012; Leal Mendez et al., 2015), French (Donaldson, 2011, 2012), and Bulgarian (Ivanov, 2009, 2012) to provide an overview of the debate around the acquisition of CLLD and the IH.

Valenzuela (2005, 2006) found divergent behavior between near-native learners and native speakers looking at CLLD and CLD (contrastive left dislocation) in L2 Spanish. The author explains that, while CLLD is used to move elements that are specific, CLD is used with non-specific or generic elements. The difference is shown in examples (15a) and (15b).

(15)a. El libro, lo leí. (CLLD)

the book cl. I-read

The book, I read it.
b. Un libro, lei.

\[(CLD)\]

\[a \text{ book I-read}\]

\[A \text{ book, I read.}\]

According to Valenzuela (2005, 2006), both structures depend on the discourse in order to be felicitous. The results from an oral grammaticality judgment task (to check if participants could identify the differences between the two structures as specific vs. non-specific), an oral sentence selection task, and a sentence completion task, showed an overgeneralization of CLLD over CLD. This divergence between native and non-native speakers was interpreted by Valenzuela as evidence for native-like acquisition at the syntactic level by the near-natives, but incomplete acquisition of the discourse features. The author concluded that this overgeneralization of CLLD was evidence in support of the IH.

Ivanov (2009, 2012), however, found convergence between near-native learners and native speakers in the use of Bulgarian object clitics. The author states that even though syntactic structures that coordinate syntax and other cognitive domains create delays in L2 acquisition, those structures and their discourse knowledge are eventually acquired in the near-native stages of acquisition. Even though the structure is simply called “clitic-doubling” in Bulgarian, Bulgarian topicalizations with clitic-doubling are comparable to Spanish CLLD (see Ivanov, 2009, 2012; Slabakova & Ivanov, 2011) in regard to their syntactic forms and their discursive functions as topics. Similarly to Spanish CLLD, the direct or indirect object in Bulgarian can move to the left periphery and become the topic of the sentence, as long as it is doubled by a clitic. If, on the other hand, the element that moves to the left periphery needs to be the focus, the sentence does not allow clitic-doubling, as Spanish FF. The author tested the knowledge that
near-natives have of Bulgarian clitics and their pragmatic functions through two separate acceptability judgment tasks. The author identifies problems in the intermediate group, but a native-like behavior on the advanced group (near-natives). Using a similar methodology to Valenzuela (2005, 2006), Ivanov (2009, 2012) tested learner’s knowledge on Bulgarian clitic constructions by using a grammaticality judgment task, to test the knowledge on the syntax of clitics, and a sentence selection task, to evaluate the knowledge at the syntax-discourse interface level. The author found convergence in both tests, and interpreted this results as a challenge to the IH.

In sum, researchers on the L2 acquisition of CLLD were facing contradictory data. While Valenzuela (2005, 2006) found evidence to support that discursive features are not acquirable without permanent optionality as predicted by the IH, Ivanov (2009, 2012) challenged these assumptions by showing data supporting convergence between native and non-native speakers at the syntax-discourse interface.

Slabakova and Ivanov (2011) gave a closer look to this contradictory situation and re-interpreted Valenzuela’s results making two important points. First, the authors pointed out that some conditions were different in Valenzuela’s and Ivanov’s studies. Slabakova and Ivanov argue that L2 speakers of Spanish are not exposed to a categorical use of CLD without clitic-doubling for generic direct objects in Spain. In fact, the control group (native speakers) also overgeneralized CLLD over CLD, showing that the theoretical distinction between both structures is not that clear, at least in Castilian Spanish. Second, since the semantic distinction in the use or absence of the clitic-doubling (CLLD vs. CLD) was native-like, Slabakova and Ivanov

---

5 Even though the structure is called differently in Bulgarian, Bulgarian topicalizations with clitic-doubling are comparable in syntactic form and discursive function as topic to Spanish CLLD (see Ivanov, 2009, 2012; Slabakova & Ivanov, 2011).
believe that the participants were paying closer attention to discourse marks, in a native-like fashion, over semantic marks. Therefore, Valenzuela’s results can be understood as a case of convergence between native and non-native grammars.

Participants’ production was introduced to the scientific debate on the syntax-discourse interface by Donaldson’s (2011, 2012) studies on the acquisition of French language. The author found convergence between native and near-native speakers of L2 French in their use of left dislocations⁶ (2011) and the focus *c’est* clefts and topic *avoir* clefts (2012). Along with being the first attempt to introduce L2 French in this debate, these studies are also relevant because the author performed the first studies using production data to test topicalizations at the interface level. Donaldson, like Ivanov (2009, 2012), interpreted these results as a challenge to the predictions made by the IH.

5.1. CLLD and FF in L2 Spanish: A test case for the IH (Slabakova et al., 2012)

New studies on the acquisition of Spanish CLLD and the IH have been performed recently. Slabakova et al. (2012) tested the knowledge of intermediate, advanced, and near-native speakers of L2 Spanish on CLLD and FF. Participants had to rate two possible answers from 1 to 4 indicating how appropriate they considered those answers to be, given a question and a paragraph explaining the context in which this question was being asked. One of the answers was always felicitous to that context and the other one was not.

The types of possible structures used by Slabakova et al. (2012) were the following. First, two types of CLLD were tested: CLLD Equivalence and CLLD Set-subset. While the former further explains the topic, the latter further explains a sub-category of the topic (see examples

---

⁶ Left dislocations in French are similar to Spanish CLLD and Bulgarian topicalizations in syntactic form and discursive function as topic (see Lopez, 2009).
16 and 17). Second, FF, which adds a new focus to the discourse. Third, some filler used as distractors.

(16) CLLD with set–subset relationship between dislocate and antecedent:

[context: What did the movers do with the furniture?]

Las sillas las dejaron en el pasillo, pero no sé dónde están las mesas.

The chairs they left in the hallway, but I don’t know where the tables are.

(17) CLLD with whole–part relationship between dislocate and antecedent:

[context: What can we do with this table? It doesn’t fit through the doorway]

Mira, las patas las podemos doblar.

Look, the legs we can fold.

Near-native learners performed native-like on both CLLD structures. Advanced learners acted native-like on CLLD Set-subset but non-native-like on CLLD Equivalence. Intermediate learners were not able to distinguish how acceptable felicitous and unfelicitous structures were. The authors interpreted the near-native learners’ results on CLLD as showing a native-like behavior. Intermediate learners seemed to have acquired the syntactic knowledge but not the discursive knowledge. Advanced learners were interpreted as a mid-step between intermediates and near-natives. Developmentally, it seemed like learners first acquired syntax, then developed discourse knowledge, and finally behaved native-like. Therefore, the authors claimed that the results were a challenge to the IH. The authors also found that all groups of learners acted native-like with respect to Spanish FF. This convergent behavior on FF between native and
non-native speakers was interpreted as a cross linguistic influence from L1 English to L2 Spanish, given the similarities between Spanish and English in the use of objects in the left periphery with prosodic stress to mark focus.

The most recent study in the acquisition of Spanish CLLD was performed by Leal Mendez et al. (2015). The authors replicated Slabakova et al.’s (2012) study and included a group of monolingual speakers (control group), a group of adult bilingual speakers whose L1 was Spanish, a group of intermediate heritage speakers of Spanish, and a group of advanced heritage speakers of Spanish. The study intended to test if L1 attrition was present in the grammar of groups of Spanish speaker who also speak English. Similarly to Slabakova et al. (2012), Leal Mendez et al. (2015) found that all three experimental groups performed similarly to the control group, challenging the predictions made by the Interface Hypothesis.

In sum, one can observe that every study on Spanish CLLD (or similar structures in other languages) has shown evidence challenging the IH (see re-interpretation of Valenzuela (2005, 2006) by Slabakova and Ivanov (2012)). While studies on anaphora resolution and subject pronouns performed by Sorace and colleagues (see section 2) provide supporting evidence for the IH, several studies on the acquisition of CLLD and similar structures (Ivanov, 2009, 2012; Donaldson, 2011, 2012), together with the theoretical criticisms reviewed in section 3, challenge the IH. Therefore, the IH and its predictions are a good area in which to research and learn more about L2 acquisition. The study by Slabakova et al. (2012) described in section 5.1. is especially important for the purposes of this thesis since, as previously explained, the empirical section of this study is based on Slabakova et al.’s study, but analyzing production data instead of data from metalinguistic tasks.
PART II

WHAT THIS THESIS INTENDS TO ADD TO THE DISCUSSION

6. Data from production tasks

The main goal of this thesis is to investigate whether or not the previous data on the acquisition of Spanish CLLD collected using non-production methodologies represents what learners produce in production tasks. The methodologies that have been used so far are acceptability judgment tasks (Valenzuela, 2005, 2006; Slabakova et al., 2012; Leal Mendez et al., 2015), sentence selection tasks (Valenzuela, 2005, 2006) and sentence completion tasks (Valenzuela, 2005, 2006).

Sorace (2011) questions the use of methodologies such as acceptability judgment tasks with language learners to test the IH. The author argues that participants may provide answers based on metalinguistic knowledge in this type of task, especially if they are not timed. Instead, as explained in section 2, Sorace and colleagues claim that research should look at processing factors rather than language knowledge. The author highlights this concern by saying that:

[...] the method typically used in generative linguistic research, acceptability judgment tasks, is not the most suitable to capture states of optionality or inconsistent integration of information in processing [...].

The acceptability judgment task is a valid tool to analyze the natural intuition that native speakers have about their own language, but it has been called into question when used with learners, due to reasons such as those expressed above. Instead of using acceptability judgment tasks with learners, this thesis will approach the issue of the acquisition of Spanish CLLD by learners by using learners’ production data, which has never been used to study the acquisition of Spanish CLLD.
Production tasks have been successfully used in the study of other syntactic structures in relation to the IH. For example, Tsimpli and Sorace (2006) used this method to collect data about the use of articles and clitics (see section 2). Belletti et al. (2007) and Montrul and Rodriguez-Louro (2006) also used a production task to study pre- and post-verbal positioning of overt and null subjects. In fact, while the latter found divergence between native and non-native speakers, the former found convergence. This shows that production tasks are not biased to support or challenge the IH. The most similar study to this thesis using a production task was done by Donaldson (2011, 2012), who found convergence between native and non-natives speakers in the use of French left dislocations.

