SECOND LANGUAGE SPANISH INTONATION:
SYSTEMIC AND REALIZATIONAL DIMENSIONS OF ITS ACQUISITION

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

By

Jorge Méndez Seijas, M.S.

Washington, DC
October 3, 2019
SECOND LANGUAGE SPANISH INTONATION:
SYSTEMIC AND REALIZATIONAL DIMENSIONS OF ITS ACQUISITION

Jorge Méndez Seijas, M.S.

Thesis Advisor: Alfonso Morales Front, Ph.D.

ABSTRACT

There is increasing interest in the acquisition of second language (L2) intonation (e.g., Estebas-Vilaplana, 2017; Graham & Post, 2018), particularly in Spanish (e.g., McKinnon, 2017; van Maastricht, 2018; Yuan, González-Fuerte, Baills, & Prieto, 2019; Zárate-Sández, 2018). Some of the previous research on L2 Spanish (e.g., Henriksen, Geeslin, & Willis, 2010; Thornberry, 2014; Trimble, 2013) has sought to determine which parameters (e.g., peak alignment) are most likely to be acquired under what conditions (e.g., short vs. long stays abroad). What is still unknown is what makes these specific parameters relatively easier or more difficult to acquire. In this dissertation, I have attempted to resolve this conundrum by couching all analyses and interpretations within the L2 Intonation Learning Theory (LILt: Mennen, 2015), a model specifically designed to analyze intonation data. This theoretical approach makes this dissertation the first formal analysis of L2 Spanish intonation.

A total of 82 Spanish learners of various proficiency levels participated in one of two studies that explored the acquisition of intonation in two different contexts: classroom (beginning, n =18; intermediate, n =21; and advanced, n =24) and study abroad (advanced, n =19). I examined the acquisition of prenuclear F0 peak alignment, the phonetic mapping of final boundary tones, and pitch range phenomena. The results indicate that the LILt is a useful and
valuable theoretical tool to analyze L2 intonation data. In my discussion, I detail how the theoretical analyses that this model allows have helped to clarify, at least to some extent, why some of the parameters under investigation (e.g., the mapping of boundary tones) appear to be more difficult to acquire than others (e.g., prenuclear peak alignment).
ACKNOWLEDGMENTS

First and foremost, I want to thank the members of my committee, as well as all the other professors in the Spanish and Linguistics departments at Georgetown University. I feel great admiration and appreciation for all they taught me through the years. Alfonso and Cristina in particular have made me a better researcher, a better student and, most importantly, a much better person. I hope to always have them in my life, as colleagues and as friends. They showed me how important it is to have professors that you look up to, professors that you trust to be there for you whenever you may need them. And Germán, always generous with his time, has also been an inspiration for me, even before I met him. I knew of his talent and excellent work as a researcher, and now I am fortunate to also know what a fantastic professor and person he is. I will never be able to thank them all enough for how rewarding, challenging, interesting, and fulfilling they made my doctoral experience.

I want to thank my wonderful family, both those that I was born into as well as those who chose me later in life. They worked tirelessly to make sure that I had everything they could not have. I also want to thank my friends, particularly Elsa Mora, for always believing in me; Tim and Janire, for making linguistics and work so much fun; and Fray, for being the best of friends. I hope, now that I have finished this dissertation, to have more time to show them how much I love them and how grateful I am to them.

Some people deserve special recognition: My mom, for supporting me unconditionally and for sacrificing so much so I could study whatever I wanted. Andrés Eloy and María Angélica, for being such wonderful human beings, and for teaching me to always be generous and positive. Anita, for being the best mother-in-law I could have dreamed of. And LeAnne, for
always being there to help me achieve my goals, and for giving me Maia. All of you make it
worth it.
Esta tesis está dedicada a mi abuela Gladys.
TABLE OF CONTENTS

Chapter 1: Introduction.......................................................................................................................... 1
  Intonation and its Relevance in Second Language Acquisition.......................................................... 1
  Models of Intonation............................................................................................................................ 2
  Research on Second Language Intonation.......................................................................................... 3
  Variables under Investigation .............................................................................................................. 7
  Statement of the Problem..................................................................................................................... 8
  Useful Terminology............................................................................................................................. 11

Chapter 2: Review of the Literature...................................................................................................... 12
  Overview............................................................................................................................................... 12
  The Linguistic Structure of Intonation: The Autosegmental-Metrical Model........................................ 12
    Timing of Implementation: Alignment ................................................................................................ 16
    Levels of Implementation: Scaling and Pitch Range ........................................................................ 17
  Spanish Intonation ............................................................................................................................... 21
    Broad-Focus Declaratives in Spanish and English: A Cross-Linguistic Comparison ...................... 24
    A Transcription of Spanish Broad-Focus Declaratives .................................................................. 26
  Second Language Spanish Intonation ................................................................................................. 30
    Second Language Spanish Intonation in a Classroom Context ....................................................... 31
    Second Language Spanish Intonation in Short-Term Study Abroad Programs ............................... 32
    Second Language Spanish Intonation in Long-Term Study Abroad Programs ............................... 35
  Pitch Range in Second Language Spanish ......................................................................................... 36
  A Framework of Phonological Acquisition: The Speech Learning Model ........................................ 37
  The L2 Intonation Learning Theory (LILt) .......................................................................................... 41
  Summary and Conclusions .................................................................................................................. 49
  Research Questions ............................................................................................................................. 53
    Study 1. Second Language Phonological Acquisition in Classroom-based Context .................... 53
    Study 2. Second Language Phonological Acquisition in Study Abroad ........................................ 54

Chapter 3. Classroom Context ............................................................................................................. 55
  Method ................................................................................................................................................ 55
  Participants .......................................................................................................................................... 55
  Recruitment ......................................................................................................................................... 56
  Materials ............................................................................................................................................ 57
Chapter 4. Study Abroad Context

Method .......................................................................................................................... 88
Participants .................................................................................................................... 89
Recruitment .................................................................................................................... 89
Materials ......................................................................................................................... 89
Language Contact Profile ............................................................................................... 90
Chapter 5. Discussion and Conclusions ................................................................. 102
  Overview ........................................................................................................... 102
  Discussion ......................................................................................................... 102
  Peak Alignment ................................................................................................. 102
  Final Boundary Tone ........................................................................................ 112
  Pitch Range ....................................................................................................... 120
  Limitations and Future Research ...................................................................... 127
  Conclusions ........................................................................................................ 129

Study 1. Second Language Phonological Acquisition in Classroom-based Context .... 129
Study 2. Second Language Phonological Acquisition in Study Abroad .................... 130

Appendix A: Background Questionnaire .............................................................. 132
Appendix B: Target Declarative Sentences in Spanish (Peak Alignment) .................. 135
Appendix C: Target Declarative Sentences in Spanish (Final Boundary Tones and Pitch Range) ........................................................................................................ 136
LIST OF FIGURES

Figure 1. Representation of monotonal phonological units: a) (L)ow and b) (H)igh. ...................... 13
Figure 2. Representation of bitonal pitch accents: a) H+L*, and b) L*+H. ................................ 14
Figure 3. Presentation of two different sentence types, as produced by one participant .......... 15
Figure 4. Representation of two different instantiations of rising pitch accents in Spanish........ 17
Figure 5. Representation of downstepping in two intonational phrases. ................................. 18
Figure 6. Representation of a circumflex contour in which the H* tone shows differences in terms of scaling......................................................... 19
Figure 7. Representation of two languages that differ in terms of pitch range......................... 20
Figure 8. Representation of two languages that differ locally in terms of pitch range.............. 21
Figure 9. Inventory of pitch accents and boundary tones obtained from the dialects of three cities, as reported by Estebas-Vilaplana and Prieto (2008). ................................................................. 23
Figure 10. Presentation of three different sections of interest within an IP: prenuclear position, nuclear position, and phrase-final level. ................................................................. 26
Figure 11. Representation of the most typical pitch accents in prenuclear and nuclear position. 27
Figure 12. Pitch accents accounted for in English and Spanish. ............................................ 43
Figure 13. Presentation of two different sentence types, as produced by one participant.......... 45
Figure 14. Representation of two different tonal sequences.................................................... 46
Figure 15. Representation of two different rising tonal contours........................................... 47
Figure 16. Representation of the alignment continuum of the peak associated with the stressed syllable. ................................................................................................................. 47
Figure 17. Presentation of the phonological aspects under study in this dissertation............... 49
Figure 18. Presentation with two different regions of interest within an IP: prenuclear position and rightmost edge................................................................. 51
Figure 19. Presentation of all the target declarative sentences used for peak alignment......... 58
Figure 20. Presentation of all the target declarative sentences used for final boundary tone and pitch range ................................................................. 59

Figure 21. Presentation of all the target declarative sentences in English .................................................. 60

Figure 22. Presentation of an analyzed and coded sentence .......................................................... 64

Figure 23. Presentation of an analyzed and coded sentence to study F0 peak alignment .......... 65

Figure 24. Presentation of a stylized sentence ................................................................................ 67

Figure 25. Representation of the linguistic measurement of pitch span for Initial Prominent Peak Minus Lows (H*i-L) ......................................................... 68

Figure 26. Representation of the linguistic measurement of pitch span for Initial Prominent High Minus Final Low (H*i-FL) .................................................. 69

Figure 27. Representation of the linguistic measurement of pitch span for Noninitial Highs Minus Lows (H*-L) ................................................................. 69

Figure 28. Representation of the linguistic measurement of pitch span for Noninitial Highs Minus Final Low (H-FL) .......................................................... 70

Figure 29. Representation of the linguistic measurement of pitch span for Initial High Minus Noninitial Highs (H*i-H) ............................................................... 71

Figure 30. Representation of the results of F0 peak alignment behavior by group ................. 103

Figure 31. A cross-linguistic analysis of English and Spanish intonation at prenuclear position ......................................................................................... 109

Figure 32. Cross-linguistic analysis of English and Spanish intonation at phrase-final position ........................................................................................................ 116

Figure 33. Direction of the difference in pitch range between English and Spanish ........... 121

Figure 34. Difference in pitch range between L1 English and L2 Spanish ................................. 122
LIST OF TABLES

Table 1. Summary of the methods utilized for the classroom context study ........................................ 74

Table 2. Descriptive statistics for participants’ F0 peak alignment on the reading task, in classroom context ................................................................................................................................. 75

Table 3. Descriptive statistics for participants’ peak alignment on the semi-spontaneous task, classroom context ................................................................................................................................. 76

Table 4. Descriptive statistics for participants’ sentence-final boundary tones, in classroom context .................................................................................................................................................. 78

Table 5. The descriptive statistics for participants’ pitch range (Span measures) in Spanish and English, in classroom context ............................................................................................................. 79

Table 6. The descriptive statistics for participants’ pitch range (Level measures) in Spanish and English, in classroom context ............................................................................................................. 82

Table 7. Summary of results of pitch span and pitch level measures ..................................................... 86

Table 8. Summary of the methods utilized for the study abroad study ................................................. 93

Table 9. Descriptive statistics for peak range alignment in study abroad (pre-test and post-test) 94

Table 10. Descriptive statistics for final boundary tones in study abroad (pre-test and post-test) ...................................................................................................................................................... 95

Table 11. Descriptive statistics for pitch range (Span measures) in the study abroad (pre-test and post-test) .............................................................................................................................................. 96

Table 12. Descriptive statistics for pitch range (Level measures) in the study abroad (pre-test and post-test) .............................................................................................................................................. 99

Table 13. Final level “paradigmatic” tone specifications derived from a set of syntagmatic features (from Dilley & Breen, 2018, with a minor correction) ................................................................. 118
Chapter 1: Introduction

Intonation and its Relevance in Second Language Acquisition

The construction of meaning in a language results from processes occurring in parallel across all linguistic components (morphological, semantic, syntactic, and so forth), and even within a single component such as phonology. For instance, in order for a sentence to be produced and for it to communicate an intended and specific meaning, there are processes triggered at the segmental level and processes that occur, at the same time, beyond the segment. The phenomena beyond the segment are referred to as “prosodic” or “suprasegmental”, and they include, among other phenomena, stress, tone, and intonation. Stress and tone affect meaning at the word level, whereas intonation affects meaning at the sentence level.

Research on second language (L2) prosody, and on L2 intonation more specifically, was relatively limited until the 1980s (Gut, 2009), and Spanish was no exception (for a review see Henriksen, 2013); however, this tendency has changed as more publications within this field have emerged in the last few years (e.g., McKinnon, 2017; Graham & Post, 2018; Van Maastricht, Krahmer, Swerts, & Prieto, 2018). Within L2 prosody, intonation is a significant component, as it has the potential to negatively impact comprehensibility as well as foreign accentedness (Munro & Derwing, 1995; Ulbrich & Mennen, 2016). Mennen (2006, p. 1), for instance, suggests that poor intonational skills in an L2 may have a “devastating effect” for communication and be “unpleasant” for all participants in a linguistic exchange.

Broadly, intonation refers to patterned fluctuations of fundamental frequency (henceforth: F0). Underlying these fluctuations, there is a linguistic structure and linguistic operations that regulate the interpolation of phonological units and how they are mapped phonetically. These units are called tonal targets (Pierrehumbert, 1980). The observable physical movements of F0
(e.g., rises and falls) in an utterance result from transitions between these tonal targets. A rise, for instance, occurs when a low (L) target is followed by a high (H) target. Through mechanisms of organization and implementation of tonal targets, intonation conveys information that can be semantic or pragmatic in nature. That is, intonation has a semantic function inasmuch as it signals sentence types, such as declarative or interrogative, but also a pragmatic function, because it can be used to express a plethora of connotations, such as politeness (Astruc, Vanrell, & Prieto, 2016), bias and presupposition (Henriksen, Armstrong, & García-Amaya, 2016), and so forth. The difficulty in teasing apart the multiple functionalities of intonational contours is one of the reasons why its study is so challenging.

**Models of Intonation**

The study of intonation has advanced greatly in the last three decades: theoretical models have been developed, transcription systems that facilitate inter- and intra-linguistic comparisons are available, systematic experimental approaches have been devised, and technological advances now allow for fine-grained laboratory analyses of phonetic detail. These innovations have paved the way for comprehensive descriptions of the intonation of typologically different languages. First language (L1) and L2 acquisition research on intonational phonology has also gained traction. This interest in the acquisition of intonation is evidenced by an increasing number of journal publications, dissertations, summer schools, and conferences exclusively dedicated to this topic.

Acquisition research, particularly in second language acquisition (SLA), has added several layers of complexity to the study of intonational phonology. It is no longer enough to describe the underlying phonological architecture of intonation without considering other factors, such as transfer from the L1, universal constraints, quantity and quality of the input, length of
stay in study abroad (SA) contexts, aptitude and motivation, or the relationship between perception and production, among many other social, cognitive, or theoretical factors.

Notwithstanding the increasing interest in L2 prosody and L2 intonation, no “fully-fledged and well-grounded” model of prosodic acquisition (Simonet, 2012, p. 741) had been developed until recently. Some models of phonological acquisition have been available for many years and have been widely used, as are the cases of the Speech Learning Model (SLM: Flege, 1995) or the Perceptual Assimilation Model for L2 (PAM-L2: Best & Tyler, 2007). Although these approaches are better suited to explain segmental phenomena, some of their generalizations and hypotheses may be applied to intonation (Mennen, 2015).

The lack of an appropriate formal model has resulted in very few theoretical attempts to analyze L2 tonal and intonational data. Among the exceptions, for instance, Teague (2011) couched her dissertation within the SLM, and tested category merger of Mandarin and English tones. Within the PAM-L2, some studies have examined tonal properties (So & Best, 2010, 2011), but no attempts have been made to use this model for the acquisition of intonation. To further our knowledge of L2 intonational phonology, at least two goals must be achieved: first, there needs to be a significant amount of research conducted on both the L1 and the L2, and second, there needs to be a theoretical model that allows for systemic and systematic examinations of intonational development in an L2.

**Research on Second Language Intonation**

The first goal, a significant amount of research, has been achieved in the cases of Spanish and English, at least to some extent. Interest in the acquisition of L2 Spanish intonation by L1 English speakers has started to percolate into the field of Spanish SLA. The studies conducted thus far can be thematically organized into four groups: first, six longitudinal studies that dealt
mostly with the L2 acquisition of tonal configurations associated with broad-focus declaratives and various question types (Craft, 2015; Henriksen, Geeslin, & Willis, 2010; Méndez Seijas, 2018; Trimble, 2013a, 2003b; Thornberry, 2014). Second, four classroom-based studies that used broad-focus declaratives to investigate the acquisition of tonal alignment in prenuclear position and the phonetic implementation of final boundary tones (Zárate-Sández, 2015, 2016, 2018a, 2018b). Third, two studies that investigated focalization, one for a population of heritage language (HL) speakers of Spanish (Kim, 2016) and one for Spanish learners whose first language was Dutch (van Maastricht, 2018). And fourth, two studies (McKinnon, 2017; Yuan, González-Fuerte, Baills, & Prieto, 2019) that explored instructional effects on L2 Spanish intonation.

All these studies have advanced our knowledge of L2 Spanish intonation by collecting and analyzing data through different experimental designs (cross-sectional and longitudinal data collection), by including different learner populations (L2 learners of various proficiency levels and HL speakers), and by analyzing development in different contexts (abroad, in classrooms, in the laboratory, and at home); however, none of these studies has framed its results or analyses within a model specifically designed for the acquisition of L2 intonation. Therefore, they have been able to report development and even attainment, or lack thereof, of the variables under investigation, but have not offered consistent theoretical interpretations of the results. Doing so would help determine not only if participants acquired particular aspects of L2 intonation, but perhaps more importantly, why they did or did not do so.

As it pertains to using a model exclusively designed to study L2 intonation, which is the second goal required to further our knowledge in this field, there is still much to be done. A model of intonation can provide a rationale to make theoretically informed predictions and
assessments of what learners have achieved or may be able to achieve at different stages of development. Furthermore, it would help to more accurately explain why some intonational parameters appear to be more difficult to acquire than others, why certain developmental patterns are likely, but others may not be, etc. (Mennen, 2015). A theoretical model could also provide clear tools and more solid arguments to help make more thorough cross-linguistic comparisons, thus allowing researchers to better determine the degree to which different factors (e.g., similarities between the L1 and the L2, frequency) affect intonational development in an L2.

Models of Intonational Phonology: The Autosegmental-Metrical Model and the L2

Intonation Learning Theory

As I pursued the second goal, that is, using a theoretical model to analyze L2 intonation data in Spanish, I assumed for this dissertation that the phonological structure of intonation is as proposed by Pierrehumbert (1980), and Beckman and Pierrehumbert (1986), works rooted in metrical phonology (Liberman, 1975) and in autosegmental phonology (Goldsmith, 1979; Leben, 1973). Ladd (2008) coined the term “Autosegmental-Metrical” (AM) to refer to this framework and to any theoretical accounts couched within it (see Dilley & Breen, 2018 for a revised version of the AM model; see Prieto, 2003 for a review of other models; see Roessig, Mücke, & Grice, 2019 for a Dynamical Systems approach; see Xu, Lee, Prom-on, & Liu, 2015 for the Parallel Encoding and Target Approximation model). The AM model, accompanied by the notation system Tones and Break Indices (ToBI), to which all my transcriptions have ascribed, has been widely used to describe Mainstream American English (MAE_ToBI: Beckman & Ayers-Elam 1997; Beckman & Hirschberg, 1994, Beckman, Hirschberg & Shattuck-Hufnagel, 2005) and Spanish (Sp-ToBI: Beckman, Diaz-Campos, McGory, & Morgan, 2002; Estebas-Vilaplana & Prieto, 2008; Face & Prieto, 2007; Hualde, 2003; Hualde & Prieto, 2015; Prieto & Roseano,
2010; Sosa, 1999), among many other languages (for a review of other Romance languages, see Frota & Prieto, 2015; for many other typologically different languages, see Jun, 2005, 2014). Although researchers using ToBI have made language-specific adjustments, all general and added features share a set of universal characteristics that seek to create homogeneity. There has even been a call to make ToBI a more standardized labeling system that could be more easily used cross-linguistically, that is, an International Prosodic Alphabet (Hualde & Prieto, 2016). The fact that the same framework and very similar notation systems can be used for different languages, both typologically similar and dissimilar, facilitates more reliable and exhaustive comparisons within and across languages, both in L1 and L2 intonation research.

As I have argued, L2 intonation research must rely on a model that enables theoretically based cross-linguistic comparisons and provides a formal apparatus to describe, compare, and analyze L2 intonation data. The L2 Intonation Learning Theory (LILt: Mennen, 2015), developed within the AM model, is a first step towards achieving these objectives. To create a basis for cross-linguistic comparisons, intonational data within the LILt are analyzed within four different dimensions, and these dimensions are to be used when determining what constitutes a relatively easy or difficult parameter to acquire. Based on Ladd’s (2008) dimensions, Mennen (2015) proposes:

i) a systemic dimension (the inventory and distribution of tonal targets in a language),

ii) a realizational dimension (how tonal targets are implemented phonetically),

iii) a semantic dimension (what meaning each tonal configuration triggers), and

iv) a frequency dimension (how frequently specific tones or tonal combinations are used in a language).
These dimensions provide a theoretical foundation upon which to base comparisons. This foundation in turn constitutes a major change in how we can conduct research on the development of L2 intonation.

**Variables under Investigation**

For this dissertation, I have chosen the most frequently studied sentence type in L2 Spanish intonation research: broad-focus declaratives. This choice has been based, among other factors, on its unmarked nature and on the fact that this structure exhibits some cross-dialectal homogeneity. That is, although Spanish intonation is characterized by great geolectal variation (for an interactive atlas of intonation, see Prieto & Roseano, 2010-2013), broad-focus declaratives present rather similar intonational patterns regardless of the dialect. This is an important asset, particularly for experimental studies like the ones included in this dissertation, which rely on data collected from participants whose Spanish instructors are from different Spanish-speaking regions. Another reason to select broad-focus declaratives is that there is abundant literature about them (for a review of Spanish intonation, see Hualde & Prieto, 2015), both pertaining to L1 grammars as well as L2 Spanish intonation. Finally, given some structural similarities and dissimilarities between broad-focus declaratives in Spanish and English (see Estebas-Vilaplana, 2008), this sentence type constitutes an ideal tool to investigate the relative ease or difficulty the acquisition of both similar and dissimilar features entails for L1 English speakers learning Spanish.

Most experimental studies conducted thus far have been mainly focused on two aspects of broad-focus declaratives: i) the alignment properties in the realization of prenuclear tonal targets, and ii) the scaling properties in the realization of boundary tones at phrase-final position. I have included these same aspects in my dissertation. Tonal target inventories, as well as their
distribution and implementation, have helped me explore deviances in the systemic dimension, and how these deviances may affect the semantic dimension of intonation (phonology-semantics relation). I have also explored how the systemic and semantic dimensions interact with the realizational dimension (particularly, the phonology-phonetics relation). I included a third object of study, a parameter seldom explored before in L2 Spanish intonation research: pitch range.

Pitch range has been defined as the distance between “some topline and some bottomline” (Patterson, 2000, p.12) in a speaker’s tonal space. Unlike previously studied aspects in L2 Spanish intonation research, all of which operate locally at specific positions within an intonational phrase, pitch range operates globally. It has been noted that pitch range exhibits cross-linguistic differences. For instance, British English has a wider pitch range than German (Mennen, Schaeffler, & Docherty, 2012), and American English a narrower pitch range than Japanese (Hanley, Snidecor, & Ringel, 1966). Like any other intonational parameter, pitch range may be acquired in an L2. Some researchers have explored the difference between English and Spanish in terms of pitch range. Hanley et al. (1966), for instance, found that Spanish presented a wider pitch range, a fact that has been indirectly confirmed by other studies (e.g., Trimble, 2013a). Another relevant justification for including pitch range is that it functions differently from the other two variables (prenuclear peak alignment and final boundary tones) under study in this dissertation: pitch range is arguably a purely realizational (phonetic) property of intonation.

**Statement of the Problem**

Research on Spanish intonation conducted thus far has provided illuminating and valuable descriptions of intonational systems at different stages of development, particularly by native English speakers. We now have some empirical evidence, from various studies, that certain intonational features appear to be more difficult to acquire than others, and hence are
developed, if at all, only when an L2 learner has reached a high or even very high proficiency level. We also know that other internal and external factors (e.g. transfer, universal constraints, quantity and quality of the input) play a role, so it is imperative that we have a solid understanding of these factors so as to better understand L2 linguistic development. This knowledge we have accrued, however, only provides evidence of which parameters or aspects of intonation are likely to be developed under what conditions (e.g., after having completed a certain number of language courses or after spending certain amount of time abroad), and at what proficiency level. To further understand the results of previous studies, and to move forward in the field of intonational phonology, the next step must be to determine what it is about specific intonational parameters that makes them relatively easier or more difficult to acquire.

This dissertation is a first attempt at solving this conundrum for L2 Spanish intonation. To do so, I collected intonational data and couched the analyses within the dimensions put forth by the LILt. This theory proposes a classification of intonational features that allows researchers to better gauge the relative weight each feature may have on different linguistic dimensions (i.e., systemic, realizational, semantic, and frequency). This classification also helps to determine if features are likely to be relatively more or less difficult to acquire in different scenarios: if they belong to a specific dimension, if affecting more than one dimension at a time entails more difficulty (and if so, what permutations), and so forth. To begin exploring this theoretical model and to assess its applicability to the acquisition of L2 Spanish intonation, I have attempted to tease apart the impact features belonging to some of the aforementioned linguistic dimensions may have on L2 Spanish learners’ (L1 English) development in two contexts: classroom and study abroad. Including both learning contexts has also allowed me to draw more robust and clean comparisons with previous research on L2 Spanish intonation.
For this dissertation, I collected data from 100 Spanish L2 learners (L1 English), who were distributed as follows: 80 participants took part in the classroom-based investigation, and 20 participants took part in the study abroad investigation. I collected data from broad-focus declaratives and used them to analyze prenuclear peak alignment and the phonetic implementation of tonal targets at phrase-final position. I also explored the relationship between proficiency level and intonational development. I was particularly interested in determining what features (from what dimension or dimensions) developed first, and if there was any discernible sequencing in the acquisition of features that could be explained using the LILt, along with a compatible phonological theory of transfer such as the SLM.

Furthermore, I studied pitch range phenomena for female advanced learners in both classroom-based and study abroad contexts. Pitch range had only been studied for L2 Spanish in Kern (1955) and Méndez Seijas (2018), but those studies were only exploratory accounts. Given our limited knowledge pertaining to the cross-linguistic differences in pitch range between English and Spanish, I also collected L1 English samples of broad-focus declaratives from the 21 female participants in the advanced classroom-based group. By doing this, I was able to make L1 English-L2 Spanish comparisons in terms of pitch range, and hence evaluate its acquisition by L2 learners of Spanish.

Although the parameters and learner populations that I included in this dissertation have been studied before (Craft, 2015; Henriksen et al., 2010; Méndez Seijas, 2018; Trimble, 2013; Thornberry, 2014; Zárate-Sández, 2015, 2016, 2018a, 2018b), results from previous studies were not theoretically-grounded, and thus offered a partial, mostly phonetic perspective of L2 intonational phenomena in Spanish. The theoretical assessment I have added provides a
complementary view that hopefully enriches the existing literature on this topic, and in doing so, continues to move the field of L2 Spanish intonation forward.

**Useful Terminology**

- **Alignment**: “Temporal implementation of fundamental frequency ($f_0$) movements with respect to the segmental string” (Prieto, 2011).
- **Boundary tone**: An abstract phonological unit aligned with the beginning or end of an intonational phrase.
- **Intonational phrase**: “The part of an utterance over which a particular intonation pattern extends” (Ladefoged, 2001, p. 273).
- **Nuclear position**: The most prominent position in an intonational phrase (IP). In English and Spanish, the nuclear position is the last stressed syllable in an IP.
- **Pitch**: The auditory property of a sound that enables a listener to place it on a scale going from low to high, without considering the acoustic properties, such as the frequency of the sound.
- **Pitch accent**: An abstract intonational unit. Pitch accents anchor to stressed syllables.
- **Pitch range**: “Some topline and some bottomline” (Patterson, 2000, p.12) in a speaker’s tonal space.
- **Prenuclear position**: All stressed syllable positions before the nuclear position.
- **Tonal level**: The relative height at which pitch accents are phonetically realized.
- **Tonal span**: The tonal distance between lows and highs in an intonational contour.
Chapter 2: Review of the Literature

Overview

In this chapter, I first provide a basic and brief explanation of the Autosegmental-Metrical (AM) approach to intonation (Beckman & Pierrehumbert, 1986; Ladd, 1999; Pierrehumbert, 1980). This explanation includes aspects that are particularly important to understand intonational phenomena, both in native grammars as well as in SLA: tonal timing and levels of tonal implementation. This theoretical presentation is followed by a review of the literature on Spanish intonation, along with a description of broad-focus declaratives. After that, I provide an overview of the literature on L2 Spanish intonation research conducted thus far. I then explore some theories of phonological acquisition that have been used in Spanish L2 research, with a special focus on the Speech Learning Model (SLM: Flege, 1995). Finally, I introduce the L2 Intonation Learning Theory (LILt: Mennen, 2015), a model designed exclusively to analyze L2 intonation data. After a section with summary and conclusions resulting from the review of the literature, I introduce the research questions that guided me in this dissertation.