Syntactic structures such as Spanish CLLD are rarely found in natural production. For example, Donaldson (2011) needed to create a corpus with 20 speakers and 10 hours of transcriptions in order to get enough tokens to be able to make generalizations. Due to the low use of this type of structure, this thesis uses a “biased” production task that will induce participants to produce the desired structures. A production task in which the researcher elicits the production of a certain structure was used by Rothman (2009), who asked the participants to translate sentences from English into Spanish in order to obtain sufficient data on pre- and post-verbal positioning of overt and null subjects. The “biased” production task used in this thesis is further explained in section 10.

7. Research questions

Considering the questions and possibilities opened by some studies in second language acquisition regarding topicalizations and the IH, such as Ivanov (2009, 2012), Donaldson (2011,
2012), Slabakova et al. (2012), or Leal Mendez et al. (2015), this study attempts to answer the following research questions:

- What are the stages in the acquisition of both syntax and discourse restrictions when acquiring CLLD in L2 Spanish?
  - Do learners acquire the skill of dislocating to the left periphery and the use of clitic-doubling simultaneously or sequentially when acquiring CLLD in L2 Spanish?
  - Do learners acquire CLLD and its correct discursive use simultaneously or sequentially in L2 Spanish?
  - Are CLLD-Equivalence and CLLD-Set-subset acquired simultaneously or sequentially in L2 Spanish?
  - Do learners of L2 Spanish acquire the discourse restrictions of Spanish CLLD without residual optionality?

- Does the data from acceptability judgment tasks on the acquisition of CLLD in L2 Spanish (Slabakova et al., 2012) represent what learners produce in production tasks?

8. Predictions

Given the scientific debate presented in sections 2 to 5 of this paper, two different predictions could be made about the results that the present study will show. This thesis may find either convergence or divergence between native and non-native speakers of Spanish. Let us describe what those results may look like in more detail taking into account the experimental groups of participants explained in section 9.

8.1. Hypothesis A: Learners’ convergent behavior
If this study finds convergence between native and non-native speakers, those results would be challenging evidence that would call the IH into question. Some group-by-group predictions can be drawn based on the data found by Slabakova et al. (2012), imagining what these data would look like in a production task. Slabakova et al.’s near-native participants show native-like behavior at the syntax-discourse interface in the use of L2 Spanish CLLD. Because of this result, the prediction for this thesis would be that near-native speakers would show evidence of mastering the syntax of CLLD and would not show optionality in the discursive use of this syntactic structure. Slabakova et al.’s advanced learners seem to distinguish between felicitous and infelicitous CLLD Set-subset examples, but not for CLLD Equivalence. This justifies predicting that advanced participants in this study would master the syntax of CLLD (see Syntax-Before-Discourse Hypothesis (Montrul & Rodriguez-Louro, 2006; Rothman, 2009) in section 3 of this thesis), but may be in the process of acquisition of all the discursive details of the correct use of Spanish CLLD. Since Slabakova et al.’s (2012) intermediate participants do not distinguish between felicitous and infelicitous sentences, the prediction for this thesis is that intermediate participants would not show any knowledge of the discursive constraints that determine the correct use of Spanish CLLD. Results may show that they have not acquired either left dislocations or clitic-doubling, or may be in the process of acquiring them.

8.2. Hypothesis B: Learners’ divergent behavior

If this study finds divergence between native and non-native speakers, those results would be supporting evidence for the IH. Based on the original interpretation of the data found by Valenzuela (2005, 2006) on the acquisition of Spanish CLLD by using acceptability judgment, sentence selection, and sentence completion tasks, some predictions can be made for
the production task used in this thesis. Valenzuela’s near natives differ in their responses to the native speakers and overextend CLLD beyond the discursive situations in which it is used by native speakers. The prediction for this thesis would be that the near-native speakers may show permanent optionality at the syntax-discourse interface. These participants would not be consistent in their responses to the cases in which CLLD is needed, and they would overextend CLLD. This means that they would produce CLLD even in contrastive focus context, where CLLD is not pragmatically licit.

Even though Valenzuela (2005, 2006) does not include other levels of proficiency in her study, predictions can be made based on other studies that show evidence supporting the IH by looking at other syntactic structures. Tsimpli and Sorace (2006) interviewed three groups of L1 Russian / L2 Greek speakers. The groups were made based on the participants’ length of stay in Greece. The authors found that speakers did not have any problems acquiring knowledge internal to language, but also found residual optionality at the syntax-discourse interface. In fact, they found that participants from all groups overextended overt subject pronouns, showing “no developmental pattern” (p. 659). Length of stay in the country does not have any effect on improving discursive accuracy. Therefore, the prediction for a production task would be a differentiation between native and non-native speakers. While the latter would show consistency in their answers, the former, at any level, would show optionality.

Valenzuela (2005, 2006) and Tsimpli and Sorace (2006) do not mention syntax explicitly, but both agree with Sorace (2011) that domains internal to language are acquirable for non-native speakers without permanent optionality. Therefore, this thesis predicts that

---

7 The authors do not clarify whether or not length of stay in the country implies different levels of proficiency.
near-natives will show native-like behavior from a syntactic point of view, but no predictions can be made about the developmental levels for the intermediate and advanced participants.
PART III

METHODOLOGY AND SCOPE

9. Participants

Twenty-five individuals participated in the study; eighteen non-native speakers and seven native speakers. Out of the seven native Spanish speakers, five of them were from Spain and two were originally from Puerto Rico. No dialectal differences were found between the Spaniards and the Puertorican. Fifteen native and non-native participants were discarded for falling outside of the scope of the study and due to technical problems.

The eighteen non-native individuals were grouped in three different levels of proficiency, following the procedure used by Slabakova et al. (2012). All of the participant completed a test based on the Diploma de Español como Lengua Extranjera (DELE, see Appendix B). The DELE test has been used in several studies (Montrul, 2002, 2004; White et al., 2004, Slabakova et al., 2012, Hoot, 2012) using many different score ranges to create or evaluate groups of participants. Due to the lack of consistent score ranges, this study follows the score ranges included in the answer key of the test. With a maximum of 50 points, counting one point for each right answer, four of the participants obtained a result between 46 and 50 points and were placed in a group coded as “near-native.” Eight of the participants obtained a result between 39 and 45 and were coded as “advanced.” The other six participants obtained a result between 25 and 38 and were coded as “intermediate.”

Additionally, all non-native participants completed the second test that Slabakova et al. (2012) used in their study to ensure the participants’ previous knowledge of clitics. In this test, participants need to answer all ten questions correctly, with five possible answers to each
question. Participants obtain points by marking the correct answers and not marking the incorrect ones. “Near-native” participants obtained results of 48 and higher, “advanced” participants obtained results of 44 and higher, and “intermediate” participants obtained results of 40 and higher.

As mentioned in previous sections, the goal of this thesis is to replicate Slabakova et al.’s (2012) study in order to be able to compare both sets of data. For this reason, the procedure of recruiting and grouping non-native participants was designed as closely as possible to the method these authors used. Table 1 compares the values and ranges used by this study and by Slabakova et al.’s.

TABLE 1: Score ranges in the DELE and clitic tests used by this thesis and by Slabakova et al. (2012) to create groups of proficiency.

<table>
<thead>
<tr>
<th></th>
<th>DELE Test</th>
<th>Clitics Test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>This thesis</td>
<td>Slabakova et al. (2012)</td>
</tr>
<tr>
<td>Near-Native</td>
<td>46-50</td>
<td>47-50</td>
</tr>
<tr>
<td>Advanced</td>
<td>39-46</td>
<td>40-46</td>
</tr>
<tr>
<td>Intermediate</td>
<td>25-38</td>
<td>30-39</td>
</tr>
</tbody>
</table>

Despite the differences in the scoring ranges, the important considerations are that 1) participants in this thesis demonstrated as much or more knowledge on clitics as those in Slabakova et al.’s (2012) study (see right half in Table 1), and 2) the near-native group was created with the same scoring range that Slabakova et al. used, which is the most important group to test the IH. Even though 46 is considered “near-native” in this study and “advanced” in Slabakova et al.’s study, none of the participants in this thesis scored 46, so all “near-natives”
actually scored between 47 and 50. All except one (Figure 1) of the figures used to present the data in Part IV are based on a different criteria to divide the participants. Therefore, the differences in the score ranges in the DELE test for the advanced and intermediate groups do not make a difference in the discussion and conclusions of this thesis.

10. Methodology

The main experimental activity is a biased oral production task, which extends beyond the methodologies used by Donaldson (2012), Ivanov (2012), Slabakova et al. (2012), and Leal Mendez et al. (2015). Instead of judging the acceptability of the sentence in the right context (Donaldson, 2012) or deciding which of the possible answers fit better in the discourse (Ivanov, 2012; Slabakova et al., 2012; Leal Mendez et al., 2015), the participants in this study have to produce the answer that they consider appropriate given the context.

As discussed in section 6, the goal of this study is to add oral production data to the academic discussion on CLLD. However, obtaining this type of construction from natural production would be difficult, since there are other possible structures that can be used in the same discursive context, like sentences following the canonical word order, clefts, etc. This study will create the production of Spanish CLLD.

Participants were presented with twenty-four trials. Each of the trials is divided into two steps. In the first step, the participants read a short paragraph from one to three sentences long that discusses a fictional situation about the participant and another person. The paragraph is written in English for the Spanish learners and in Spanish for the native Spanish speakers. The participant is asked to read the paragraph out loud and to himself or herself to make sure they understand it. The objective is to make sure that the participants understand the details of the
fictional situation, regardless of their native language. In the second step, the participant reads a question asked by the other fictional person, and he or she has to use a given group of words to begin constructing an answer to that question. The initial prompt provides the bias mentioned above. Given the context and the question, the initial group of words requires the formation of a CLLD. (18) is one of the trials presented to the participants:

(18) Ves a tu amigo Mario en Valencia los viernes. Tu amiga Alicia te pregunta:

You see your friend Mario in Valencia on Fridays. Your friend Alicia asks:

Alicia: ¿Dónde ves a Mario los viernes?

Where do you see Mario on Fridays?

Usted: A Mario…

Mario…

The expected answer from the control group of native speakers would be a sentence similar to:


Mario, I see him in Valencia.

The two sentences in the beginning of example (18) is the context in which the fictional situation and conversation develop. Alicia’s question provides the discourse and the topic, a Mario “Mario,” which is also the beginning of the answer that the participant needs to complete. A Mario being the first constituent of the answer forces the participants to use it as a direct object due to the context and the personal “a,” which in Spanish indicates that the element is accusative and human. Therefore, participants are forced to reply with a left dislocation. This way, the only possible grammatical constructions that the participant can use are CLLD, FF, clefts, and focal
adverbs (see section 4). However, even if the participant can form those types of sentences, he or she still has to choose which one of them is felicitous given a certain context and discourse.

The twenty-four trials are divided in four groups following the four categories used by Slabakova et al. (2012), depending on the type of answer expected. The category CLLD Equivalence is formed by those six trials in which the given group of words that begin the answer is the same as the topic of the question (e.g. ¿A Mario? - A Mario). The category CLLD Set-Subset is formed by those six trials in which the given group of words that begin the answer is a sub-category of the topic of the question (e.g. ¿A tus amigos? - A Mario). The category Focus is formed by those six trials in which the given group of words that begin the answer is different than the topic of the question with the intention of adding information that is not included in the question (e.g. ¿A Mario? - ¡A Enrique!). The category Fillers is formed by random trials used as distractors. These four types of trials are mixed in a random order, as shown in Appendix A.

11. Considerations in the Design

This thesis focuses on the dislocation of direct objects that refer to humans. The case marker “a” required for a constituent with accusative case that refers to a human is key in the design of the experiment. Including the preposition “a” in the beginning of the prompt offered to the participant forces him or her to structure his or her answer with CLLD in the topic trials, and with FF, clefts, or focal adverbs in the focus trials. If the study used non-human direct objects or indirect objects without the preposition “a,” other structures (such as passive voice) could be used to create a felicitous answer. However, those structures would not provide the type of data that this study is seeking because they may not be topicalizations or focalizations.
Other syntactic features are common in all the trials. All the direct objects with which participants have to start their answers are [+specific], [+definite], [+animate], [+human], and masculine. All the questions that the participants have to answer are presented in present simple tense. The goal is to avoid these features becoming variables that influence the results. This way, the results will be based only on the variables “topicalizations vs. focalization” and “level of proficiency.”

To avoid the vocabulary presenting a problem of understanding, especially for the intermediate learners, all the words in the questions and answers can be found in *A Frequency Dictionary of Spanish* (Davies, 2006), which includes the 5,000 most common words in Spanish. Also, the length of the dislocated direct objects is controlled and limited to a maximum of seven syllables to avoid the same possible problem. The trials were then reviewed by several professors of Spanish for appropriateness before being presented to the participants.
12. Results by Level of Proficiency

12.1. Dislocation and Clitic-Doubling

Figure 1 compares the evolution of the skill of dislocating the direct object to the left periphery and the skill of using clitic-doubling. “Dislocation” refers to the responses in which the given group of words for the response was used as the direct object of the sentence. “Clitic-doubling” refers to the responses in which a clitic pronoun was used as a duplication of the direct object. A speaker must perform these two actions – dislocating and clitic-doubling - in order to form a CLLD construction.

Figure 1 is based only on those trials designed to test the topic condition, for which the native responses are CLLD categorically, so the syntax is taken into account without considering discourse. The green bars in the figure show that, from left to right, intermediate learners perform left dislocation at a rate of 83%, and advanced learners at a rate of 99%. On the other hand, the red bars show that, also from left to right, clitic-doubling was performed at a rate of 21% by intermediates, at a rate of 61% by advanced learners, and at a rate of 87% by near-natives. Native speakers show 100% for both dislocations and clitic-doubling, in green and red respectively.
12.2. Codes at the Data Set

The codes “CLLD,” “CLEFT,” and “FF” represent the type of sentences performed. A response was coded as “CLLD” if there was a left dislocation of the direct object and a clitic-doubling duplicating the direct object (e.g. *A Pedro lo veo en el parque* “Pedro, I see him in the park). A response was coded as “CLEFT” if there was a verb copula and the given group of words was not used as the subject and if there was no clitic pronoun duplicating the direct object (e.g. *A Dave es a quien conozco* “Dave is the one that I know”). No response included a copula with a clitic-doubling, but the second category was included as a requisite in case some participant responded using non-native like innovation. A response was coded as “FF” if there was a left dislocation of the direct object, no clitic pronoun duplicating the direct object and the subject of the sentence was located after the verb or it was null (e.g. *A tío Paco voy a visitar* “Uncle Paco I am going to visit”).

**FIGURE 1.** Dislocations and clitic-doubling performed under topic condition, by level of proficiency.
The non-native like utterances were coded in two different ways: “Invalid / Ungrammatical” and “CAN+.” A large number of responses presented a pattern that would have disappeared from the data if coded simply as “Invalid / Ungrammatical.” Many responses by the non-native participants presented a left dislocation, no clitic-doubling, and a pre-verbal overt subject (e.g. #A Luis yo voy a llevar “Luis I am going to bring”). These types of sentences are considered ungrammatical (Dominguez, 2004; Hualde, 2012), but they are coded as “CAN+,” meaning “canonical sentence plus a left dislocation” because these sentences seem to be a developmental step previous to the use of focus fronting and CLLD. Considering these sentences as simply “ungrammatical” would hide part of the acquisition process.

12.3. Topic and Focus Conditions

Figure 2 shows the type of sentences performed by the participants from the three different levels of proficiency for topic and focus conditions. Results show an increase in the percentage of CLLD used for the topic condition, 28% for the intermediates, 61% for the advanced group, 88% for the near-native participants, and 100% for the natives (second, fourth, sixth, and eight bars from the right, CLLD in blue). On the other hand, the focus condition shows an unexpected decrease in the contrastive focus constructions from the intermediates with a 42% to the near-natives with a 33%. Native speakers use contrastive focus constructions in 80% of the cases (first, third, fifth, and seventh bars from the right, focus constructions in two different tones of orange).
The data presented in figure 2 may be understood as a higher degree of accuracy by learners at lower levels of proficiency in the focus condition, since the intermediate 42% is closer to the native 80% than the near-native 33%. This counterintuitive result is simply the result of the study design, since the results from the DELE test may not group the participants properly for the objectives of this thesis. When preparing the data set, it was obvious that there were three different tendencies that were not represented in the results of the DELE. There is a group of learners that underuse clitic-doubling, a second group that uses the clitic-doubling properly in the topic condition but overextend it in the focus condition, and a third group that, comparable to the native speakers, used a high rate of clitic-doubling for the topic condition and a low rate for the focus condition. Therefore, grouping participants by their results in the DELE test obscured the real developmental pattern.
Table 2 shows the sample answers of three different non-native participants to the same two statements: ¿Cuándo ves a nuestros amigos de Salamanca? “When do you see our friends from Salamanca?” (in a topic case) and ¡Quiero cuidar a David! “I want to take care of David!” (for a focus case). While the first participant did not use clitic-doubling in either response, the second one used it in both, and the third one used it only for the topic condition as native speakers do. The focus example for participant #1 is not coded as felicitous or infelicitous since the lack of knowledge of clitic-doubling does not allow one to know if the participant has any discursive knowledge.

### TABLE 2: Samples responses by one native and three non-native participants to the same pair of questions, one of them topic condition and the other one focus condition.

<table>
<thead>
<tr>
<th>Part.</th>
<th>Condition</th>
<th>C-D</th>
<th>Felicitous?</th>
<th>Samples response</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>Topic</td>
<td>No</td>
<td>No</td>
<td>#A Manuel voy a visitar la semana que viene. &quot;#Antonio I am going to visit next week”</td>
</tr>
<tr>
<td></td>
<td>Focus</td>
<td>No</td>
<td>?</td>
<td>?¡No! ¡A Antonio vas a cuidar! &quot;#No! Antonio you are going to take care of”</td>
</tr>
<tr>
<td>#2</td>
<td>Topic</td>
<td>Yes</td>
<td>Yes</td>
<td>A Manuel lo veo la semana que viene. “Manuel, I am going to visit him next week”</td>
</tr>
<tr>
<td></td>
<td>Focus</td>
<td>Yes</td>
<td>No</td>
<td>¡No! ¡A Antonio le vas a cuidar! &quot;#No! Antonio, you are going to take care of him”</td>
</tr>
<tr>
<td>#3</td>
<td>Topic</td>
<td>Yes</td>
<td>Yes</td>
<td>A Manuel lo veo la semana que viene. “Manuel, I am going to visit him next week”</td>
</tr>
<tr>
<td></td>
<td>Focus</td>
<td>No</td>
<td>Yes</td>
<td>¡No! ¡A Antonio es a quien vas a cuidar! “No! Antonio is the one you are going to take care of!”</td>
</tr>
<tr>
<td>Native</td>
<td>Topic</td>
<td>Yes</td>
<td>Yes</td>
<td>A Manuel le voy a visitar la semana que viene. “Manuel, I am going to visit him next week”</td>
</tr>
<tr>
<td></td>
<td>Focus</td>
<td>No</td>
<td>Yes</td>
<td>¡No! ¡A Antonio es al que vas a cuidar la próxima semana! “No! Antonio is the one you are going to take care of!”</td>
</tr>
</tbody>
</table>
The data shows that there is an overlap between tendencies of performance and levels of proficiency. Each participant was regrouped in new categories: “underextender” if the participant used less than 50% of clitic-doubling in both the topic and the focus conditions, “overextender” if the participant used 50% or more of clitic-doubling in both the topic and the focus condition, and “native-like speaker” if the participant used less than 50% of clitic-doubling in the focus condition and 50% or more of clitic-doubling in the topic condition.