The Linguistic Structure of Intonation: The Autosegmental-Metrical Model

Until the 1970s, intonational phonology as we currently understand it did not exist. Approximations to the study of intonation were “instrumental” or “impressionistic”, and could not be described as phonological, but rather as “phonetic” and “proto-phonological” (Ladd, 2008, p. 10). For an approach to be phonological, Ladd notes, it ought to distinguish between two levels, one phonetic and one phonological. At the phonological level, there are categorically distinct and autonomous entities whose organization and functioning are regulated by linguistic rules. At the phonetic level, there is a linguistic sound continuum best described in terms of
physical parameters. The Autosegmental-Metrical (AM) model, within these terms, is a model of intonational phonology because it clearly distinguishes the phonological level from the phonetic level. It does so by separating the relatively continuous fluctuations of F0 (phonetic level) from discrete intonational units (phonological level). These intonational units at the phonological level are called “tonal targets”, and they are represented by two tones: A H(igh) tonal target and a L(ow) tonal target. Despite being autonomous, these intonational units are connected to the segments insomuch as their interpolation in the phonological level results in the F0 movements that shape intonational contours in the phonetic level. Therefore, a rising contour would be the by-product of a sequence of L and H tonal targets, whereas a valley may result from either a sequence of L tones or by de-accentuation of part of an intonation phrase (IP). Figure 1 depicts the tones L and H and how they are phonetically implemented:

![Figure 1. Representation of monotonous phonological units: a) (L)ow and b) (H)igh.](image)

The shaded section represents a syllable carrying lexical stress, while the white region represents an unstressed syllable. The horizontal line represents the acoustic correlate (F0 trajectory) each of these tones draws when implemented phonetically.

Tonal targets in a language are not only monotonous (H, L), as bitonal units (e.g., L+H) and even larger units (e.g., LHL) have been proposed. These singletons or tone groupings constitute “pitch accents” or “boundary tones”, which are respectively linked to stressed syllables and phrasal edges via association rules. Following ToBI, pitch accents and boundary tones are annotated differently: Pitch accents are associated with a stressed syllable and are
represented with a starred tone symbol (T*), whereas boundary tones are represented by a tone followed by a percentage sign (T%). Whenever bitonal or larger elements are described, the starred element is the one more strongly associated with the stressed syllable, whereas the other element is described as leading it (e.g., a low tone leading a high tone: L+H*) or trailing it (e.g., a high tone trailing a low tone: L*+H) (Figure 2).

![Figure 2](image)

*Figure 2. Representation of bitonal pitch accents: a) H+L*, and b) L*+H.*

The star symbol indicates stronger association with the stressed syllable.

Intonation can by itself be responsible for the conveyance of the intended semantic or pragmatic meaning of an utterance. In many Romance languages, for instance, the most common mechanism to distinguish between yes/no questions and broad-focus declaratives is by contrasting intonational units, rather than via morphology (e.g., a particle) or syntax (e.g., word order). Figure 3 illustrates how intonation helps trigger these sentence type distinctions: the unmarked and most common ending for yes/no questions (a), on the one hand, is a falling-rising contour in nuclear position (the last stressed syllable of an IP), which results of the underlying bitonal pitch accent L*+H, followed by an H boundary tone: L*+H H% (Prieto & Roseano, 2010-2013). Broad-focus declaratives (b), on the other hand, tend to exhibit no tonal excursion or a low pitch accent followed by an L boundary tone: L* L%. Given that in most cases in Spanish, the syntax of yes/no questions and broad-focus declarative sentences is the same, intonation is thus the most important marker of sentence type. In Figure 3 (below),
representations of both a prototypical yes/no question (a) and of a broad-focus declarative sentence (b) are given in order to show graphically what this tonal difference looks like.

Figure 3. Presentation of two different sentence types, as produced by one participant. The image above (a) is a yes/no question sentence, while the image below (b) is a broad-focus declarative sentence. The blue square is provided to facilitate comparisons between the final F0 movements in these two sentences.

The number of pitch accents and boundary tones that comprise intonational inventories varies from one language to another, and even from one dialect to another. Inventories may range from one pitch accent, as in Japanese (Beckman & Pierrehumbert, 1986), to much larger
inventories as those of English (Beckman & Pierrehumbert, 1986; Pierrehumbert, 1980) or Spanish (Beckman, et al., 2002; Estebas-Vilaplana & Prieto, 2008; Hualde & Prieto, 2015). Although relatively limited in terms of number of elements, the nature of tonal inventories is further complicated when tonal targets undergo language-specific or even dialect-specific implementation rules. That is, when each individual tonal target is mapped phonetically, speakers must carry out precise timing (the realization of tonal targets in time and their alignment with respect to stressed syllables or phrasal edges) and scaling (realization of tonal targets in space, within the confines of a speaker’s F0 minima and F0 maxima) specifications.

**Timing of Implementation: Alignment.** The timing of a pitch accent refers to what is called “alignment”. Pitch accents are anchored to strong metrical positions (stressed syllables). To signal this association, pitch accents usually show some acoustic correlate on or around these stressed syllables. Some rising contours in Spanish constitute a perfect example of this association. It has been reported (Prieto & Torreira, 2007; Prieto, van Santen, & Hirschberg, 1995) that for rising contours in prenuclear position in broad-focus declaratives, the trailing H tone is often delayed with respect to the stressed syllable, that is, the F0 peak is not realized within the stressed syllable, but rather in a posttonic position (in some transcriptions, the delayed peak is represented by the symbol “greater than”: >) (see Figure 4a for a representation of a delayed peak). When a rising contour appears at nuclear position, however, the peak is instead aligned with the stressed syllable, that is, it occurs within the confines of the stressed syllable. Figure 4b, which also represents a low-high sequence similar to that presented in 4a, shows different alignment configurations. These two examples illustrate how the same tonal sequence can result in different realizations depending on how the F0 peaks align with stressed syllables or with subsequent unstressed syllables.
Figure 4. Representation of two different instantiations of rising pitch accents in Spanish. In a) the rise begins within the stressed syllable but peaks in posttonic position, whereas in b) the rise occurs at full within the stressed syllable.

Levels of Implementation: Scaling and Pitch Range. The level of implementation of a tone refers to the relative F0 height at which tonal targets and boundary tones are realized and perceived by speakers. The AM model is a phonological model that bases its analyses on the notion of levels (L and H) rather than configurations (e.g., rising contour, falling contour). In other words, the AM model theorizes that speakers distinguish between pitch levels by controlling or perceiving categorical differences in pitch height, thus creating tonal contrasts (interpolation of tones that are perceptually low or high) and configurations that have specific semantic or pragmatic readings in their languages or linguistic communities. Pitch variations that occur within one level are referred to as variations in “scaling”.

The F0 level within intonational units tends to gradually decrease, particularly in declarative utterances, with each subsequent H and L tone occurring at a lower F0 range than the ones preceding them. The gradual decrease in F0 is called “downstepping”, a phenomenon that has been reported in many languages (e.g., Liberman & Pierrehumbert, 1984, for English; van den Berg, Gussenhoven & Rietveld, 1992, for Dutch) including Spanish (Prieto, Shih & Nibert, 1996). Phrase-final changes in pitch level also have a demarcative function, as speakers reset their pitch to a higher range after they have finished an IP, from which it once again decreases until the following phrase’s end (Figure 5).
This figure illustrates how the level of F0 is reset to indicate that an IP has ended, and another has begun. This operation also serves syntactic purposes, as IPs delimit syntactic constituents.

Besides demarcating intonational phrases, the level at which pitch is realized may change within intonational phrases, either globally or at specific locations. In most dialects of Spanish (and in most studied languages, for that matter), scaling variations within levels do not trigger sentence distinctions. Rather, variations in terms of scaling within the domains of a tonal category are deemed gradient, hence phonetic, and have thus been associated with emphasis and prominence (Bolinger, 1986; Pierrehumbert, 1980). However, research in some Romance languages has revealed that changes in scaling may elicit linguistic distinctions (Savino & Grice, 2011; Vanrell, 2006, 2011). Vanrell (2006), for instance, conducted a study to determine what intonational cues were being used in Majorcan Catalan to distinguish between two seemingly identical contours that were found in two different sentence types. In this case, an H leading tone aligned with the pretonic syllable and an L* tone associated with the last stressed syllable defined yes/no questions and wh-questions respectively. Vanrell’s conclusion was that the leading H tone in these sentence types is somehow different in terms of scaling, and this difference triggers a sentence type distinction. Venezuelan Andean Spanish exhibits a similar phenomenon. In this variety, broad-focus declaratives and yes/no questions present the same phrase-final intonational contour: a circumflex tonal configuration (Méndez Seijas, 2010). This circumflex configuration results from a sequence of L-H* L% tones. To differentiate these two sentence types, speakers rely on scaling differences. In other words, they realize the high tone
even higher when signaling interrogativity (see Figure 6). Some researchers (Méndez Seijas, 2010; Mora, 1996) have even suggested that the difference between broad-focus declaratives and yes/no questions in Venezuelan Andean Spanish might be achieved through realizing a global wider pitch range across the sentence for interrogatives and a comparatively narrower global pitch range for broad-focus declaratives. If this scenario proved to be correct, it would mean that scaling differences have phonologized in this variety of Spanish.

Figure 6. Representation of a circumflex contour in which the H* tone shows differences in terms of scaling.

The solid line represents the final contour of a declarative sentence in Venezuelan Andean Spanish, whereas the dotted line represents the final contour of a yes/no question in the same dialect.

How high or how low tones are mapped with respect to other tones in an IP, a characteristic referred to as “pitch range”, constitutes a feature of intonation worth exploring as well. Patterson (2000) defines pitch range as the distance between “some topline and some bottomline” (p.12) from the F0 utilized by a speaker. It has been reported that pitch range is a language-specific feature (e.g., Chen, Gussenhoven, & Rietveld, 2004). To illustrate one way in which languages differ in terms of pitch range, Figure 7 shows two hypothetical languages, one represented by a curved dotted line and another represented by a curved solid line. In this example, the language represented by a solid line has an overall wider pitch range. When speakers of this language realize H tonal targets, they do so at a higher F0 range. Therefore, the distance between those H tones and the L tones is larger than in the language represented by the
dotted line. The distance between Hs and Ls is referred to as “tonal span” and, as illustrated here, is also subject to cross-linguistic differentiation.

![Diagram of two languages with solid and dotted lines]

Figure 7. Representation of two languages that differ in terms of pitch range.

The language represented with the solid curved line has an overall wider pitch range than the language represented by the dotted curved line. The tonal level of the H tones is higher in the language represented by the solid line. The tonal span in the language represented by the dotted line is smaller (adapted from Mennen et al., 2014). However, cross-linguistic differences in terms of pitch range may be much more complicated than originally thought. Mennen et al. (2012) and Mennen et al. (2014), for instance, have proposed that cross-linguistic differences can occur locally. That is, it is possible that two languages present the same pitch range at certain locations within an intonational phrase but not within others. In Figure 8, I present a graphic representation of this type of contrast:

---

Figure 8. Representation of two languages that differ locally in terms of pitch range. The language represented with the curved solid line has a wider pitch range at the beginning of the intonational phrase, but a narrower pitch range in the end (adapted from Mennen et al., 2014).

Naturally, parameters like F0 peak alignment, scaling, and pitch range play a role not only in linguistic systems in L1 acquisition and native speakers’ developed linguistic systems. L2 speakers must also acquire these parameters, which may in some respects be similar to or different from their L1, at a phonological or phonetic level, or even at both levels.

Spanish Intonation

Spanish is, along with English, one of the most widely described languages in terms of intonation. Studies on Spanish intonation span back to the 1940s (Navarro Tomás, 1944), and include general descriptions (Hualde & Prieto, 2015) as well as detailed dialectal accounts (Sosa, 1999; Prieto & Roseano, 2010), many of which are even available online (e.g., Interactive Atlas of Spanish Intonation: Prieto & Roseano, 2010-2013). Extensive research has been devoted to specific phenomena such as wide vs. narrow focus (Face, 2001, 2002), peak alignment (Prieto & Torreira, 2007; Prieto et al., 1995), phrasing (Frota, D’Imperio, Elordieta, Prieto, & Vigário, 2007; Prieto, 2007), and so forth. Also, researchers have increasingly started to explore developmental data, both in L1 acquisition (Armstrong, Barrachina, Esteve-Gilbert, & Prieto,
2016; Lleó, 2016; Prieto & Esteve-Gilbert, 2018) as well as in L2 acquisition (Henriksen et al., 2010; McKinnon, 2017; Méndez Seijas, 2018; Trimble, 2013; van Maastricht, 2018; Zárate-Sánchez, 2015, 2016, 2018a, 2018b). Moreover, the study of intonational phenomena related to language contact has been gaining traction, with analyses of Spanish in contact with English (Alvord, 2006), Basque (Elordieta & Irurtzun, 2016), and Quechua (Muntendam & Torreira, 2016), among other languages.

The two most frequent sentence types analyzed in the literature are broad-focus declaratives and yes/no questions, so much so that ambitious projects such as the Atlas Multimedia Romanesque Space Prosody have investigated them exclusively, describing patterns of these two sentence types in many dialectal areas in Venezuela, Chile, Argentina, Costa Rica, Bolivia, Cuba, and Spain (Martínez-Celdrás & Fernández-Planas, 2003-2015). In Figure 9, I present an inventory adapted from Estebas-Vilaplana and Prieto (2008), who obtained it from their analysis and descriptions of the dialects of three cities: Madrid, Seville, and Mexico City. An exhaustive inventory of all the pitch accents and boundary tones that exist in Spanish and all its dialects would undoubtedly be larger than the one in Figure 9. However, this figure serves as an example of the many elements that may comprise a tonal system.
### Pitch accents

- **monotonal:**
  - L*

- **bitonal:**
  - L+H
  - L+H
  - H+L
  - L+H

### Boundary tones

- **monotonal:**
  - L%
  - L%
  - H%
  - M%
  - M%
  - H%
  - H%

- **bitonal:**
  - H%H
  - L%H
  - L%H
  - H%M
  - H%M

- **triton:**
  - L%H

---

*Figure 9.* Inventory of pitch accents and boundary tones obtained from the dialects of three cities, as reported by Estebas-Vilaplana and Prieto (2008).

The cities used in this study were Madrid, Seville, and Mexico City.

Tonal inventories, like segmental inventories, exhibit variability. Therefore, it is likely that some of these tonal targets in Figure 9 are used in some dialects but not in others. This should not come as a surprise. As an official language in more than 20 countries on both sides of the Atlantic, commonly coexisting with other languages, and with over 430 million native
speakers, it is safe to say that variation may be recognized as one of the defining characteristics of the Spanish language, both from a segmental as well as a suprasegmental perspective. In many cases, the languages (indigenous or European) that Spanish is or has been in contact with are undisputable catalysts of this variation, yet other factors play a role as well (e.g., language change). Despite this variation, there are features considered to be so widespread that they can arguably be used to describe Spanish intonation in general terms. Given that the target sentences of this dissertation are broad-focus declaratives, the next sections describe the most general features that characterize this particular sentence type in Spanish. In these next sections, I also provide comparisons with English broad-focus declaratives, when relevant.

**Broad-Focus Declaratives in Spanish and English: A Cross-Linguistic Comparison.**

There are some agreed-upon features of broad-focus declaratives that set Spanish apart, in some respects but not in others, from English. Below, I present a comparison between Spanish and English. This presentation of similarities and dissimilarities is crucial to further explore the likely difficulties of L2 Spanish acquisition by L1 English speakers. For clarity purposes, I reproduce these same characteristics in Figure 10, showing specifically where within an intonation phrase these similarities and dissimilarities take place.

Similarities:

i) Both Spanish and English display a downstepped F0 trajectory from the beginning to the end of intonational phrases (Ladd, 1990, 1993, for English; Prieto et al., 1996 for Spanish).

ii) In Spanish and English, phrasal accents tend to be aligned with the rightmost content word of an intonational phrase (Pierrehumbert, 1980 for English; Sosa, 1999 for Spanish).
iii) The final boundary tone of both Spanish and English is oftentimes the same, L% (see Hualde, 2003; Sosa, 1999 for Spanish; Pierrehumbert, 1980 for English).

iv) In nuclear position, there is some consensus that both Spanish and English have an underlying L* (Pierrehumbert, 1980 for English; Sosa, 1999 for Spanish;). However, there is some disagreement in the Spanish transcriptions that have been proposed.

Dissimilarities:

i) In Spanish, unlike English, most stressed words carry a pitch accent (Hualde & Prieto, 2015).

ii) In prenuclear position, Spanish declaratives present a rising pitch movement. This rising contour begins within the stressed syllable, and the peak is realized in posttonic position (delayed peak)\(^2\) (Garrido, Lliisterri, de la Mota & Rios, 1993; Prieto et al.,1995). In English, there is usually peak alignment (Pierrehumbert, 1980). That is, the peak is aligned with the stressed syllable, at least to a greater degree than in Spanish. The exact location of the F0 peak may vary due to various factors such as the length of the word or its stress position (Estebas-Vilaplana, 2007); the position of the word within the intonational phrase; speech rate or distance (in number of syllables) between stressed syllables (Silverman & Pierrehumbert, 1990); dialectal variation (Arvaniti & Garding, 2007), etc.

iii) Although both English and Spanish use the same boundary tone for phrase-final intonation, L%, it has been widely reported in English that final rises in declaratives are possible (Armstrong, Piccinini, & Ritchardt, 2015; Bolinger, 1998), a phenomenon that

\(^2\) Delayed peaks do not occur in all dialects (e.g., Argentine Spanish: Colantoni, 2011).
has been called “High Rising Terminal” (HRT), more popularly referred to as “uptalk” (Warren, 2016). This phenomenon has been, albeit scantly, reported for Spanish as well (Vergara, 2017).

\[
\begin{array}{lll}
\text{Broad-focus declarative sentences} \\
\text{A comparison between Spanish and English} \\
\hline
\text{Prenuclear position} & \text{Nuclear position} & \text{Phrase-final level} \\
\text{(pitch accent)} & \text{(pitch accent)} & \text{(boundary tone)} \\
\text{Dissimilar} & \text{Similar} & \text{Similar} \\
\text{(Delayed peak in Spanish, aligned peaks in English)} & \text{(Both use an L*)} & \text{(Both Spanish and English use an underlying L%)} \\
\hline
\text{Globally:} \\
\text{Similar: downstepping, phrasal accent is the rightmost stressed syllable in an IP} \\
\text{Dissimilar: in Spanish, unlike in English, most stressed words carry a pitch accent} \\
\end{array}
\]

Figure 10. Presentation of three different sections of interest within an IP: prenuclear position, nuclear position, and phrase-final level.

In the table, I indicate whether these sections are similar or dissimilar in Spanish and English.

**A Transcription of Spanish Broad-Focus Declaratives.** The transcriptions and interpretations of the most common pitch accents in prenuclear and nuclear position in broad-focus declaratives, which are depicted in Figure 11a and 11b respectively, may seem straightforward. However, their phonological notations are much more complicated than they seem at first glance, oftentimes resulting in conflicting opinions on which symbols should be used. In this section, I review the different transcriptions that have been put forth for pitch accents in these two positions.
Figure 11. Representation of the most typical pitch accents in prenuclear and nuclear position. Figure 11a shows a prenuclear position. 11b shows a nuclear position. In 11a, there is a delayed peak. In 11b, there is a low valley.

Prenuclear pitch accents, whose most common realization is a rise with a delayed peak, (11a) have been described as monotonal (Prieto et al., 1995) and as bitonal (see Beckman, Díaz-Campos et al., 2002; Estebas-Vilaplana & Prieto, 2008; Hualde & Prieto, 2015; Prieto & Roseano, 2010). Nuclear accents have also been described in more than one way: L*, *, and H* (Hualde & Prieto, 2015; Estebas-Vilaplana, 2008). The reason why multiple symbols have been used is that, in intonation research, it is sometimes very difficult to distinguish between a) the behavior of F0 movements resulting from specific underlying pitch accents and, b) the scaling or pitch range variations caused by language-specific tune-text mappings or by expressive functions of intonation. As will be seen in the discussion that follows, understanding the reasoning and implications behind the selection of these symbols will also help us explore how these different interpretations and transcriptions may affect cross-linguistic comparisons.

The first proposed notation for prenuclear accents in Spanish was a monotonal pitch target, H* (Prieto et al., 1995), a decision that can be explained by the fact that the tonal implementation of this H* tone is often perceived as high (Face & Prieto, 2007). This interpretation has also been recently reported in L2 Spanish perception by L1 English learners (Zárate-Sández, 2015). Interpreting it and transcribing it this way would mean that, at least in prenuclear position, Spanish uses the same prenuclear accent as that proposed for English
(Pierrehumbert, 1980). However, the observed intonational contour in prenuclear position in these two languages is different. This phonetic difference between English and Spanish was first attributed to language-specific implementation rules: in English, the peak is frequently aligned with the stressed syllable, whereas in Spanish it is normally delayed, that is, it reaches its highest F0 point in posttonic position. This phonological analysis, however, has been contested (Face, 2001; Prieto, 1995), and the phonetic difference observed in the F0 trajectory is now mostly attributed to two distinct underlying pitch accents: the tonal movement in English results from an underlying H* pitch accent whereas in Spanish there is an underlying bitonal L+H pitch accent.

The specifics of the bitonal underlying pitch accent in Spanish have been described in various ways as well: Estebas-Vilaplana (2007) proposed that it was an L* followed by an H word-edge tone. In a different description, the delayed F0 peak has been represented as L* followed by an H that was realized postonically: L*→H (Estebas-Vilaplana, 2008). In one of the most recent revisions of Spanish ToBI (Prieto & Roseano, 2010), however, yet another symbol was proposed: L+<H*, indicating that the element most strongly associated with the stressed syllable is H, not L, thus explaining the perception as a high tone that had been reported in Face and Prieto (2007).

The notation and interpretation of F0 peak alignment in prenuclear position in Spanish is a particularly relevant factor in examining stress and the acquisition of Spanish intonation by L1 English speakers. There are various acoustic correlates to stress in Spanish (pitch, duration, and intensity), but not all of them are always present (Ortega-Llebaria & Prieto, 2011). Pitch is one correlate of stress that does contribute to the perception of prominence (Hualde, 2012). Therefore, the delayed peak described for Spanish, and the fact that it differs from English, may lead to the incorrect interpretation of a lexical item, which can potentially cause communication
breakdowns. For instance, a misinterpretation of F0 peak alignment properties in Spanish may result in Spanish learners (L1 English) perceiving posttonic syllables as tonic. That is, they may hear a word like “bebe” (first syllable carrying lexical stress), meaning “she/he drinks”, but perceive “bebe” (as if the second syllable were stressed), a word meaning “baby”. This may occur because, in English, F0 peaks tend to be more closely aligned with stressed syllables.

Although it may be crucial in the interpretation of lexical stress at the word level, differences in the degree of peak alignment in English and Spanish do not carry semantic weight at the sentence or utterance level. In other words, the alignment of the prenuclear F0 peak may signal new information or a focalized syntactic constituent, but it does not by itself change the semantic content of a sentence. For instance, given a sentence like “Jaime loves Cercei”, speakers can use alignment properties to signal focalization, and thus convey different pragmatic interpretations: the original sentence, “Jaime loves Cercei”, can thus be transformed into “JAIME loves Cercei” (Jaime, not anyone else) or “Jaime LOVES Cercei”, meaning that his feelings are of love, and nothing less, or Jaime loves CERCEI (Cercei, not Brianne of Tarth or anyone else).

Nuclear pitch accents in Spanish tend to exhibit a valley followed by a phrase-final low tone. The most commonly used transcription to explain the F0 movement is L* at nuclear position, followed by an L boundary tone: L* L%. However, Hualde and Prieto (2015) point out that nuclear position in broad-focus declaratives oftentimes presents no tonal excursions whatsoever, and thus the authors suggest that it might be appropriate to have no tone symbols in the transcription. Instead, there should only be a star indicating that this is a stressed position: *. Estebas-Vilaplana (2008) has proposed that this valley before the L% boundary tone is actually not an L*, but rather an H* whose phonetic implementation seems low because of downstepping.
These disagreements in interpretation demonstrate how difficult it is to describe the intonational system in a single language, let alone compare systems cross-linguistically: some of the transcriptions proposed, for instance, would make us think that pitch accents at nuclear position in English and Spanish are different (Spanish: L* or * vs. English H*) while others would make us conclude that speakers of these two languages underlyingly have the same pitch accent in this position (H*). For transcriptions in this dissertation, I have adopted the most widely accepted L*<H for prenuclear pitch accents and L* for final boundary tones, respectively, thus facilitating comparisons with other research conducted in the area of L2 Spanish intonation.

**Second Language Spanish Intonation.** Research on L2 Spanish intonation has been gaining traction in the last few years. The studies conducted thus far can be thematically organized into four groups: first, six longitudinal studies that dealt mostly with L2 acquisition of the tonal configurations associated with broad-focus declaratives and various question types (Craft, 2015; Henriksen, Geeslin, & Willis, 2010; Méndez Seijas, 2018; Thornberry, 2014; Trimble, 2013a, 2003b). Second, classroom-based studies of broad-focus declaratives investigating the acquisition of production and perception of tonal alignment in prenuclear position and final boundary tones (Zárate-Sández, 2015, 2016, 2018a, 2018b). Third, studies that investigated focalization in Spanish as a HL (Kim, 2016) and as an L2 (van Maastricht, 2018). And fourth, two investigations that explored pedagogical interventions in the teaching of intonation (Craft, 2015; McKinnon, 2017; Yuan et al., 2019). The contributions these studies have made are significant, covering different experimental designs (cross-sectional and longitudinal), different learner populations (from all proficiency levels, as well as HL speakers), insights into different intonational features (sentence type marking, alignment, pitch range, focus), and even pedagogical interventions.
The studies that are most relevant for this dissertation can also be classified according to the context in which learning occurred: on the one hand, classroom-based research (Zárate-Sández, 2015, 2016, 2018a, 2018b), and on the other hand, studies dealing with the development of intonational phonology in study abroad programs (Craft, 2015; Henriksen et al., 2010; Méndez Seijas, 2018; Thornberry, 2014; Trimble, 2013). Despite their differences, some of the aspects under investigation in all these studies allow for comparisons to be made and very tentative conclusions to be drawn. To better understand and interpret their results in a more unified fashion, I have organized and explained these studies as follows: first, classroom-based studies; second, short-term programs ranging from 5-7 weeks (Craft, 2015; Henriksen et al., 2010; Méndez Seijas, 2018); and third, long-term programs ranging from 15-16 weeks (Trimble, 2013; Thornberry, 2014). The following section concludes with a brief overview of research on pitch range in Spanish as an L2.

**Second Language Spanish Intonation in a Classroom Context.** Zárate-Sández (2015) is the only researcher to have explored developmental data in a classroom context. In a cross-sectional design, which included 70 L2 Spanish learners from three proficiency levels (intermediate, advanced, and highly advanced), he analyzed production and perception data and compared their data to L1 English speakers, L1 Spanish speakers, and HS of Spanish. His study focused on the phonetic implementation of prenuclear accents as well as final boundary tones. His results showed that intermediate-proficiency learners’ intonational behavior tended to be similar to the L1’s (English), whereas very high-proficiency learners resembled HSs’ production data. That is, highly advanced speakers’ results were neither like the L1 nor like the L2, but rather somewhere in between.
A significant contribution of Zárate-Sández (2015) was the inclusion of an imitation task to measure perception. His results, for instance, showed that intermediate proficiency learners did not perceive delayed peaks in prenuclear position, a feature much later acquired. Given the relationship there is between perception and production, these results in perception can be used to partially explain why these same intermediate proficiency learners could not produce delayed peaks in Spanish. Unsurprisingly, highly advanced speakers who could produce these tonal patterns were also able to imitate them in the perception task. The decision to adopt an imitation task to tap into learners’ perception does not come without peril, though, as it could be argued that intermediate speakers’ perception cannot be assessed with an instrument that relies on the production of a structure that they do not yet master. Put differently, his production data may be confounding his perception data.

In terms of boundary tones, Zárate-Sández (2015) found that there were also instances of HRT, also called uptalk, particularly for participants of intermediate proficiency. This was explained in terms of transfer from the L1. As proficiency increased, the height of the F0 at phrase-final position decreased, thus indicating that there was improvement towards native-like levels. The fact that this study did not collect data from beginning learners, however, leaves a gap in our understanding of intonational development in classroom contexts. That is, the results indicate that learners of intermediate proficiency had not developed native-like implementation of final boundary tones, but that highly proficient speakers had.