Figure 3 shows that 67% of the intermediates and 38% of the advanced learners were underextenders (grey sections of the bars) while 50% of the near-natives and 13% of the advanced learners were native-like speakers (green sections of the bars). Therefore, the underextension state can be considered the developmental stage before the native-like stage, with overextension being the step in between (red sections of the bars). This shows that the three tendency groups are not random, but they represent a series of steps in the process of acquisition of the syntactic constructions studied and their discursive restrictions. Grouping participants by tendency, instead of level of proficiency, offers the possibility of creating groups of participants that are developmentally similar, which may show the evolution of the conditions and constructions studied in this thesis more accurately.
Two additional comments must be made about the grouping of participants. The first comment is on the DELE test (see Appendix B), which did not group the participants efficiently for the purposes of this thesis. As any other proficiency test, results are imprecise. The test is based mostly on lexical knowledge, while this study looks at syntax. Additionally, the vocabulary tested is specific to Spain, which does not represent the vocabulary that language learners of Spanish hear in North America. The second comment is on the factor that determines the new tendency groups. This study did not find any pattern within the new groups, but there may be extra-linguistic factors - such as study abroad or careers relating to the Spanish language - that participants of the same tendency may have in common. This thesis was not designed to find those sociolinguistic elements, so those are left for future research.

13. Results by Tendency

13.1. Topic and Focus Conditions by Tendency
After understanding that tendency would be more informative than level of proficiency for the objectives of this thesis, the data was rearranged following this new criterion. Figure 4 shows the type of constructions produced in topic and focus conditions, this time organized by tendency. While the underextenders (n=7) use CLLD at a rate of 6% for the topic condition and 9% for the focus conditions (first two bars from the right, CLLD in blue), the overextenders (n=8) use CLLD over 90% of the topic stimuli and 88% of the focus stimuli (third and fourth bars from the left, CLLD in blue). The rate of clitic-doubling stays high (83%) among the native-like speakers (n=3) (sixth bar from the left, CLLD in blue), while the rate of contrastive focus constructions increases up to a 72% (fifth bar from the left, focus constructions in orange), showing a more similar situation to the results from the native speakers with 100% of CLLD in the topic condition (first bar from the right, CLLD in blue) and 81% of contrastive focus constructions in the focus condition (second bar from the left, focus constructions in orange). See Table 2 in section 12.3. for sample responses that represent a general overview of the performance of the participants at the three tendency groups.
Given that the results of the native speakers and the native-like speakers are similar, but not quite the same, the individual results of the native-like speakers are presented in detail in Figure 5. Participant #1 uses CLLD at a rate of 100% for the topic condition (sixth bar from the left, CLLD in clue) and clefts at a rate of 83% in the focus condition (fifth bar from the left, clefts in orange), very similarly to the average of the native speakers with 100% and 81% respectively (first and second bar from the right, CLLD in blue and clefts in orange). Participant #2 uses CLLD in 100% of the topic conditions cases (fourth bar from the left, CLLD in blue) and focus fronting in 50% of the focus condition cases (third bar from the left, FF in red). Finally, participant #3 uses CLLD at a rate of 50% for the topic condition (second bar from the left, CLLD in blue) and focus fronting at a rate of 83% for the focus condition (first bar from the left, FF in red).
Table 3 shows the sample answer to one of the focus condition stimulus performed by one of the native speakers, participant #1, and participant #2 from figure 5. While participant #1 performed a native like cleft with a copular construction, participant #2 performed a FF with a null subject.

**TABLE 3: Sample responses for a concrete focus condition stimulus, by type/number of participant.**

<table>
<thead>
<tr>
<th>Focus Condition Stimulus</th>
<th>Answer</th>
<th>Participant</th>
</tr>
</thead>
</table>
| ¡Quiero cuidar a David!  | ¡No! ¡A Antonio es al que vas a cuidar la próxima semana!  
“I want to take care of David!” | Native |
|                          | ¡No! ¡A Antonio es a quien vas a cuidar!  
“No! Antonio is the one that you are going to take care of!” | #1 |
|                          | ¡No! ¡A Antonio vas a cuidar!  
“No! Antonio you are going to take care of!” | #2 |
13.2. Equivalence and Set-Subset Topic Conditions

As explained in previous sections, the stimuli to test the topic condition was subdivided in two categories: Equivalence (e.g. ¿A Pedro? - A Pedro) and set-subset (e.g. ¿A tus amigos? - A Pedro). Figure 6 shows the rate of CLLD performed under those two types of stimuli by the participants from the different tendencies of performance. While the underextenders use CLLD in 7% of the equivalence stimuli (second bar from the left, CLLD in blue) and 5% of the set-subset stimuli (first bar from the left, CLLD in blue), the overextenders use CLLD at a rate of 92% and 88% (forth and third bars from the left, CLLD blue) and the native-like speakers at a rate of 78% and 89% (sixth and fifth bars from the left, CLLD in blue).

FIGURE 6. Type of structure performed under the equivalence and the set-subset topic conditions, by tendency.
PART V
DISCUSSION AND CONCLUSIONS

14. Revisiting the First Research Question

The first research question is subdivided in four questions that ask about many relevant points on the developmental process in the acquisition of Spanish CLLD and its discursive restrictions: 1) the acquisition of the syntax of left dislocation and the clitic-doubling, 2) the acquisition of the discursive restrictions on Spanish CLLD, 3) the acquisition of two types of CLLD (Equivalence and Set-subset), and 4) residual optionality in the acquisition of CLLD and its discursive restrictions. These four points will be addressed in this first section of the discussion.

- Do learners acquire the skill of dislocating to the left periphery and the use of clitic-doubling simultaneously or sequentially when acquiring CLLD in L2 Spanish?

The information in Figure 1 can explain the acquisition process of left dislocations and clitic doubling. On the one hand, the rate of left dislocations for the topic condition is at a rate of 83% even for the intermediate learners, and this becomes almost categorical at an advanced level (99%) and categorical at a near-native level (100%), exactly like native speakers. On the other hand, the rate of clitic-doubling for the topic condition rises from 28% for intermediates, to 61% for advanced, and to 88% for near-natives. This may be interpreted as evidence that the ability to use left dislocations and the ability to use clitic-doubling are acquired sequentially and not simultaneously. While the former is almost perfected by an advanced level, near-natives, as a group, still do not always use clitic-doubling. Therefore, left dislocations seem to be acquired first and then clitic-doubling.
It is worth mentioning that the personal “a,” the case marker that indicates that a constituent is accusative and human, is not very perceptually salient for native and non-native speakers of Spanish. This could play a role in the responses of some of the participants, especially those at the lower levels of proficiency. The fact that some participants reply using the given accusative constituent as nominative, may mean that they did not perceive that the case marker required them to use the given group of words as a direct object. However, they may still have understood the same constituent as a direct object in the question due to the context.

- Do learners acquire CLLD and its correct discursive use simultaneously or sequentially in L2 Spanish?

Figure 4 shows data on the type of construction used in topic and focus conditions by tendency groups. The increase of CLLD from 6% for the topic condition and 10% for the focus condition at the underextension stage to 90% and 88% respectively at the overextension stage, may be evidence of the acquisition of the clitic-doubling at this level. However, the similar increase in the use of CLLD for both conditions show that the discursive restrictions in the use of CLLD, FF, and clefts have not been acquired. In other words, the overextenders know how to use the clitic-doubling, but do not distinguish when it is felicitous to use it and when it is not.

Also in Figure 4, native-like speakers show a sudden decrease in the use of CLLD for the focus condition, with an increase of FF and clefts. This could be understood as evidence of acquisition of the discursive restrictions of the focus condition as different from the topic condition. In other words, these participants seem to know not only how to use clitic-doubling, but also when it is appropriate and when it is not. However, as a group, “native-like” participants
show results exactly like the native control group for the topic condition, but there is a difference in the percentage for the focus condition.

Let us remember that the three participants coded as “native-like” were grouped as such because the showed a 50% or more use of clitic-doubling at the topic condition and less than 50% at the focus condition. Further, Figure 5 divides the data of the three participants individually and compares them to the results from the native speakers. Two of the participants (#2 and #3) fulfill the “native-like” criteria, but a detailed analysis shows that their performance is in fact much different than the native speakers due to their use of FF. Even though FF is described in the literature as a possible option for the focus condition, this construction is categorically avoided by native speakers in this study. The other participant (#1), however, performed very much native like, with categorical use of CLLD for the topic condition and a 83% use of clefts in the focus condition (natives used clefts 81% for focus condition).

To sum up, while CLLD is used almost native-like by participants at a syntactic level even before they are “native-like” or near-natives, the ability of distinguishing between topic and focus conditions happens much later. Turning back to the research question, L2 learners in this study seem to acquire the syntax of CLLD first and then its discursive restrictions.

- Are CLLD-Equivalence and CLLD-Set-subset acquired simultaneously or sequentially in L2 Spanish?

Figure 6 shows the constructions performed under the equivalence and set-subset topic conditions, categorized by tendency. The evolution from 7% use of CLLD-Equivalence and 5% use of CLLD-Set-subset at the underextension stage to 92% and 88% respectively at the overextension stage shows that these structures have been acquired by the time learners reach
this second stage. However, there seems to be no evidence of any differentiation between one kind of CLLD and the other one. Even though CLLD-Equivalence shows percentages slightly higher than CLLD-Set-subset for the “underextenders” and “overextenders,” this prediction is in fact falsified by the “native-like” participants with rates of 78% and 89% respectively. Therefore, CLLD-Equivalence and CLLD-Set-subset do not seem to show any pattern of acquisition based on the data set used for this study.

- Do learners of L2 Spanish acquire the discourse restrictions of Spanish CLLD without residual optionality?

Figure 4 seems to suggest optionality by the “native-like” participants. The percentages in the use of contrastive focus constructions for the focus condition differ, with an 81% for the natives and 72% for the “native-like” non-natives. The results for the topic condition may seem even more relevant since CLLD seems to be categorical for the native speakers while “native-like” non-natives use it 83% of the time. However, these differences are a result of merging individual data into one group. A closer analysis of the individual results presented in figure 5 shows that participant #1 used CLLD categorically for the topic condition and clefts for the focus condition at a rate of 83%. It is important to note that clefts, or contrastive focus constructions in general, are not categorical for native speakers either. Percentages are in fact very similar, with 83% of cleft for participant #1 and 81% of clefts for the native group. Only a further statistical analysis would clarify if that difference of 2% is significant or not. This small difference seems to indicate that participant #1 acquired CLLD, clefts, and their discursive restrictions without optionality or instability, but it cannot be claimed with certainty due to the lack of statistics.
14.1. A Note on the Contrastive Focus Constructions

The focus condition is not the main center of interest for this thesis. It is only included to discover whether non-native speakers can acquire the restrictions of the topic condition by contrasting topic and focus. However, this section will briefly analyze 1) the use of the “canonical sentence plus a left dislocation” construction and 2) the use of FF instead of clefts by non-natives.

Many of the participants produced sentences with a left dislocation of the direct object, no clitic-doubling, and pre-verbal overt subject. Although this type of sentence is normally considered ungrammatical (Dominguez, 2004; Hualde, 2012; see section 12.1.), this study codes them as “CAN+” instead of simply “ungrammatical.” By doing this, the data set is able to show a part of the acquisition process that would be hidden otherwise. CAN+ seems to be a first step in the development of FF and CLLD. It may also be the case that FF is not different from CAN+ with a null subject in the participants’ mental grammar. In fact, the high percentage of CAN+ for the underextenders suggests that their percentage of FF may actually be representing CAN+ constructions with a null subject. The design of the experiment of this thesis does not allow for clarification of this issue.

A second note must be made about the use of FF and clefts. While the native speakers used clefts categorically when CLLD was not used in the focus condition, non-native speakers (except one single participant) used FF, and never clefts. This means that 1) the focus condition may actually be acquired after the topic condition (unlike Slabakova et al., 2012 and Tsimpil & Sorace, 2006), and that 2) FF is not a construction used by these native speakers in Spanish
(unlike Slabakova et al., 2012; Leal Mendez et al., 2015; Ivanov, 2009, 2012; see Leal et al., 2014 for a similar conclusion).

The canonical sentences with a left dislocation and the use of FF and clefts are not the center of this thesis since they are part of the focus condition. However, addressing these questions might be a good avenue for future studies.

15. Revisiting the Second Research Question

This section discuss a possible answer to the second research question presented in section 7. The goal is to compare the results of this study with the results presented by Slabakova et al. (2012). The second research question states:

- Does the data from acceptability judgment tasks on the acquisition of CLLD in L2 Spanish (Slabakova et al., 2012) represent what learners produce in production tasks?

As explained in Part I, Slabakova et al. (2012) used grammaticality judgment tasks to answer questions similar to those posed in this paper. However, the offline methodology used by Slabakova et al. is limited because it gives participants the opportunity to use metalinguistic knowledge in making their judgments. The experiment of this thesis is based on Slabakova et al.’s study, but uses a method in which metalinguistic analysis is minimized, which helps results to be based more on linguistic intuition rather than on analysis (Sorace, 2011, 2012). A second limitation in Slabakova et al.’s (2012) study is that their results are only helpful to understand the development of the acquisition of the discursive restriction of Spanish CLLD since sentences are presented to the students, and always in a given discursive context, but do not offer any information about the acquisition process of the syntax of CLLD. The first sub-section of the first research question of this thesis is not answered by Slabakova et al. Even though the
acquisition of syntax is not the focus of their article, this thesis provides data that shows the
general picture of the acquisition of, not only discourse knowledge, but also the development of L2 syntax.

There are, however, other elements in which Slabakova et al.’s (2012) study and the present thesis are comparable. Slabakova et al. provide results on the questions about permanent optionality at the syntax-discourse interface and about the development of CLLD-Set-subset and CLLD equivalence. While Slabakova et al. state that both near-native and advanced learners show discourse sensitivity, the data for this thesis only found three very advanced participants, the “native-likes,” that showed that sensitivity. Also, while Slabakova et al. show that all near-natives performed native-like as a group, this thesis only found one learner who performed exactly like native speakers. These differences might be due either to the level of the participants (participants in this thesis may have had a lower level of proficiency overall) or to the difficulty of the task (the production task used in this thesis may make the task harder). In any case, the important element that both studies have in common is that both found non-native participants who performed like native speakers at the syntax-discourse interface. It is worth re-stating that the same results have been found in many other studies that used offline methodologies (Slabakova et al., 2012; Leal Mendez et al., 2015; Donaldson, 2011, 2012; Ivanov, 2009, 2012). Many studies using methodologies that do not allow much of a metalinguistic analysis (Montrul and Rodriguez-Louro, 2006; Donaldson, 2011, 2012; Rothman, 2009), like the one used in this thesis, seem to find convergence between native and non-native speakers at very advanced levels as well.
A second element to compare between Slabakova et al.’s (2012) study and this paper is the evolution of CLLD-Set-subset and CLLD-Equivalence. According to the authors, near-native participants performed native-like in both subcategories of Spanish CLLD while advanced participants performed native like only in the set-subset category. This is interpreted by the authors as a developmental process in which CLLD-Set-subset is acquired before CLLD-Equivalence. The data set set for this thesis shows that CLLD-Equivalence seems to be performed in higher percentages at the lower stages of acquisition and CLLD-Set-subset shows more correctness at the “native-like” stage. In conclusion, and unlike Slabakova et al., no clear pattern of acquisition is observed for CLLD-Equivalence and CLLD-Set-subset.

16. Revisiting the Thesis’ Predictions

Section 8 predicts two possible scenarios for what the data could look: divergence and convergence. Data shows that one participant performed native-like, which supports the prediction called “learners’ convergent behavior,” even though the predicted and actual scenarios are different. First, the prediction estimated that all near-natives would perform native-like both on the syntactic and discursive knowledge, but results show that only one individual actually did so. Second, the prediction estimated that advanced learners would be able to perform native-like in the CLLD-Set-subset stimuli, but data shows no difference between both types of CLLD. However, the prediction is supported in the sense that it assumes that non-native participants would perform native-like at the syntax-discourse interface, even before the highest level of proficiency. The participant coded as “native-like” due to his tendency that responded almost exactly like the group of native speakers is in fact an “advanced” participant according to the DELE test, and not native-like. This shows that permanent optionality can disappear even before
the last stage of acquisition if participants are grouped in the traditional categories of level of proficiency.

17. Implications for the Interface Hypothesis

The results shown in part IV of this thesis have some implications for the debate about the L2 endstate and the debate about second language acquisition at the interfaces. As explained in section 2, the main claim of the IH (Sorace, 2011, 2012; Tsimili & Sorace, 2006; Sorace & Serratrice, 2009) is that second language learners cannot felicitously use some constructions in real time due to the difficulties in processing at the external interfaces, in which a linguistic element (in this case, syntax) interfaces with a non-linguistic one (in this case discourse). According to this theory, very advanced L2ers stay at a state of permanent optionality or instability that can never be overcome.

Data from this thesis calls the IH into question. Section 13 shows that one participant performed like the natives, with his responses being categorical in the topic condition and non-categorical in the focus condition, but with similar percentages to those shown by the native speakers. The two main characteristics of the latest version of the IH were taken into account in the design of the experiment. First, the experimental task was designed to test the interface between a linguistic domain (syntax) and a non-linguistic one (discourse). According to Tsimili and Sorace (2006) and Sorace and Serratrice (2009) this type of external interface is the origin of problems for L2 learners. Second, participants completed a production task instead of a grammaticality judgment task to avoid a metalinguistic analysis. The goal of this was to take a step away from the more introspective methods, since processing is suggested by Sorace (2011, 2012) as the root of the problems that non-native speakers have at the external interfaces. Despite
the limitations proposed by the supporters of the IH, a non-native speaker was able to perform native-like.

The conclusion that can be drawn from these results is that, even though the acquisition of the discursive restrictions of Spanish CLLD takes place in a very advanced stage of L2 acquisition, the discourse features are actually acquirable for L2ers. Montrul and Rodriguez-Louro (2006) and Rothman (2009) already indicated similar conclusions in a different theory called the Syntax-Before-Discourse Hypothesis. Results from this thesis support the Syntax-Before-Discourse Hypothesis, challenging the IH.

18. Limitations

There are three main ways in which the results of this study are limited. First, a further statistical analysis would have been needed to determine if the obtained percentages would have actually directed the research to the same conclusions. For example, it is not possible to determine if the differences in percentage in the use of clefts between the native speaker groups and the non-native native-like individual are significant. If that difference were found to be significant, the conclusions of this study would be very different from what has been stated throughout the previous sections. Second, and connected to the first limitation, a larger number of participants would have provided much stronger results. The results of each individual participant have too much weight in each group and overall data. Third, and related to the previous limitation, only one non-native participant was able to reach a native-like performance. Even though this shows that the native-like state is achievable for non-native learners, this participant is, in fact, an outlier. A larger sample would show to what extent this native-like endstate is reachable. Fourth, the stimuli were randomized in two different orders for the
non-native speakers versus the native control group. Even though this may not have had any effect on the results of the study, it is worth noting it and avoiding it the future.

19. Conclusions

This section is a summary of the main findings of this thesis. The information is presented in bullet points with the intention of highlighting the most important information:

- The main finding is that, even taking into account the main restrictions of the IH (such as the differentiation between internal and external interfaces (Tsimpli & Sorace, 2006; Sorace & Serratrice, 2009) and even designing an experiment that allows less analysis (Sorace, 2011, 2012), at least some non-native participants seem able to overcome the residual instability at the syntax-discourse interface.

- Discourse knowledge seems to be acquired after syntax is acquired, which supports with the Syntax-Before-Discourse Hypothesis (Montrul & Rodriguez-Louro, 2006; Rothman, 2009).

- These results are similar to those from many other studies in generative second language acquisition (Slabakova et al., 2012; Leal Mendez et al., 2015; Donaldson, 2011, 2012; Ivanov, 2009, 2012).