*Second Language Spanish Intonation in Short-Term Study Abroad Programs.* In a 7-week short-term SA program in León, Spain, Henriksen et al. (2010) analyzed sentences belonging to three different types, namely declarative statements, yes/no questions, and wh-questions. This study was the first attempt to study L2 Spanish intonation. Its methodological
instruments, particularly a set of contextualized sentences for participants to read out loud, has continued to be used, with slight modifications, in subsequent research on L2 intonation. In this investigation, the authors sought to determine what changes occurred over time, especially in terms of prenuclear realizations and sentence-final contours (nuclear position + boundary tone). As it pertains to declarative sentences, the post-test results for their five L2 Spanish learners (L1 English) showed improvements in some respects, but also overall lack of consistency. For prenuclear contours, for instance, some participants had started to implement a delayed rise in prenuclear position by week seven, but there was great variability. In terms of sentence-final configurations, some learners were using a fixed set of intonational contours consistently, often target-like, while others increased their inventory of pitch accents and final boundary tones to the detriment of consistency. Certain patterns associated with the L1, such as final rises, did not disappear in most cases. As Henriksen (2013) indicates, a drawback of Henriksen et al. (2010) is that this study relied on the description of intonational contours without providing any specific measurements of individual pitch gestures. In other words, the results were presented in terms of configurations (e.g., rise, fall), thus not allowing for these results to be extrapolated and compared to other studies that provide more precise descriptions of specific phonetic and phonological detail.

Craft (2015), in her dissertation, studied the acquisition of some intonational features of declarative sentences and yes/no questions by L2 Spanish learners of advanced proficiency while on a 6-week stay abroad in Valencia, Spain. The participants were 8 learners (L1 English) that the author placed into two groups: an experimental group (five) and a control group (three). The experimental group received explicit instruction on prosody. Her results indicated that only the experimental group evidenced improvements. These improvements happened across the board,
with evidence of delayed prenuclear rises and frequent target-like final contours, despite also reporting many instances of uptalk. The three participants in the control group did not show any significant changes in intonation.

Finally, Méndez Seijas (2018) also used a pre-test/post-test experimental design to study intonational development in a short-term SA program, in this case a 5-week program in Barcelona, Spain. His results indicated that some improvements in both prenuclear and nuclear configurations were indeed possible, and that some aspects seemed to be more difficult to develop than others. For instance, prenuclear alignment was not consistent, and its use did not change over time. Like Henriksen et al. (2010) and Craft (2015), it was likely for participants to have rising intonation at the end of broad-focus declaratives. This study, however, used spontaneous data, which might to some extent explain why final low boundary tones were not more frequent. A relevant contribution of this study was its theoretical analysis, which constitutes a first attempt to frame results within a model of L2 intonational phonology. The results of this study were based on only 4 participants, and hence its scope was extremely limited.

These three studies offer a blurry and very inconclusive picture of SA development in a short-term program: their results are difficult to compare because they used different types of data (controlled vs. spontaneous) and based their results on incompatible theoretical approaches: Méndez Seijas (2018) framed his results within a model of intonation whereas Henriksen et al. (2010) and Craft (2015) mostly relied on the phonetic description of configurations. Furthermore, these studies based their results on data from very small participant pools (four to eight participants). More theoretically principled research and bigger participant pools are needed in short-term SA intonation research.
Second Language Spanish Intonation in Long-Term Study Abroad Programs. Trimble (2013) studied intonational development of L2 Spanish advanced learners (L1 English) in a study abroad context. In this case, a semester-long program in Mérida, a city in the Venezuelan Andes. This study was the first of its kind, providing developmental data from participants who spent a whole semester abroad. Trimble included declarative and interrogative sentences in his study, and took into account factors such as style/formality, interaction with native speakers, etc. Of particular interest was the development of interrogatives, because this variety of Spanish presents a somewhat peculiar intonation for this sentence type. Results in this investigation indicated that in the post-test, particularly in broad-focus declaratives, most learners had modified their intonational contours, adopting a preferred pattern, which resembled the target form. They used this adopted pattern consistently. It is also noteworthy that five out of the nine learners that participated in the study developed an expanded pitch range which mirrored that of the target language in this particular variety of Spanish.

The other study that investigated development in a long-term SA program was Thornberry (2014). He analyzed data from 11 L2 Spanish learners (L1 English) of intermediate proficiency during a semester abroad in Buenos Aires, Argentina, and also focused on the acquisition of broad-focus declaratives and yes/no questions. His results indicated that only two of the learners’ intonational contours approximated the target language by the end of the semester abroad, in both declarative sentences as well as in yes/no questions. The other nine speakers’ contours remained unchanged after the SA program. It must be noted, though, that the learners who were enrolled in this program reported to have had rather limited and sporadic interactions with Spanish speakers outside of their homestays, which could in part explain why only two of the nine participants evidenced clear signs of improvement.
As was the case with short-term SA programs, these two studies on long-term programs also provided somewhat contradictory results. On the one hand, Trimble (2013) obtained positive results, as most of his participants seem to have improved, particularly in terms of declarative intonation. Thornberry (2014), on the other hand, had dispiriting results. Despite the contrasting findings, two important conclusions can be drawn: first, that quantity and quality of input seem to play a major role in the development of L2 intonation in SA, and second, that some participants were able to improve so much that they were able to produce targetlike forms by the time the SA program ended.

**Pitch Range in Second Language Spanish.** In L2 intonation research, it has been shown that learners may realize their pitch accents in a globally expanded or reduced tonal space as compared to the target language. For instance, Backman (1979) studied 12 L1 Venezuelan Spanish speakers learning English and found an overall narrower pitch range in the learners’ data. The acquisition of language-appropriate pitch range in an L2, however, is more complex as differences are not always global. Mennen, et al. (2012) found that cross-linguistic differences in pitch range may also be local. That is, it is possible that two languages present the same pitch range at certain locations of an intonational phrase but not at others. Therefore, L2 learners must first perceive how the target language diverges with respect to the L1 and where within the IPs, and then accordingly adjust their pitch range either globally or locally (Mennen et al., 2014).

Until very recently, pitch range had barely been studied in Spanish. Only two studies (Hanley et al., 1966; Kelm, 1995) conducted more than 20 years ago had focused on this aspect of intonation. Hanley et al. (1966), who conducted a cross-linguistic comparison between Japanese, Latin American Spanish, and American English, found that Spanish had a wider pitch range than American English, but narrower than Japanese. Kelm (1995), for his part, focused on
the acquisition of Spanish and English pitch range by L1 English speakers and L1 Spanish speakers, respectively. The author suggested that learners, irrespective of their L1, had a narrower pitch range in their respective target language. Kelm hypothesized that this narrowing in pitch range might happen regardless of the L1 or L2. More recently, Méndez Seijas (2018) included this variable in his study but could draw no conclusions given his small sample size. Despite his limited data, the author reported that some modifications had been observed from pre-test to post-test. To better understand changes in pitch range, however, more data needed to be analyzed, both in Spanish as well as in English.

Other than the aforementioned studies, there have been only a handful of researchers who have reported any information on pitch range phenomena. Trimble (2013) and Thornberry (2014), for instance, mentioned pitch range when reporting modifications from pre-test to post-test: some of the participants in their studies increased their overall pitch range after a semester abroad, a change arguably implemented to mirror target language levels.

A Framework of Phonological Acquisition: The Speech Learning Model

To varying degrees of success, some models of phonological acquisition have been able to include aspects (e.g. transfer, universal constraints, cognitive variables, social variables, age-related effects) that are deemed very relevant in the acquisition of intonation. The Speech Learning Model (Flege, 1995) and the Perceptual Assimilation Model for L2 (PAM-L2: Best & Tyler, 2007), for instance, have integrated the notion of transfer, age of arrival (AOA), and language experience, among other factors. Unlike generative developments that have focused mostly on the internal architecture of phonological systems, these two models focus on how internal and external factors affect acquisition, particularly in cross-linguistic comparisons.
between learners’ L1s and L2s. These models have provided researchers with tools to better understand and predict areas of difficulty in L2 development.

The SLM, probably the most widely used model of phonological acquisition, treats discrete entities as perceptual phonetic categories. When a learner is exposed to an L2, there is an interaction between the sound system in the L1 and the L2. Some of the L2 sounds may be erroneously categorized as belonging to perceptually similar categories in the L1, thus slowing or even impeding the formation of a new L2 category. Some other sounds may instead be treated as new sounds, which is likely to facilitate the creation of a new category. The SLM has been used almost extensively to study the acquisition of segmental phonology (for a review of studies on vowels in L2 Spanish, for instance, see Colantoni, Steele, & Escudero, 2015) but seldomly to the acquisition of intonation (e.g., Teague, 2011). To illustrate how the SLM operates, I briefly present Spanish/English differences in terms of vowel categories (different vowel inventories). There are certain vowels in English that do not exist in the learners’ L1, but which they may perceive as “similar” to some existing vowel category in their native system. For instance, English, on the one hand, has four non-low front vowels /i, ɪ, e, ɛ/. Spanish, on the other hand, has two /i, e/. For Spanish speakers learning English, the similarity between /i, ɪ/ and /e, ɛ/ may result in equivalence classification (Flege, 1987) of these two pairs into two categories rather than four.

Based on similarities and dissimilarities like these between Spanish and English’s non-low front vowels, the SLM can be used to make predictions as to which sounds may be easier or more difficult to perceive. Sounds that are “similar” may lead to category merger, and their discreteness thus more difficult to acquire. Sounds that are unlike any L1 category are instead more easily perceived as different and may therefore more likely lead to the emergence of a new
L2 category. Intonational phenomena can also be analyzed in terms of discrete units (phonemelike tones: L and H) that form phonetic categories (e.g., rises and valleys). Their acquisition is also likely to undergo some of the same processes explored for segmental units such as vocalic or consonantal phonemes.

The SLM is, to a great extent, a model based on linguistic transfer; however, it does not state that transfer is the only explanation for all deviations in an L2, like earlier models such as Contrastive Analysis (Lado, 1957) did. Rather, it presents very specific conditions under which transfer can take place. The ability to form new linguistic categories in the SLM is not blocked after a certain age, as the more radical version of the Critical Period Hypothesis (Lennenberg, 1967) would have suggested. Instead, the SLM states that the ability to acquire native-like categories never expires. This does not mean that age is not a factor. Researchers focus on age-effects related to AOA and the fact that phonological categories strengthen with time, facts that may negatively impact the perception of new categories in the target language, thus making them more resistant to acquisition. As a basis for analysis within the SLM, Flege (1995, p. 239) proposed a series of postulates and hypotheses:

Postulates:

1) *The mechanisms and processes used in learning the L1 system, including category formation, remain intact over the life span, and can be applied to L2 learning.*

2) *Language-specific aspects of speech sounds are specified in long-term memory representations called “phonetic categories”.*

3) *Phonetic categories established in childhood for L1 sounds evolve over the life span to reflect the properties of all L1 and L2 phones identified as a realization of each category.*
4) Bilinguals strive to maintain contrasts between L1 and L2 phonetic categories, which exist in a common phonological space.

Hypotheses:

1) Sounds in the L1 and L2 are related perceptually to one another at a position-sensitive allophonic level, rather than at a more abstract phonemic level.

2) A new phonetic category can be established for an L2 sound that differs phonetically from the closest L1 sound if bilinguals discern at least some of the phonetic differences between the L1 and the L2 sounds.

3) The greater the perceived phonetic dissimilarity between an L2 sound and the closest L1 sound, the more likely it is that phonetic differences between the sounds will be discerned.

4) The likelihood of phonetic differences between the L1 and the L2 sounds, and between L2 sounds that are noncontrastive in the L1, being discerned decreases as AOL increases.

5) Category formation for an L2 sound may be blocked by the mechanism of equivalence classification. When this happens, a single phonetic category will be used to process perceptually linked L1 and L2 sounds (diaphones).

6) The phonetic category established for L2 sounds by a bilingual may differ from a monolingual's if: 1) the bilingual's category is "deflected" away from an L1 category to maintain phonetic contrast between categories in a common L1-L2 phonological space; or 2) the bilingual's representation is based on different features, or feature weights, than a monolingual's.

7) The production of a sound eventually corresponds to the properties represented in its phonetic category representation.
The SLM, however, does not make any explicit predictions with regards to prosodic development (Colantoni et al., 2015). This inability to adapt this model to the study of prosody explains why Simonet (2012, p. 741) stated that there was no “fully-fledged and well-grounded” model of prosodic acquisition. However, this has changed thanks to emergence of the L2 Intonation Learning Theory (Mennen, 2015), a framework that offers a classification of intonational parameters that allows for their study within frameworks of phonetic/phonological acquisition like the SLM.

**The L2 Intonation Learning Theory (LILt)**

The L2 Intonation Learning Theory (Mennen, 2015) is different from other models of phonological acquisition in that it is exclusively developed to analyze intonational data. This is significant given the fact that intonation is in many respects different from segmental phonology. First, unlike segments, no models have provided clear foundations on which to base cross-linguistic analyses of intonational phenomena. Second, in intonation, unlike segmental phonology, categorical distinctions are not as clear-cut (Grice, Ritter, Niemann, & Roettger, 2017; Gussenhoven, 1999). These differences between intonational and segmental phonology justify an independent model of intonation. While other phonological models of acquisition such as the SLM or the PAM-L2 offer hypotheses that can be extended to intonation, these hypotheses need to be somehow adapted so they can be used to interpret intonational data (Mennen, 2015).

To create a basis for analyses and cross-linguistic comparisons for the LILt, Mennen (2015) proposes that we analyze intonational data within four different dimensions, which can be used to explore what features are relatively easier or more difficult to acquire. Building on Ladd’s (2008) dimensions, Mennen proposes:
i) a systemic dimension (the inventory and distribution of pitch accents and boundary tones in a language),

ii) a realizational dimension (how tonal targets are implemented phonetically),

iii) a semantic dimension (what meaning each tonal configuration carries), and

iv) a frequency dimension (how frequently specific tones or tonal combinations are used in a language).

The systemic dimension refers to the inventory and distribution of pitch accents and boundary tones in a language. Given the widely attested cross-linguistic differences and similarities that have been found and described in phonological research (e.g., number and type of vowels), it is only natural to find this same variability in terms of the categorical elements that form part of an intonational system. An example of this can be seen in the figure below (Figure 12), where widely accepted pitch accent inventories for Spanish (Estebas-Vilaplana & Prieto, 2008) and English (Beckman & Pierrehumbert, 1986) are presented.
Figure 12. Pitch accents accounted for in English and Spanish.

In the red box in can be seen that, according to these authors, there are pitch accent gaps, that is, pitch accents that exist in the inventory of only one of the two tonal systems.

Inside the red box, two gaps are highlighted, showing that although these tonal inventories are very similar, there are two pitch accents that are not shared: Unlike English, the Spanish inventory has L<H*. This pitch accent occurs in prenuclear position in Spanish, a position where English uses H* instead. This systemic difference could, to some extent, explain the observable phonetic differences in peak alignment between English and Spanish. In light of this cross-linguistic dissimilarity and the theoretical underpinnings and hypotheses of the LILt and a model like the SLM, it might be possible to speculate as to what types of challenges this could entail for L2 Spanish learners whose L1 is English.