- Data shows no difference in the acquisition of the two different types of Spanish CLLD - Equivalence and Set-subset, unlike other studies (Slabakova et al., 2012).

- Unlike other studies (Slabakova et al., 2012; Leal Mendez et al., 2015; Donaldson, 2011, 2012; Ivanov, 2009, 2012), this thesis presents data that show the full acquisition process of Spanish CLLD. For example, no other study shows that left dislocations start to be used by non-native speakers at lower levels of proficiency than is the clitic-doubling.
1 (Focus Condition)
Your friend Elena thinks that you know William from Foo Fighters. You want to correct her and tell her that it is Dave the one that you actually know.
- Elena: Conoces a William de Foo Fighters, ¿verdad? (You know W. from FF, right?)
- You: ¡No! ¡A Dave…

2 (Topic Condition - Equivalence)
You see Pedro at the park a lot. Susana asks:
- Susana: ¿Dónde ves a Pedro? (Where do you see P.?)
- You: A Pedro...

3 (Focus Condition)
You do not go to work to take your son Luis to the doctor. Your boss asks a question, but you want to tell him that you are taking Luis, not Pablo, to the doctor:
- Your boss: ¿Por qué llevas al doctor a tu hijo Pablo? (Why are you taking P. to the doctor?)
- You: ¡No! ¡A Luis…

4 (Topic Condition - Equivalence)
You will take your son Antonio to your friend Sandra’s house at 3 p.m. She calls to ask:
- Sandra: ¿A qué hora traes a Antonio? (What time are you bringing A.?)
- You: A Antonio...

5 (Topic Condition - Equivalence)
You will interview some professors at your university on Thursday for your project for you class project with Maria. She calls to ask:
- Maria: ¿Cuándo entrevistas a los profesores? (When do you interview the professors?)
- You: A los profesores...

6 (Topic Condition - Equivalence)
You see you friend Mario in Valencia on Fridays. Your friend Alicia asks:
- Alicia: ¿Dónde ves a Mario los viernes? (Where do you see M. on Fridays?)
- You: A Mario…
7 (Topic Condition - Set-subset)
You will interview the economists tomorrow and the politicians on Friday, for you class project with Maria. She calls to ask:
- Maria: ¿Cuándo entrevistas a los participantes de nuestro proyecto? (When do you interview the participants for our project?)
- You: A los economistas…

8 (Topic Condition - Equivalence)
Eduardo went to your house but could not fix the bathroom. You decided that you are not going to pay him until he fixes the bathroom. Juan asks:
- Juan: ¿Cuándo pagas a Eduardo? (When are you paying E.?)
- You: A Eduardo...

9 (Filler)
You like to speak French, but it is very difficult. Your friend Belen asks:
- Belen: ¿Qué lengua te gusta hablar? (What language do you like to speak?)
- You: A mí...

10 (Topic Condition - Set-subset)
You see Paul McCartney in concert every year. Jose asks:
- Jose: ¿Con qué frecuencia ves a algún músico famoso en concierto? (How often do you see a famous musician in concert?)
- You: A Paul McCartney...

11 (Filler)
You have a book for Alex for his birthday. Elisa asks:
- Elisa: ¿Qué tienes para Alex por su cumpleaños? (What do you have for A. for this birthday?)
- You: Para Alex...

12 (Filler)
Your friend Enrique likes playing the guitar. Veronica asks:
- Veronica: ¿Qué le gusta a tu amigo Enrique? (What does E. like?)
- You: A Enrique...

13 (Topic Condition - Equivalence)
You will visit your old friend Manuel next week. Sara asks:
- Sara: ¿Cuándo visitas a tu amigo Manuel? (When do you visit your friend M?)
- You: A Manuel...
14 (Filler)
You have not seen your friends from Buenos Aires in years. However, you know that Ivan is getting married soon. Your friend Jacinto asks:
- Jacinto: ¿Ves a tus amigos de Buenos Aires últimamente?
  (Have you seen your friends from B.A. lately?)
- You: De Iván...

15 (Topic Condition - Set-subset)
You will visit your cousin Francisco in Guanajuato, Mexico tomorrow. Your sister calls to ask:
- Your sister: ¿Visitas a alguno de nuestros familiares en Guanajuato?
  (Have you visited some of our family members in Guanajuato?)
- You: A Francisco…

16 (Focus Condition)
You bring your old friend Ruben to a party. Laura comments, but you want to tell her that you bring Ruben, not Carlos:
- Carlos: ¡Gracias por traer a Carlos a la fiesta! (Thanks for bringing C. to the party?)
- You: ¡No! ¡A Rubén...

17 (Focus Condition)
You will visit your uncle Paco next Saturday. Your sister asks, but you want to correct her and tell her that you will not visit your uncle Jose, but your uncle Paco:
- Your sister: ¿Cuándo visitas al tío José? (When are you visiting uncle J.?)
- You: ¡No! ¡Al tío Paco...

18 (Topic Condition - Set-subset)
You will visit your old friend from Salamanca, Manuel, next week. Sara asks:
- Sara: ¿Cuándo ves a nuestros amigos de Salamanca? (When do you see our friends from S.?)
- You: A Manuel...

19 (Focus Condition)
Your sons are called David and Antonio. Your friend Natalia thinks that she will be watching (cuidar) David next week, but she will actually watch Antonio. You want to correct her:
- Natalia: ¡Quiero cuidar a David! (I want to take care of D.!
- You: ¡No! ¡A Antonio...
20 (Focus Condition)
Lucia loves Javier, but not Miguel. Your friend Pancho thinks that she loves Miguel, but you
want to correct her:
- Pancho: ¡Lucía quiere mucho a Miguel! (L. loves M!)
- You: ¡No! ¡A Javier...

21 (Filler)
You found a job, but you do not work for Daniel because he pays very little. Your friend
Fernando asks:
- Fernando: ¿Trabajas para Daniel? (Do you work for D.?)
- You: Para Daniel...

22 (Topic Condition - Set-subset)
Of you friends in Valencia, you see Mario on Saturdays and Marcos on Sundays. Alicia asks:
- Alicia: ¿Cuándo ves a tus amigos de Valencia? (When do you see your friends from V.?)
- You: A Mario...

23 (Topic Condition - Set-subset)
Your sons are called David y Antonio. You will take David to the doctor on Thursday, and
Antonio on Friday. Pepe asks:
- Pepe: ¿Cuándo llevas a tus hijos al doctor? (When are you taking your children to the doctor?)
- You: A David...

24 (Filler)
In your restaurant, you meet with the servers every week and with the cooks only once a month.
Your partner asks:
- Your partner: ¿Nos reunimos con todos los empleados del restaurante cada semana?
  (Do we meet with all our employees from the restaurant every week?)
- You: Con los cocineros...
Appendix B: DELE Test

SPANISH PROFICIENCY TEST (DELE)

Part 1 - Multiple Choice Test

Each of the following sentences contains a blank indicating that a word or phrase has been omitted. Select the choice that best completes the sentence.

1. Al oír del accidente de su buen amigo, Paco se puso ___________.
   a. alegre  b. fatigado  c. hambriento  d. desconsolado

2. No puedo comprarlo porque me ____________ dinero.
   a. falta  b. dan  c. presta  d. regalan

3. Tuvo que guardar cama por estar ____________.
   a. enfermo  b. vestido  c. ocupado  d. parado

4. Aquí está tu café, Juanito. No te quemes, que está muy ____________.
   a. dulce  b. amargo  c. agrio  d. caliente

5. Al romper los anteojos, Juan se asustó porque no podía ____________ sin ellos.
   a. discurrir  b. oír  c. ver  d. entender

6. ¡Pobrecita! Está resfriada y no puede ____________.
   a. salir de casa  b. recibir cartas  c. respirar con pena  d. leer las noticias

7. Era una noche oscura sin ____________.
   a. estrellas  b. camas  c. lágrimas  d. nubes

8. Cuando don Carlos salió de su casa, saludó a un amigo suyo: -Buenos días, _____________.
   a. ¿Qué va?  b. ¿Cómo es?  c. ¿Quién es?  d. ¿Qué tal?

9. ¡Qué ruido había con los gritos de los niños y el ____________ de los perros!
   a. olor  b. sueño  c. hambre  d. ladrar
10. Para saber la hora, don Juan miró el ____________.
   a. calendario   b. bolsillo   c. estante   d. despertador

11. Yo, que comprendo poco de mecánica, sé que el auto no puede funcionar sin __________.
    a. permiso   b. comer   c. aceite   d. bocina

12. Nos dijo mamá que era hora de comer y por eso__________.
    a. fuimos a nadar  b. tomamos asiento  c. comenzamos a fumar  d. nos acostamos pronto

13. ¡Cuidado con ese cuchillo o vas a __________ el dedo!
    a. cortarte  b. torcerte  c. comerte  d. quemarte

14. Tuvo tanto miedo de caerse que se negó a __________ con nosotros.
    a. almorzar  b. charlar  c. cantar  d. patinar

15. Abrió la ventana y miró: en efecto, grandes lenguas de __________ salían llameando de las casas.
    a. zorros  b. serpientes  c. cuero  d. fuego

16. Compró ejemplares de todos los diarios pero en vano. No halló ____________.
    a. los diez centavos  b. el periódico perdido  c. la noticia que deseaba  d. los ejemplos

17. Por varias semanas acudieron colegas del difunto profesor a ___________ el dolor de la viuda.
    a. aliviar  b. dulcificar  c. embromar  d. estorbar

18. Sus amigos pudieron haberlo salvado pero lo dejaron _____________.
    a. ganar  b. parecer  c. perecer  d. acabar

19. Al salir de la misa me sentía tan caritativo que no pude menos que___________ a un pobre mendigo
    que había allí sentado.
    a. pegarle  b. darle una limosna  c. echar una mirada  d. maldecir
20. Al lado de la Plaza de Armas había dos limosneros pidiendo _____________.
   a. pedazos  b. paz  c. monedas  d. escopetas