The realizational dimension refers to how pitch accents and tones are implemented phonetically. This is the dimension that has been explored the most (Mennen, 2015), probably because realizational deviations from the norm are relatively easy to observe and test experimentally. A few studies will suffice to understand what can be done to explore phenomena
in this dimension. Backman (1979), for instance, compared eight Venezuelan Spanish speakers learning English and found that their peaks were displaced leftward (alignment) and their pitch range was too narrow (pitch range). Also, Trofimovich and Baker (2006), in a study of L2 English (L1 Korean) found significant differences in terms of stress timing compared with their control group. Another study that can be catalogued here is Mennen et al. (2014), who explored the pitch range of advanced L2 German speakers (L1 English). Finally, the differences in scaling and pitch range between English and Spanish that I have described in this chapter would also constitute parameters to be studied within the realizational dimension.

The semantic dimension refers to the relation that exists between tones and the meanings they convey. That is, speakers use particular pitch accents in specific combinations at precise locations to convey specific semantic or pragmatic information (e.g., interrogativity, presupposition). Failure to realize the right element in the right place may lead to miscommunications or even to breakdowns in communication. This dimension is directly related to the systemic and the realizational dimensions because deviations in either of those dimensions may result in intonational contours that do not correspond with the message that learners are trying to express. Examples such as the difference between (a) yes/no questions and (b) broad-focus declaratives in Spanish, as reproduced in Figure 13, are useful to understand this interaction between dimensions.
Figure 13. Presentation of two different sentence types, as produced by one participant. The image above (a) is a yes/no question sentence, while the image below (b) is a broad-declarative sentence. The blue square is provided to facilitate comparisons between the final F0 movements in these two sentences.

Another example will further illustrate this is found in Figure 14 (below): two contours are shown which differ only in one tone, the boundary tone: L*H H% and L*H L%. The first contour (L*H H%) is associated with neutral yes/no questions, whereas the second contour (L*H L%) is used to express obvious contrast (Estebas-Vilaplana & Prieto, 2008). The steep fall at the
end of the contour may be caused by a systemic deviance or by a realizational deviance (and, in any case, this also causes a problem in the semantic dimension).

![Diagram](image)

**Figure 14.** Representation of two different tonal sequences.

In a) an unmarked yes/no question, and in b) a contour used to express “obvious contrast”

The frequency dimension refers to how frequently or infrequently certain pitch accents or tonal movements are used in a given language as compared to other languages that also have the same elements in their tonal inventory. As it pertains to Spanish and English, the bitonal pitch accent L*+(<)H offers a perfect example: the contour that results from this pitch accent is very frequent in Spanish, as it is the most common underlying accent in prenuclear position in broad-focus declaratives. In English, this contour is infrequent and marked (Arvaniti & Garding, 2007), so much so that in Dainora’s (2012) corpus, for instance, it only represented 0.05% of the total labeled items.

This cross-linguistic difference may lead to perception and interpretation problems. As Ortega-Llebaria, Gu and Fan (2013) reported in a perception task they conducted, stimuli of *mama* were oftentimes interpreted as *mama* by advanced L2 Spanish (L1 English) speakers. This interpretation is partly a consequence of frequency: although the Spanish learners also had the L*H pitch accent in their L1 inventory, they perceived this sequence as a high tone aligned with the last syllable of the word. This is a logical and rather expected interpretation given the very low frequency of L*H and the high frequency of H* in their L1 system.
Classifications into these four categories are not always clear. For instance, although the realizational dimension includes F0 peak alignment, this feature sometimes triggers phonological contrasts and must therefore be studied within the systemic dimension. Rising contours are an example of how complicated classifications within these dimensions might sometimes be. The rising contours that result from L+ H may be realized in various ways, showing either early rise (Figure 15a) or late rise (Figure 15b). Furthermore, as Figure 16 depicts, differences can also be observed in terms of alignment of the H tone with respect to the stressed syllable (Figures 16 and 17). These differences can be phonological in some languages and phonetic in others.

Figure 15. Representation of two different rising tonal contours.

In a) it can be observed in both cases the rise begins “early”, that is, within the stressed syllable (the gray area represents the stressed syllable). In b) the rise begins immediately after the stressed syllable, that is, it begins “late”.

Figure 16. Representation of the alignment continuum of the peak associated with the stressed syllable.

Although always a rising contour, the peak may be reached either within the stressed syllable or outside of its domain.

In terms of the theoretical assumptions that should be used to interpret differences and similarities in the four dimensions, Mennen (2015), in agreement with SLM and PAM-L2, states
that a) many of the difficulties for learners may be caused by the (mis-)perception of tonal categories, b) comparisons should not be restricted to the phonological level, as phonetic detail may be crucial to understanding intonational development, c) age of arrival may influence the degree to which a learner may acquire some aspects of intonation (though it must be determined whether all dimensions are equally affected), d) there should be a relationship between experience and performance, and e) linguistic interaction is bi-directional, that is, as both the L1 and the L2 share the same linguistic space, it is expected that both be somehow influenced by the interaction. In other words, acquisition of an L2 may have an impact on the L1 system.

Given these four dimensions and the theoretical underpinnings of the SLM and the PAM-L2, Mennen (2015) offers a framework that could facilitate cross-linguistic comparisons. The ultimate goal of this theory is to further our understanding of the relative difficulty that intonational parameters may pose for L2 learners. Having a model as well as clear dimensions upon which to base predictions and analyses makes it possible to study, compare, and contrast learners of different language backgrounds and different proficiencies.

Specifically applied to L2 Spanish intonation, and for the purposes of this dissertation, I have mostly focused on two dimensions: the systemic dimension and the realizational dimension. These two dimensions represent the phonetic and phonological aspects of intonation. I have used these two dimensions to classify the variables under study (prenuclear peak alignment, the height of final boundary tones, and pitch range phenomena). Given that the meaning of intonation is of utmost importance when studying phonological contrasts, I have made references to the semantic dimension when pertinent. The classification I have made using these dimensions has helped me determine what features in which categories are apparently more difficult to acquire. In Figure
17 below I present the main dimensions in which each variable was most likely to present deviations from target-like intonation.

<table>
<thead>
<tr>
<th>Prenuclear peak alignment</th>
<th>Final boundary tone</th>
<th>Pitch range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Systemic dimension/Realizational dimension</td>
<td>Realizational dimension</td>
<td>Realizational dimension</td>
</tr>
</tbody>
</table>

*Figure 17. Presentation of the phonological aspects under study in this dissertation.*

Below each aspect, I list the potentially affected dimension(s).

**Summary and Conclusions**

Intonation research has been generating a wide breadth of knowledge these past few years, with an increasing interest in both L1 and L2 acquisition. This interest has spread to the field of Spanish L2 acquisition, resulting in a number of publications and dissertations (Craft, 2015; Henriksen et al., 2010; Kim, 2016; Méndez Seijas, 2018; Thornberry, 2014; Trimble, 2013; van Maastricht, 2018; Zárate-Sánchez 2015, 2016, 2018a, 2018b). Researchers have looked into various aspects of intonation (e.g., tonal configuration, the effects of instruction) in different learning contexts (i.e., study abroad, classroom). Practically all of these studies on L2 Spanish intonation have zeroed in, directly or indirectly, on one particular sentence type: broad-focus declaratives. Other sentence types (e.g., yes/no questions) or aspects (e.g., focalization) have been explored as well, albeit to a lesser extent. The interest in broad-focus declaratives is not fortuitous; rather, it is based on two main factors: first, this sentence type is unmarked, hence a logical starting point for inquiry in a subfield of study that is still relatively new. Second, broad-focus declaratives in Spanish have been widely described in the literature, which has provided us with a solid understanding of both its underlying structure as well as its implementation characteristics. Because of these two reasons, the target structure in this dissertation is also
broad-focus declarative sentences. I explored broad-focus declaratives in two learning contexts (classroom and study abroad) and included various proficiency levels (low, intermediate, and advanced). By including different learning contexts and learner populations, I have been able to better interpret not only what the results of this dissertation indicate, but more importantly, I have been able to draw clearer comparisons with other research previously conducted.

Within an intonational unit such as a broad-focus declarative sentence, there are various sections of interest. In this dissertation I have further explored two main aspects: prenuclear F0 peak alignment and scaling of final boundary tones. For both of these aspects, previous research had identified particular tonal behaviors and different linguistic functions. My conclusions will have hopefully added to the knowledge gathered regarding these aspects.

As explained before, languages may vary in their phonetic and phonological dimensions. For example, the tonal configurations and realizations of Spanish intonational features coincide in some respects with English intonation. Divergences, however, are also possible. Up until now, no thorough theoretically based explorations had been made so as to determine the difficulty similar and dissimilar cross-linguistic aspects may entail in the acquisition of L2 Spanish intonation. In part, the reason why this had not been attempted was a lack of theoretical tools exclusively designed to analyze intonational phenomena. The L2 Intonation Learning Theory (Mennen, 2015) lays the groundwork for such cross-linguistic comparisons. Analyzed through the lens of the LILt, it is possible to classify cross-linguistic differences and similarities in terms of dimensions. I have couched all my analyses within the LILt because this model allows for comparisons that go far beyond descriptive analyses. In Figure 18, I show two of the regions of interest in this dissertation: prenuclear position and sentences’ rightmost edge. I then detail in
what respect English and Spanish differ in these positions. Finally, I indicate on which
dimensions deviances in L2 Spanish are more likely to occur.

<table>
<thead>
<tr>
<th>Prenuclear position (pitch accent)</th>
<th>Sentence’s rightmost edge (final boundary tone)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dissimilar at the systemic level (L&lt;\text{H}^* \text{ vs. } \text{H}^*)</td>
<td>Dissimilar at the realizational level (Zárate-Sández, 2015, reports that Spanish has lower realizations of F0)</td>
</tr>
<tr>
<td>Dissimilar at the realizational level (delayed peak vs. aligned peak)</td>
<td></td>
</tr>
<tr>
<td>Systemic dimension</td>
<td>Realizational dimension</td>
</tr>
</tbody>
</table>

*Figure 18.* Presentation with two different regions of interest within an IP: prenuclear position and rightmost edge.

In the table, I indicate: i) whether these sections are similar or dissimilar in Spanish and English, and ii) the dimension(s) where deviances are more likely to occur.

The comparison between Spanish and English tonal configurations in prenuclear position and in phrase-final position presented in Figure 19 can serve as an example of how dimensions can change our cross-linguistic analyses and interpretations of intonational phenomena. If we analyze data from L2 Spanish learners (L1 English) without these dimensions, we may conclude that both prenuclear accents and phrase-final configurations in Spanish and English are dissimilar. This conclusion is not necessarily erroneous. If configurations in these two positions are different, we may then expect that their tonal configurations likely be as easy or as difficult to
acquire. However, research on L1 English and L2 Spanish learners conducted thus far has indicated that this is not the case. The theoretical grounding offered by the LILt may provide further arguments as to why we should consider that although tonal configurations in some positions are indeed superficially different, as is the case I just presented for Spanish and English, tonal configurations in each of these positions are not necessarily functionally analogous. That is, in light of the dimensions proposed by the LILt, prenuclear pitch accents and final boundary tones in Spanish can no longer be treated as equivalent, because one of them (prenuclear pitch accents) affects an extra dimension: the systemic dimension. Without these dimensions, we can only provide a somewhat superficial, phonetic analysis of the data. These dimensions offer a new understanding of the various ways in which “difference” can be deconstructed, which may begin to explain why for L2 Spanish learners (L1 English) one aspect (scaling of the boundary tone) seems to be more resistant to acquisition than another (prenuclear pitch alignment).

To further understand the L2 acquisition of intonation, I have added another variable to this dissertation: pitch range. The acquisition of pitch range has been studied before (Mennen et al., 2014) in L2 English. In this study, Mennen and colleagues explored moderate to advanced L2 English learners (L1 German). Although neither age of arrival nor time of residence were controlled for in their study, the authors were able to determine that some learners approximated their productions to language-appropriate pitch ranges. While references to pitch range changes have been made in the case of L1 English speakers learning Spanish (Kelm, 1995; Haley et al., 1966; Méndez Seijas, 2018; Thornberry, 2015; Trimble, 2013), no studies before this dissertation had included a theoretical analysis based on a model created exclusively to analyze intonation. Besides being an innovation within the field of L2 Spanish intonation, studying pitch range
within the LIL has allowed me to explore a parameter of intonation that clearly affects only one dimension of those proposed by Mennen (2015): the realizational dimension. No reliable data pertaining to pitch range is available for either American English or Spanish. Therefore, to be able to make cross-linguistic comparisons, I have included a comparison between L1 English and L2 Spanish.

I used production data for all the experiments of this dissertation, both from a classroom-based context as well as from a short-term study abroad program. Including these two contexts helped me explore intonational development both cross-sectionally and longitudinally. The target structure was broad-focus declaratives, and within this sentence type, I investigated the acquisition of L2 Spanish intonation in terms of 1) prenuclear F0 peak alignment, 2) the scaling of the final boundary tones, and 3) pitch range. I based most of my experimental design on previous research (Henriksen et al., 2010; Mennen et al., 2012, 214; Thornberry, 2014; Trimble, 2013; Zárate-Sández, 2015) so that connections could be easily made with results from other studies.

To analyze all the variables mentioned above, I have divided this dissertation into two independent studies: a classroom-based investigation (Chapter 3) and a study abroad investigation (Chapter 4). The research questions that guided me in these studies were the following:

**Research Questions**

**Study 1. Second Language Phonological Acquisition in Classroom-based Context.**

1. Prenuclear pitch alignment: Does the realization of Spanish prenuclear peak alignment change as L2 learners advance in proficiency?
2. Sentence-final boundary tones: Does the scaling of final boundary tones change as L2 learners advance in proficiency?

3. Pitch range: Do advanced L2 learners of Spanish evidence differences in pitch range between L2 Spanish and L1 English?

**Study 2. Second Language Phonological Acquisition in Study Abroad.**

1. Prenuclear pitch alignment: Do advanced L2 learners of Spanish evidence changes in the realization of prenuclear peak alignment after a short-term 5-week study abroad program?

2. Sentence-final boundary tones: Do advanced L2 learners of Spanish evidence changes in the realization of final boundary tones after a short-term 5-week study abroad program?