21. Siempre maltratado por los niños, el perro no podía acostumbrarse a ___________ de sus nuevos amos.
   a. las caricias  b. los engaños  c. las locuras  d. los golpes

22. ¿Dónde estará mi cartera? La dejé aquí mismo hace poco y parece que el necio de mi hermano ha vuelto a _____________.
   a. dejarla  b. deshacérmela  c. escondérmela  d. acabármela

23. Permaneció un gran rato abstraído, los ojos clavados en el fogón y el pensamiento _____________.
   a. en el bolsillo  b. en el fuego  c. lleno de alboroto  d. Dios sabe dónde

24. En vez de dirigir el tráfico estabas charlando, así que tú mismo ___________ del choque.
   a. sabes la gravedad  b. eres testigo  c. tuviste la culpa  d. conociste a las víctimas

25. Posee esta tierra un clima tan propio para la agricultura como para _____________.
   a. la construcción de trampas  b. el fomento de motines  c. el costo de vida  d. la cría de reses

26. Aficionado leal de obras teatrales, Juan se entristeció al saber ___________ del gran actor.
   a. del fallecimiento  b. del éxito  c. de la buena suerte  d. de la alabanza

27. Se reunieron a menudo para efectuar un tratado pero no pudieron _____________.
   a. desavenirse  b. echarlo a un lado  c. rechazarlo  d. llevarlo a cabo

28. Se negaron a embarcarse porque tenían miedo de _____________.
   a. los peces  b. los naufragios  c. los faros  d. las playas

29. La mujer no aprobó el cambio de domicilio pues no le gustaba _____________.
   a. el callejero  b. el puente  c. esa estación  d. aquel barrio

30. Era el único que tenía algo que comer pero se negó a _____________.
a. hojearlo  b. ponérselo  c. conservarlo  d. repartirlo
**Part 2 - Cloze Test**

In the following text, some of the words have been replaced by blanks numbered 1 through 20. First, read the complete text in order to understand it. Then reread it and choose the correct word to fill each blank from the answer sheet. Mark your answers by circling your choice on the answer sheet, not by filling in the blanks in the text.

**El sueño de Joan Miró**

Hoy se inaugura en Palma de Mallorca la Fundación de Joan Miró, en el mismo lugar en donde el artista vivió sus últimos treinta y cinco años. El sueño de Joan Miró se ha __________(1). Los fondos donados a la ciudad por el pintor y su esposa en 1981 permitieron que el sueño se __________(2); más tarde, en 1986, el Ayuntamiento de Palma de Mallorca decidió __________(3) al arquitecto Rafael Moneo un edificio que __________(4) a la vez como sede de la entidad y como museo moderno. El proyecto ha tenido que __________(5) múltiples obstáculos de carácter administrativo. Miró, coincidiendo __________(6) los deseos de toda su familia, quiso que su obra no quedara expuesta en ampulosos panteones de arte o en __________(7) de coleccionistas acaudalados; por ello, en 1981, creó la fundación mallorquina. Y cuando estaba __________(8) punto de morir, donó terrenos y edificios, así como las obras de arte que en ellos __________(9).

El edificio que ha construido Rafael Moneo se enmarca en __________(10) se denomina “Territorio Miró”, espacio en el que se han __________(11) de situar los distintos edificios que constituyen la herencia del pintor.

El acceso a los mismos quedará __________ (12) para evitar el deterioro de las obras. Por otra parte, se __________(13), en los talleres de grabado y litografía, cursos __________(14) las distintas técnicas de estampación. Estos talleres también se cederán periódicamente a distintos artistas contemporáneos, __________ (15) se busca que el “Territorio Miró” __________(16) un centro vivo de creación y difusión del arte a todos los __________(17).

La entrada costará 500 pesetas y las previsiones dadas a conocer ayer aspiran __________(18) que el centro acoja a unos 150.000 visitantes al año. Los responsables esperan que la institución funcione a __________(19) rendimiento a principios de la __________(20) semana, si bien el catálogo completo de las obras de la Fundación Pilar y Joan Miró no estará listo hasta dentro de dos años.
<table>
<thead>
<tr>
<th>Cloze Test Answer Sheet</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. a. cumplido</td>
</tr>
<tr>
<td>2. a. inició</td>
</tr>
<tr>
<td>3. a. encargar</td>
</tr>
<tr>
<td>4. a. hubiera servido</td>
</tr>
<tr>
<td>5. a. superar</td>
</tr>
<tr>
<td>6. a. por</td>
</tr>
<tr>
<td>7. a. voluntad</td>
</tr>
<tr>
<td>8. a. al</td>
</tr>
<tr>
<td>9. a. habría</td>
</tr>
<tr>
<td>10. a. que</td>
</tr>
<tr>
<td>11. a. pretendido</td>
</tr>
<tr>
<td>12. a. disminuido</td>
</tr>
<tr>
<td>13. a. darán</td>
</tr>
<tr>
<td>14. a. sobre</td>
</tr>
<tr>
<td>15. a. ya</td>
</tr>
<tr>
<td>16. a. será</td>
</tr>
<tr>
<td>17. a. casos</td>
</tr>
<tr>
<td>18. a. a</td>
</tr>
<tr>
<td>19. a. total</td>
</tr>
<tr>
<td>20. a. siguiente</td>
</tr>
</tbody>
</table>
Appendix C: Clitics Test

INSTRUCTIONS:
Each dialogue below is incomplete. Choose any answer that you consider to sound natural and appropriate. There may or may not be more than one correct answer.

1. Pedro: ¿Cuánto cuesta la bicicleta, está barata?
   Vendedor: Está en oferta, ________.
   *(There might be more than one right answer)*
   - □ a) puedo vender en 200
   - □ b) puedo vendérsela en 200
   - □ c) se puedo vender en 200
   - □ d) se la puedo vender en 200
   - □ e) puedo se la vender en 200

2. José: Le voy a enviar las instrucciones a Pablo.
   Alfonso: Te pido un favor, ¿________ a mí también?
   *(There might be more than one right answer)*
   - □ a) podrías me las enviar
   - □ b) me las podrías enviar
   - □ c) me podrías enviar
   - □ d) podrías enviármelas
   - □ e) podrías enviar
3. María: ¿José le trajo los libros al profesor?
Javier: Sí, _________.
(There might be more than one right answer)
- a) los se trajo
- b) trajo
- c) se trajo los
- d) se los trajo
- e) trájoselos

4. Pilar: ¡Qué buen chiste! _________.
(There might be more than one right answer)
- a) Se quiero contar a Pedro
- b) Se lo quiero contar a Pedro
- c) Quiero contárselo a Pedro
- d) Quiero se lo contar
- e) Quiero contar a Pedro

5. Lucía: ¿Quién te contó el secreto?
Mario: Patricia _________.
(There might be more than one right answer)
- a) dijomelo
- b) dijo
- c) me dijo lo
- d) lo me dijo
- e) me lo dijo
(There might be more than one right answer)

- [ ] a) quiero me la comprar
- [ ] b) me la quiero comprar
- [ ] c) quiero comprar
- [ ] d) me quiero comprarla
- [ ] e) quiero comprármela

7. Francisca: ¿Quién me puede pasar la sal, por favor?  
Andrés: Yo _______.  
(There might be more than one right answer)

- [ ] a) te la paso
- [ ] b) pásotela
- [ ] c) te paso la
- [ ] d) la te paso
- [ ] e) paso

8. Paulina: Parece que nadie se ganó el premio.  
Josefa: ¡Por el contrario! Yo _______.  
(There might be more than one right answer)

- [ ] a) me lo gané
- [ ] b) gané
- [ ] c) lo me gané
- [ ] d) me gané lo
- [ ] e) ganéme lo
9. Alondra: ¿Me puedes traer la revista "Somos" mañana?
Sebastián: No estoy seguro si _______.
(There might be more than one right answer)
- □ a) pueda traer
- □ b) te la pueda traer
- □ c) pueda te la traer
- □ d) pueda traértela
- □ e) te pueda traer

10. Pablo: Me encanta esta camisa pero es muy cara
Juana: No te preocupes, ___________.
(There might be more than one right answer)
- □ a) Te compro la
- □ b) Compro
- □ c) Te la compro
- □ d) La te compro
- □ e) Cómpriotela
## Appendix D: Data Set