3. Pitch range: Do L2 advanced learners of Spanish evidence changes in pitch range after a short-term 5-week study abroad program?
Chapter 3. Classroom Context

Method

In this chapter, I present the methods section, the analysis section, and the results section for the classroom study. A description of the participant pool is followed by the recruitment protocol, all the different materials used to collect data to answer the research questions, and the procedure. This is followed by the coding schema for all the variables under study. In the analysis section that follows, there is detailed information about the statistical analyses that were implemented, as well as the rationale for their use. Finally, statistical analyses for all the variables under study are presented in this order: a) F0 peak alignment, b) final boundary tone, and c) pitch range.

Participants. Eighty Spanish learners volunteered to partake in the classroom-based study in this dissertation. To be retained for analysis, participants had to a) be a native speaker of English, b) be enrolled in a beginning, intermediate, or advanced Spanish course at Georgetown University, c) not be a heritage speaker of any language, and d) not have participated in a Spanish study abroad program. Furthermore, for a participant to be retained for analysis, she or he had to have at least 80% of analyzable sentences in the contextualized reading task.

The data from 17 participants was discarded for not meeting at least one of the aforementioned criteria: 4 participants were discarded because they were L1 speakers of a language other than English (Korean, n = 1; Mandarin, n = 2; Tagalog, n = 1), 4 participants were discarded because they were heritage speakers (Spanish, n = 3; Yoruba, n = 1), and 9 L1 English participants were discarded because the quality of their recordings was not high enough for segmentation and analysis. The final number of participants whose data was retained for analysis was 63. These 63 participants were divided into three groups. The placement into these...
groups was based on the class in which they were enrolled: participants enrolled in a second-semester Spanish class were placed into the beginning group \((n = 18)\), participants enrolled in a fourth-semester Spanish class were placed into the intermediate group \((n = 21)\), and participants enrolled in a sixth-semester Spanish class were placed into the advanced group \((n = 24)\).

For the study of pitch range, which only included learners from the advanced group, there was an additional criterion for inclusion: only female participants were retained for analysis. The decision to include only female participants followed previous studies on pitch range (Mennen et al., 2012; Mennen et al. 2014). The reason for this was to avoid gender effects on F0 range, a distortion that could have been particularly acute in the present study given the disparity in the distribution of gender within the advanced group (female, \(n = 21\); male, \(n = 3\)). Although one possible solution might have been to normalize the F0 values into Equivalent Rectangular Bandwidth or semitones, for instance, no previous study on pitch range has used normalized data to include both genders. Therefore, such a conversion could have compromised the analyses, comparisons, and conclusions of this dissertation.

**Recruitment.** All L2 Spanish learners enrolled in regular second semester, fourth-semester, and sixth-semester Spanish courses at Georgetown were notified of the opportunity to participate in this study via three possible routes: 1) an email sent by the director of the Spanish program, 2) an announcement made, in person, by the principal researcher in one of their Spanish classes, or 3) word of mouth. After receiving information about the study, learners were instructed to contact the principal researcher to schedule a meeting for the data collection process. Participants were offered extra credit in their Spanish classes for the exam in which they had the lowest grade. Participation in this empirical study was voluntary.
**Materials.** Following previous L2 Spanish intonation work dealing with production data (Henriksen et al., 2010; Thornberry, 2014; Trimble, 2013; Zárate-Sández, 2015), I relied on various tasks and instruments for this study. To gather demographic and linguistic information, I included a background questionnaire (Appendix A). Two tasks were used to collect intonational data: a contextualized reading task to collect controlled utterances, and an information gap task used to collect semi-spontaneous utterances. Both tasks are described in turn.

**Contextualized Reading Task.** For this task, I created two series of sentences that contained only high-frequency words, one series to study timing phenomena (prenuclear F0 peak alignment, 4 sentences total), and another series to study scaling phenomena (boundary tones and pitch range, 6 sentences total), for a total of 10 experimental sentences. All the words used in these sentences were retrieved from textbooks of basic Spanish used at Georgetown University. All 10 experimental sentences were used in a contextualized reading task like that of Henriksen et al. (2010). To facilitate phonetic analysis and coding, most words in these sentences were comprised of voiced consonantal segments. By selecting mostly voiced segments, the resulting curves of F0 drawn by the data processing software Praat (Boersma & Weenink, 2011) were almost always guaranteed to be almost continuous lines.

The four sentences used to study peak alignment (Appendix B) were constructed following these criteria: a) the target words where I would measure peak alignment were paroxytone, that is, the unmarked word type in Spanish, b) the target words constituted the first stressed word in the sentence, c) the target syllables did not contain a high vowel, d) the target syllables did not contain consonantal clusters in either the onset or the coda. The distance between the critical stressed syllables in prenuclear position and the following stressed syllable in the sentence was kept constant at two unstressed syllables. The configuration of words as well
as the distance between stressed syllables are relevant considerations when creating this type of stimuli because peak timing is affected by these factors (Prieto, 1995, 2011; Prieto et al., 1995).

The target sentences, shown below in Figure 19, were read twice by each participant.

<table>
<thead>
<tr>
<th>Target Sentences</th>
<th>Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manolo me manda flores</td>
<td>Manolo sends me flowers</td>
</tr>
<tr>
<td>Él bebe la leche</td>
<td>He drinks the milk</td>
</tr>
<tr>
<td>Mi hermana camina al trabajo</td>
<td>My sister walks to work</td>
</tr>
<tr>
<td>El pelo castaño me gusta</td>
<td>I like brown hair</td>
</tr>
</tbody>
</table>

*Figure 19. Presentation of all the target declarative sentences used for peak alignment. Sentences are presented in Spanish (left column) as well as their translation into English (right column). The target syllable is bolded and underlined.*

To study final boundary tones and pitch range phenomena, I created another six stimuli in Spanish (Appendix C). All these stimuli were 7-10 syllable broad-focus declarative sentences. The words in the sentences where all high-frequency words and were all also retrieved from the beginning Spanish textbook used at Georgetown University. To facilitate phonetic analysis and coding of these sentences, most words were formed by voiced consonantal segments. By selecting mostly voiced segments, the resulting curves of F0 drawn by the data processing software Praat were almost always continuous lines. In terms of word selection at sentence-final position, the stimuli were constructed using words that differed in terms of stress placement: one oxytone, four paroxytones and one proparoxytone. This variety in the last word of the stimuli, in my estimation, would make results more generalizable. The six sentences are presented in Figure 20 below.
<table>
<thead>
<tr>
<th>Target Sentences</th>
<th>Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>El clima era muy cálido</td>
<td>The weather was very warm</td>
</tr>
<tr>
<td>Manuela tomó una limonada</td>
<td>Manuela drank a lemonade</td>
</tr>
<tr>
<td>Fueron a ver a una amiga</td>
<td>(They) went to see a friend</td>
</tr>
<tr>
<td>Manuela leyó una novela</td>
<td>Manuela read a novel</td>
</tr>
<tr>
<td>Comieron algunas bananas</td>
<td>(They) ate some bananas</td>
</tr>
<tr>
<td>Bebieron algo en un bar</td>
<td>(They) drank something at a bar</td>
</tr>
</tbody>
</table>

*Figure 20.* Presentation of all the target declarative sentences used for final boundary tone and pitch range.

Sentences are presented in Spanish (left column) as well as their translation into English (right column).

To study the advanced learners’ pitch range phenomena, I created another 4 sentences in English (Appendix D). These sentences, which were 10-11 syllables long, were also comprised of mostly voiced segments so as to facilitate segmentation. The last words in these sentences are always paroxytone. After the pilot study, I modified the original sentences I had created. The previous sentences were too short and were thus more likely to have only one peak. To be able to make a comparison with Spanish, I needed longer sentences where more modulation was possible. As Hualde and Prieto (2015) have indicated, unlike in Spanish, not every word in English carries a pitch accent. The final English sentences that I used are presented in figure 21 below:
Target sentences in English

My brother ran eleven miles this morning

My mother made a wonderful dinner

The rain this morning flooded my bedroom

I want to buy new running shoes this summer

Figure 21. Presentation of all the target declarative sentences in English.

These sentences were used in the within-subjects comparison of pitch range.

All the sentences used to study both peak alignment and boundary tones/pitch range were included in a contextualized reading task. Participants had to read the sentences as they were presented one by one along with a “context”. This “context” provided extra information intended to help participants imagine a real situation in which the sentences could have been said. An example is presented below:

Contexto: Tus padres te preguntan si tus hermanos pequeños ya desayunaron.

Tú dices: Comieron algunas bananas

Context: Your parents ask you if your little sisters have already eaten

Participants read: They ate some bananas

The 10 target declarative sentences (4 sentences to study peak alignment, and 6 sentences to explore final boundary tones and pitch range) were interspersed along with 15 sentences of different types (yes/no questions and exclamative sentences) that served as distractors. The total
number of sentences participants read in Spanish as part of this task was 25. The advanced participants read a total of 4 sentences in English.

**Information Gap Task.** This task was used to collect semi-spontaneous data, and to compare peak alignment results in controlled and semi-spontaneous speech. It was not to be used to study final boundary tones or pitch range because semi-spontaneous data are characterized by isolated words and very short intonational phrases whose endings are oftentimes not clearly marked through intonation, but instead through vowel lengthening and pauses (Méndez Seijas, 2018). Using semi-spontaneous data could have obscured the results of final tonal boundary tones and pitch range rather than enrich them. This was a particularly important decision given that two of the groups included in this dissertation were of rather low proficiency (beginning and intermediate learners).

Although most studies in intonation research collect controlled data in laboratory settings, there has been a call for phonologists to carry out more analyses of spontaneous speech (Face, 2010), that is, naturally occurring speech. In the last decade, many researchers (e.g., Colantoni, 2011; Face, 2003; Méndez-Seijas, 2018) have conducted investigations using spontaneous speech, and it seems to be a trend that will continue to grow. Some researchers have made comparisons between intonation in controlled and spontaneous speech, and they have reported differences. Face (2003), for instance, concluded that the number of accented syllables in an utterance may be affected by speaking style. However, previous results in L2 Spanish intonation research have indicated that patterns of broad-focus declaratives have been similar in both controlled and spontaneous styles (e.g., Zárate-Sández, 2015). Zárate-Sández (2015), for instance, observed that broad-focus declaratives in his study were not affected, probably because of their unmarked nature, and also because they were short and syntactically simple.
The instrument that I used to collect semi-spontaneous data is an increasingly common tool in L2 Spanish intonation research: an information gap task based on the popular game Twenty Questions. This game has already been used in several studies (Henriksen et al., 2010; Thornberry, 2014; Trimble, 2013). Its use was originally inspired in sociolinguist interviews conducted in Simonet (2009). Twenty Questions is a guessing game. To play this game, one person has a card with a name, a city, or a thing and another person must ask questions to try to guess who or what is on the card. The person in possession of the card can only respond with “yes” or “no” answers. In my data collection sessions, however, I adopted the same modifications as other researchers (e.g., Henriksen et al., 2010; Trimble, 2013): I or my research assistant asked L2 Spanish participants to always give further information in their response. For instance, if the participant had a card with a female singer on it and they were asked whether the image of the card was of a woman, participants could then elaborate a response like this: “Yes, it is a woman, and she has a beautiful voice”. These responses were instances of declarative sentences. We played the game long enough with each participant so that they would produce at least ten to fifteen declarative sentences. Participants played both the role of the person asking questions, and the other role of responding to questions. By both asking and responding to questions, participants did not know what type of sentences I was interested in collecting.

Procedure. The data collection was conducted in one session that lasted approximately 60 minutes total. First, participants were invited to read the consent form, ask any questions they may have, and sign it if they agreed to participate. Then, they completed a language background questionnaire (10 minutes), a contextualized reading task (30 minutes), and an information gap task (20 minutes). Advanced learners only completed the contextualized reading task twice: once in Spanish and then again in English. For the contextualized reading task and the information gap
task, the L2 Spanish participants’ responses were recorded with a Shure SM58-X2U cardioid Dynamic Microphone. All data were collected by the researcher or a research assistant in a quiet office located in one of the buildings of the university.

Coding. The L2 Spanish participants’ recordings for the contextualized reading and the semi-spontaneous tasks were analyzed using Praat. The 10 target declarative sentences (4 sentences to study F0 peak alignment, and 6 sentences to explore final boundary tones and pitch range phenomena) in the contextualized reading task were identified in the audio files of each participant’s recording. Additionally, the 4 English sentences were identified for the advanced group. Then a total of 4 sentences in the semi-spontaneous task were identified. The sentences that were of the best quality were selected for each participant. This process resulted in 810 sentences. Out of these 894 sentences, 32 had to be eliminated because of poor audio quality, pervasive creaky voice, or because reading was unnatural.

For the transcription of tonal events, the coding was made following the specifications of the notation system Tones and Break Indices (ToBI: Beckman & Ayers-Elam 1997), particularly its version for Spanish (Estebas-Vilaplana & Prieto, 2008). In Figure 22, a transcription of the sentence “Manuela tomó una limonada” (“Manuela drank a lemonade”) is shown. At the top, this image shows the waveform. Then, a representation of the curve traced by the fluctuations of F0. Using Pratt, I created three independent tiers which I completed while analyzing each sentence. These tiers, as recommended by ToBI, were an orthographic tier, a break index tier, and a tone tier. The orthographic tier is a transcription of the words read or said by the participants, the break index tier signals the prosodic groupings in terms of conjoining or disjoining, and the tone tier is used to transcribe the underlying tones associated with the observable tonal movements.
Figure 22. Presentation of an analyzed and coded sentence.

Beneath the curve representing the changes in F0, three tiers can be observed: an orthographic tier, a break index tier, and a tone tier.

**Peak Alignment.** To study F0 peak alignment, I manually calculated the distance between the onset of the target stressed syllable and the peak of F0, which usually occurred within the first unstressed syllable. This distance was expressed in milliseconds. Given the controlled nature of the stimuli in this dissertation (e.g. no variation in intervening unstressed syllables between stressed syllables, no relevant differences in syllable structure) and its reduced number, I decided against normalizing the data for differences in speech rate, a consideration taken into account in other research on intonation (e.g., Dilley & Brown, 2007; Zárate-Sández, 2015). Below in Figure 23, a simplified sample sentence is analyzed for F0 peak alignment. The target syllable in this particular sentence was the second syllable in *Manolo*. At the top, this image shows the
waveform and then a representation of the curve traced by the fluctuations of F0. Then, also using Pratt, I added a fourth tier in which I recorded the distance, in ms, between the onset of the stressed syllable and the F0 peak. This procedure has been used before in other studies of intonation (e.g., Prieto, 2009). Whenever there was no clear F0 peak, but rather a plateau, I recorded the right elbow of the plateau instead. Similar procedures have been adopted in other studies on peak alignment (e.g., Thornberry, 2014)

![Figure 23](image1.png)

*Figure 23.* Presentation of an analyzed and coded sentence to study F0 peak alignment.