<table>
<thead>
<tr>
<th>#</th>
<th>Code</th>
<th>Type</th>
<th>Level</th>
<th>(T)Inv/Ung</th>
<th>(T)DO</th>
<th>(T)C-D</th>
<th>Tendency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>00 03</td>
<td>N</td>
<td>49/47</td>
<td>0%</td>
<td>100%</td>
<td>100%</td>
<td>Native</td>
</tr>
<tr>
<td>2</td>
<td>00 04</td>
<td>N</td>
<td>50/49</td>
<td>0%</td>
<td>100%</td>
<td>100%</td>
<td>Native</td>
</tr>
<tr>
<td>3</td>
<td>00 05</td>
<td>N</td>
<td>49/49</td>
<td>0%</td>
<td>100%</td>
<td>100%</td>
<td>Native</td>
</tr>
<tr>
<td>4</td>
<td>00 06</td>
<td>N</td>
<td>46/49</td>
<td>0%</td>
<td>100%</td>
<td>100%</td>
<td>Native</td>
</tr>
<tr>
<td>5</td>
<td>00 07</td>
<td>N</td>
<td>44/49</td>
<td>0%</td>
<td>100%</td>
<td>100%</td>
<td>Native</td>
</tr>
<tr>
<td>6</td>
<td>091 0</td>
<td>N</td>
<td>49/48</td>
<td>0%</td>
<td>100%</td>
<td>100%</td>
<td>Native</td>
</tr>
<tr>
<td>7</td>
<td>33 33</td>
<td>N</td>
<td>50/50</td>
<td>0%</td>
<td>100%</td>
<td>100%</td>
<td>Native</td>
</tr>
<tr>
<td>8</td>
<td>14 32</td>
<td>L2</td>
<td>NN 50/50</td>
<td>0%</td>
<td>100%</td>
<td>100%</td>
<td>Over CD</td>
</tr>
<tr>
<td>9</td>
<td>14 33</td>
<td>L2</td>
<td>NN 47/48</td>
<td>0%</td>
<td>100%</td>
<td>100%</td>
<td>NativeLike</td>
</tr>
<tr>
<td>10</td>
<td>43 21</td>
<td>L2</td>
<td>NN 48/48</td>
<td>0%</td>
<td>100%</td>
<td>100%</td>
<td>Over CD</td>
</tr>
<tr>
<td>11</td>
<td>55 55</td>
<td>L2</td>
<td>NN 48/49</td>
<td>0%</td>
<td>100%</td>
<td>50%</td>
<td>NativeLike</td>
</tr>
<tr>
<td>12</td>
<td>00 01</td>
<td>L2</td>
<td>ADV 41/45</td>
<td>0%</td>
<td>100%</td>
<td>8.33%</td>
<td>No CD</td>
</tr>
<tr>
<td>13</td>
<td>00 08</td>
<td>L2</td>
<td>ADV 45/50</td>
<td>0%</td>
<td>100%</td>
<td>91.66%</td>
<td>Over CD</td>
</tr>
<tr>
<td>14</td>
<td>021 0</td>
<td>L2</td>
<td>ADV 39/49</td>
<td>0%</td>
<td>100%</td>
<td>91.66%</td>
<td>Over CD</td>
</tr>
<tr>
<td>15</td>
<td>040 7</td>
<td>L2</td>
<td>ADV 41/44</td>
<td>0%</td>
<td>100%</td>
<td>100%</td>
<td>Over CD</td>
</tr>
<tr>
<td>16</td>
<td>040 8</td>
<td>L2</td>
<td>ADV 42/49</td>
<td>0%</td>
<td>100%</td>
<td>8.33%</td>
<td>No CD</td>
</tr>
<tr>
<td>17</td>
<td>061 6</td>
<td>L2</td>
<td>ADV 41/50</td>
<td>8.33%</td>
<td>91.66%</td>
<td>91.66%</td>
<td>Over CD</td>
</tr>
<tr>
<td>18</td>
<td>44 44</td>
<td>L2</td>
<td>ADV 43/50</td>
<td>0%</td>
<td>100%</td>
<td>100%</td>
<td>NativeLike</td>
</tr>
<tr>
<td>19</td>
<td>99 34</td>
<td>L2</td>
<td>ADV 41/47</td>
<td>8.33%</td>
<td>100%</td>
<td>0%</td>
<td>No CD</td>
</tr>
<tr>
<td>20</td>
<td>080 9</td>
<td>L2</td>
<td>INT 28/42</td>
<td>16.66%</td>
<td>16.66%</td>
<td>0%</td>
<td>No CD</td>
</tr>
<tr>
<td>21</td>
<td>100 7</td>
<td>L2</td>
<td>INT 37/47</td>
<td>0%</td>
<td>100%</td>
<td>75%</td>
<td>Over CD</td>
</tr>
<tr>
<td>22</td>
<td>14 65</td>
<td>L2</td>
<td>INT 29/41</td>
<td>16.66%</td>
<td>83.33%</td>
<td>0%</td>
<td>No CD</td>
</tr>
<tr>
<td>23</td>
<td>42 39</td>
<td>L2</td>
<td>INT 37/47</td>
<td>0%</td>
<td>100%</td>
<td>66.66%</td>
<td>Over CD</td>
</tr>
<tr>
<td>24</td>
<td>47 54</td>
<td>L2</td>
<td>INT 37/42</td>
<td>0%</td>
<td>100%</td>
<td>25%</td>
<td>No CD</td>
</tr>
<tr>
<td>25</td>
<td>99 17</td>
<td>L2</td>
<td>INT 35/40</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
<td>No CD</td>
</tr>
<tr>
<td>#</td>
<td>(T)Inv/Ung</td>
<td>(T)CAN+</td>
<td>(T)FF/CLEFT</td>
<td>(T)CLLD</td>
<td>(F)Inv/Ung</td>
<td>(F)CAN+</td>
<td>(F)FF/CLEFT</td>
</tr>
<tr>
<td>----</td>
<td>------------</td>
<td>---------</td>
<td>-------------</td>
<td>---------</td>
<td>------------</td>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>1</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>2</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>3</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>4</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>100%</td>
<td>16.66%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>5</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>6</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>7</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>8</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>9</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
<td>0%</td>
<td>50% / 0%</td>
</tr>
<tr>
<td>10</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>11</td>
<td>0%</td>
<td>0%</td>
<td>50% / 0%</td>
<td>50%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>12</td>
<td>0%</td>
<td>91.66%</td>
<td>0%</td>
<td>8.33%</td>
<td>16.66%</td>
<td>66.66%</td>
<td>16.66% + 0%</td>
</tr>
<tr>
<td>13</td>
<td>0%</td>
<td>0%</td>
<td>8.33% / 0%</td>
<td>91.66%</td>
<td>0%</td>
<td>0%</td>
<td>0% / 16.66</td>
</tr>
<tr>
<td>14</td>
<td>0%</td>
<td>0%</td>
<td>8.33% / 0%</td>
<td>91.66%</td>
<td>16.66%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>15</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>100%</td>
<td>16.66%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>16</td>
<td>0%</td>
<td>8.33%</td>
<td>83.33% / 0</td>
<td>8.33%</td>
<td>0%</td>
<td>0%</td>
<td>100% / 0%</td>
</tr>
<tr>
<td>17</td>
<td>0%</td>
<td>8.33%</td>
<td>0%</td>
<td>91.66%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>18</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
<td>0%</td>
<td>0% / 83.33%</td>
</tr>
<tr>
<td>19</td>
<td>16.66%</td>
<td>58.33%</td>
<td>25% / 0%</td>
<td>0%</td>
<td>16.66%</td>
<td>83.33%</td>
<td>0%</td>
</tr>
<tr>
<td>20</td>
<td>16.66%</td>
<td>33.33%</td>
<td>50% / 0%</td>
<td>0%</td>
<td>83.33%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>21</td>
<td>0%</td>
<td>0%</td>
<td>25% / 0%</td>
<td>75%</td>
<td>0%</td>
<td>0%</td>
<td>33.33% / 0%</td>
</tr>
<tr>
<td>22</td>
<td>16.66%</td>
<td>0%</td>
<td>83.33% / 0%</td>
<td>0%</td>
<td>16.66%</td>
<td>16.66%</td>
<td>50% / 0%</td>
</tr>
<tr>
<td>23</td>
<td>0%</td>
<td>8.33%</td>
<td>25% / 0%</td>
<td>66.66%</td>
<td>0%</td>
<td>0%</td>
<td>16.66% / 0%</td>
</tr>
<tr>
<td>24</td>
<td>0%</td>
<td>0%</td>
<td>75% / 0%</td>
<td>25%</td>
<td>16.66%</td>
<td>16.66%</td>
<td>33.33% / 0%</td>
</tr>
<tr>
<td>25</td>
<td>0%</td>
<td>0%</td>
<td>100% / 0%</td>
<td>0%</td>
<td>16.66%</td>
<td>0%</td>
<td>83.33% / 0%</td>
</tr>
<tr>
<td>#</td>
<td>(E)Inv/Ung</td>
<td>(E)CAN+</td>
<td>(E)FF/CLEFT</td>
<td>(E)CLLD</td>
<td>(SS)Inv/Ung</td>
<td>(SS)CAN+</td>
<td>(SS)FF/CLEFT</td>
</tr>
<tr>
<td>----</td>
<td>------------</td>
<td>---------</td>
<td>-------------</td>
<td>---------</td>
<td>-------------</td>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>1</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>2</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>3</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>4</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>5</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>6</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>7</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>8</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>9</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>10</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>11</td>
<td>0%</td>
<td>0%</td>
<td>66.66% / 0%</td>
<td>33.33%</td>
<td>0%</td>
<td>0%</td>
<td>33.33% / 0%</td>
</tr>
<tr>
<td>12</td>
<td>0%</td>
<td>83.33%</td>
<td>0%</td>
<td>16.66%</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
</tr>
<tr>
<td>13</td>
<td>0%</td>
<td>0%</td>
<td>16.66% / 0%</td>
<td>83.33%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>14</td>
<td>0%</td>
<td>0%</td>
<td>16.66% / 0%</td>
<td>100%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>15</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>16</td>
<td>0%</td>
<td>0%</td>
<td>83.33% / 0%</td>
<td>16.60%</td>
<td>0%</td>
<td>16.66%</td>
<td>83.33% / 0%</td>
</tr>
<tr>
<td>17</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>100%</td>
<td>16.66%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>18</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>19</td>
<td>16.66%</td>
<td>50%</td>
<td>33.33% / 0%</td>
<td>0%</td>
<td>16.66%</td>
<td>66.66%</td>
<td>16.66% / 0%</td>
</tr>
<tr>
<td>20</td>
<td>33.33%</td>
<td>33.33%</td>
<td>33.33%</td>
<td>0%</td>
<td>0%</td>
<td>33.33%</td>
<td>66.66% / 0%</td>
</tr>
<tr>
<td>21</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
<td>0%</td>
<td>50% / 0%</td>
</tr>
<tr>
<td>22</td>
<td>16.66%</td>
<td>0%</td>
<td>83.33% / 0%</td>
<td>0%</td>
<td>16.66%</td>
<td>0%</td>
<td>83.33% / 0%</td>
</tr>
<tr>
<td>23</td>
<td>0%</td>
<td>16.66%</td>
<td>16.66% / 0%</td>
<td>66.66%</td>
<td>0%</td>
<td>0%</td>
<td>33.33% / 0%</td>
</tr>
<tr>
<td>24</td>
<td>0%</td>
<td>0%</td>
<td>83.33% / 0%</td>
<td>16.66%</td>
<td>0%</td>
<td>0%</td>
<td>66.66% / 0%</td>
</tr>
<tr>
<td>25</td>
<td>0%</td>
<td>0%</td>
<td>100% / 0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>100% / 0%</td>
</tr>
</tbody>
</table>
REFERENCES