Beneath the curve representing the changes in F0, four tiers can be observed: an orthographic tier, a break index tier, a tone tier, and a distance tier.

**Final Boundary Tones.** For the study of final boundary tones, one data point was extracted from each target sentence in the contextualized reading task. Following the same procedure as Henricksen (2014) and Zárate-Sández (2015), the height of F0 was captured at the last non-spurious F0 point detected in the pitch track by Praat. The value was first recorded in Hertz, and then converted into Equivalent Rectangular Bandwidth (ERB) units using the formula

\[
\text{ERB} = \frac{4000 \times \text{Hz}}{1 + 7 \times \log_{10}(\text{Hz})}
\]
developed by Glasberg & Moore (1990). ERB units, instead of raw values in Hertz, have been the preferred measurement in some of the previous studies looking at pitch height phenomena in Spanish (e.g., Alvord, 2006; Ladd, 2008; Simonet, 2008; Henriksen, 2010) because they neutralize differences in scaling caused by physiological differences between male and female participants.

**Pitch Range.** Analysis and coding of pitch range was based on insights from Mennen et al. (2012), Mennen et al. (2014) and Patterson (2000). These authors compared the different techniques that have been used to measure pitch range. These measures are either linguistic (F0 values associated with specific tonal targets) or long-term distributional (LTD) (4 standard deviations above or below the mean, 95th-5th percentile, etc.). Their conclusion was that linguistic measures were far superior. I therefore followed Mennen et al. (2012) and Mennen et al. (2014), and I identified the following absolute landmark, which would be measures of pitch level:

i) Initial value (I): the first value of F0 at the very beginning of a sentence.

ii) Final value (FL): the last tonal event in a sentence. I recorded the final non-spurious point detected by Praat, regardless of whether it was in a falling or rising (uptalk) contour.

iii) Local Minima or Valleys (L): All instances of L, excluding the first and last points in a sentence. This measurement included instances of starred and unstarred Ls.

iv) Local Maxima or Peaks (H). All instances of H, excluding the first and last points in a sentence. This measurement included instances of starred and unstarred Hs.

v) High Prominent Peak (H*i). The first F0 peak in an intonational phrase.
To code for these values, I stylized every sentence using the stylization function in Praat. This procedure helped me identify all the inflection points in each of the sentences. After this procedure, each inflection point was manually labeled and F0 data were recorded for the aforementioned inflection points: (I), (L), (H*)i), (H), and (FL). In Figure 24, I present an image that illustrates this process. Approximately 50-55 data points were recorded per participant per language, for a total of 2,268 data points. Regions where there was creaky voice or interruptions in the speech flow were discarded from the analyses.

These recorded data points are used to measure participants’ pitch level at different points within an intonational phrase. I repeated the same procedure to obtain pitch range measurements in English.

![Figure 24. Presentation of a stylized sentence.](image)

Each of the green dots signals an inflection point in the curve representing the changes in F0. These inflection points are detected by the sound processing software Praat. The red arrow is pointing the first value, coded as “Initial value (I)”.
Following Mennen and colleagues (2012), the linguistic formulas used to measure pitch span were: \((H^*i-L)\), \((H^*i-FL)\), \((H^-L)\), \((H^*FL)\), and \((H^*i-H^*)\). Below, each one of them is further analyzed:

- **Initial Prominent Peaks Minus Lows** \((H^*i-L)\) explores pitch range in the first section of the intonational phrase, comparing the highest initial point to the mean Ls.

\[ (H^*i-L) \]

\[ (H^*i-FL) \]

\[ (H^*FL) \]

\[ (H^*i-H^*) \]

\[ (H^*i-L) \]

\[ (H^*i-FL) \]

\[ (H^*FL) \]

\[ (H^*i-H^*) \]

\[ (H^*i-L) \]

\[ (H^*i-FL) \]

\[ (H^*FL) \]

\[ (H^*i-H^*) \]

\[ (H^*i-L) \]

\[ (H^*i-FL) \]

\[ (H^*FL) \]

\[ (H^*i-H^*) \]

\[ (H^*i-L) \]

\[ (H^*i-FL) \]

\[ (H^*FL) \]

\[ (H^*i-H^*) \]

\[ (H^*i-L) \]

\[ (H^*i-FL) \]

\[ (H^*FL) \]

\[ (H^*i-H^*) \]

\[ (H^*i-L) \]

\[ (H^*i-FL) \]

\[ (H^*FL) \]

\[ (H^*i-H^*) \]

\[ (H^*i-L) \]

\[ (H^*i-FL) \]

\[ (H^*FL) \]

\[ (H^*i-H^*) \]

\[ (H^*i-L) \]

\[ (H^*i-FL) \]

\[ (H^*FL) \]

\[ (H^*i-H^*) \]

\[ (H^*i-L) \]

\[ (H^*i-FL) \]

\[ (H^*FL) \]

\[ (H^*i-H^*) \]

\[ (H^*i-L) \]

\[ (H^*i-FL) \]

\[ (H^*FL) \]

\[ (H^*i-H^*) \]

\[ (H^*i-L) \]

\[ (H^*i-FL) \]

\[ (H^*FL) \]

\[ (H^*i-H^*) \]

\[ (H^*i-L) \]

\[ (H^*i-FL) \]

\[ (H^*FL) \]

\[ (H^*i-H^*) \]

\[ (H^*i-L) \]

\[ (H^*i-FL) \]

\[ (H^*FL) \]

\[ (H^*i-H^*) \]

\[ (H^*i-L) \]

\[ (H^*i-FL) \]

\[ (H^*FL) \]

\[ (H^*i-H^*) \]

\[ (H^*i-L) \]

\[ (H^*i-FL) \]

\[ (H^*FL) \]

\[ (H^*i-H^*) \]

\[ (H^*i-L) \]

\[ (H^*i-FL) \]

\[ (H^*FL) \]

\[ (H^*i-H^*) \]

\[ (H^*i-L) \]

\[ (H^*i-FL) \]

\[ (H^*FL) \]

\[ (H^*i-H^*) \]

\[ (H^*i-L) \]

\[ (H^*i-FL) \]

\[ (H^*FL) \]

\[ (H^*i-H^*) \]

\[ (H^*i-L) \]

\[ (H^*i-FL) \]

\[ (H^*FL) \]

\[ (H^*i-H^*) \]

\[ (H^*i-L) \]

\[ (H^*i-FL) \]

\[ (H^*FL) \]

\[ (H^*i-H^*) \]

\[ (H^*i-L) \]

\[ (H^*i-FL) \]

\[ (H^*FL) \]

\[ (H^*i-H^*) \]

\[ (H^*i-L) \]

\[ (H^*i-FL) \]

\[ (H^*FL) \]

\[ (H^*i-H^*) \]

\[ (H^*i-L) \]

\[ (H^*i-FL) \]

\[ (H^*FL) \]

\[ (H^*i-H^*) \]

\[ (H^*i-L) \]

\[ (H^*i-FL) \]

\[ (H^*FL) \]

\[ (H^*i-H^*) \]

\[ (H^*i-L) \]

\[ (H^*i-FL) \]

\[ (H^*FL) \]

\[ (H^*i-H^*) \]

\[ (H^*i-L) \]

\[ (H^*i-FL) \]

\[ (H^*FL) \]

\[ (H^*i-H^*) \]

\[ (H^*i-L) \]

\[ (H^*i-FL) \]

\[ (H^*FL) \]

\[ (H^*i-H^*) \]

\[ (H^*i-L) \]

\[ (H^*i-FL) \]

\[ (H^*FL) \]

\[ (H^*i-H^*) \]

\[ (H^*i-L) \]

\[ (H^*i-FL) \]

\[ (H^*FL) \]

\[ (H^*i-H^*) \]

\[ (H^*i-L) \]

\[ (H^*i-FL) \]

\[ (H^*FL) \]

\[ (H^*i-H^*) \]

\[ (H^*i-L) \]

\[ (H^*i-FL) \]

\[ (H^*FL) \]

\[ (H^*i-H^*) \]

\[ (H^*i-L) \]

\[ (H^*i-FL) \]

\[ (H^*FL) \]

\[ (H^*i-H^*) \]

\[ (H^*i-L) \]

\[ (H^*i-FL) \]

\[ (H^*FL) \]

\[ (H^*i-H^*) \]

\[ (H^*i-L) \]

\[ (H^*i-FL) \]

\[ (H^*FL) \]

\[ (H^*i-H^*) \]

\[ (H^*i-L) \]

\[ (H^*i-FL) \]

\[ (H^*FL) \]

\[ (H^*i-H^*) \]

\[ (H^*i-L) \]

\[ (H^*i-FL) \]

\[ (H^*FL) \]

\[ (H^*i-H^*) \]

\[ (H^*i-L) \]

\[ (H^*i-FL) \]

\[ (H^*FL) \]

\[ (H^*i-H^*) \]

\[ (H^*i-L) \]

\[ (H^*i-FL) \]

\[ (H^*FL) \]

\[ (H^*i-H^*) \]

\[ (H^*i-L) \]

\[ (H^*i-FL) \]

\[ (H^*FL) \]
Figure 26. Representation of the linguistic measurement of pitch span for Initial Prominent High Minus Final Low (H*i-FL).

In the figure, the blue dot represents the Initial Prominent High, and the red dot represents the Final Low

- **Noninitial Prominent Peaks Minus Lows** (H*-L) evaluates the pitch range in the middle of an intonational phrase, and it excludes both the initial H and the final L.

Figure 27. Representation of the linguistic measurement of pitch span for Noninitial Highs Minus Lows (H*-L).

In the figure, the red dots represent the Noninitial Highs and the blue dots represent the Lows.
• **Noninitial Prominent Peaks Minus Final Lows** (H\*FL) measures the span between the noninitial Highs and the Final Low, which tends to occur at lower F0 levels.

![Diagram of pitch span measurement](image)

*Figure 28.* Representation of the linguistic measurement of pitch span for Noninitial Highs Minus Final Low (H-FL).

In the figure, the blue dots represent the Noninitial Highs and the red dot represents the Final Low.

• **Initial Prominent Peaks Minus Noninitial Prominent Peaks** (H*i-H). measures the span between the initial Highs and the noninitial Highs. Given than there tends to be downstepping, the second value is oftentimes lower than the first.
Figure 29. Representation of the linguistic measurement of pitch span for Initial High Minus Noninitial Highs (H*i-H).

In the figure, the red dot represents the Initial High and the blue dots represent the Noninitial Highs.

Analysis. A between-subjects design was implemented to answer research questions 1 and 2, and a within-subjects design was implemented to answer research question 3. Research question 1 sought to determine if peak alignment differed as an effect of proficiency, and if so, in what direction. For the classroom-based study, I compared the results of three different proficiency groups (beginning, intermediate, and advanced) in the reading task, and two groups (intermediate and advanced) in the semi-spontaneous task. The beginning group was excluded from the analysis in the semi-spontaneous task because most of the utterances from beginning participants in this task were isolated words rather than phrases. Therefore, it was not possible, in the vast majority of cases, to have peaks that clearly occurred in prenuclear position.

Research question 2 sought to determine if there was a difference in the relative height of final boundary tones that could have occurred as an effect of proficiency, and if so, in what direction this difference happened. For this research question, I only explored data from the reading task. The semi-spontaneous task led to speaking that was more conversational in nature. Hence, there were no clear phrase-final prosodic breaks, but rather isolated words or short
sentences that presented the characteristic intonation of unfinished speech (F0 that remained at a mid-point, instead of the lows that signal the end of the broad-focus declaratives).

Research question 3 looked into pitch range phenomena. More specifically, it sought to determine if there were differences in pitch level and pitch span that occurred as an effect of proficiency. To answer this research question, both absolute landmarks and the tonal difference between these landmarks were explored. These linguistic measures correspond, respectively, with pitch level and pitch span.

For all analyses, I first checked if the data met the assumptions for parametric tests. This was done by first visually inspecting the data and then conducting Shapiro-Wilk tests. For research questions 1 and 2, Shapiro-Wilk tests were run for each of the categories of the independent variable (i.e., each proficiency group). For research question 3, I first calculated the difference between each participant’s pitch range in English and Spanish, and then determined if the distribution of those differences were normal with a Shapiro-Wilk test. Research question 1 (reading task only) and research question 2 compared the performance of three different proficiency groups. In both cases, the assumption of normality was violated, so non-parametric Kruskal-Wallis tests were conducted. In the case of a statistical omnibus test, follow-up Mann-Whitney tests were conducted. A Bonferroni correction was implemented to control for multiple comparisons. Research question 1 (semi-spontaneous task) compared only two proficiency groups. In this case, an independent t-test was conducted after checking that the data met the assumptions for that test. Finally, for research question 3, I conducted paired-sample t-tests for data that met the assumptions for parametric tests. Whenever there were violations of the assumption of normality, Wilcoxon Signed Ranks tests were run. A Bonferroni correction was implemented to control for multiple pairwise comparisons.
Cohen’s $d$ was calculated for all parametric tests with the calculators from Lenhard and Lenhard (2016). Because Cohen’s $d$ assumes a normal distribution of the data, it was not calculated for non-parametric tests. Instead, the following effect size formula was used for non-parametric tests: $r = z/\sqrt{N}$ (Rosenthal, 1994). Interpretations of effect sizes were adopted from Plonsky and Oswald (2014). For mean differences between groups, $d$ values of .40 were considered small, .70 medium and 1.00 large. For mean differences within groups, $d$ values of .60 were considered small, 1.00 medium and 1.40 large. For correlation coefficients, $r$s around .25 were considered small, .40 medium, and .60 large.

**Summary of Methods.** A summary of the methods that were utilized to answer the three research questions for the classroom-based context is presented in Table 1.
Table 1. Summary of the methods utilized for the classroom context study

<table>
<thead>
<tr>
<th>Research question</th>
<th>Research Question 1</th>
<th>Research Question 2</th>
<th>Research Question 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Does the realization of Spanish prenuclear peak alignment change as L2 learners advance in proficiency?</td>
<td>Does the scaling of final boundary tones change as L2 learners advance in proficiency?</td>
<td>Do advanced L2 learners of Spanish evidence differences in pitch range between L2 Spanish and L1 English?</td>
</tr>
<tr>
<td>Parameter</td>
<td>Peak alignment</td>
<td>Final boundary tone</td>
<td>Pitch range</td>
</tr>
<tr>
<td>Participants</td>
<td>Beginning (n = 18), intermediate (n = 21), advanced (n = 24)</td>
<td>Beginning (n = 18), intermediate (n = 21), advanced (n = 24)</td>
<td>Only advanced female participants retained for analyses (n = 21)</td>
</tr>
<tr>
<td>Experimental design</td>
<td>Cross-sectional</td>
<td>Cross-sectional</td>
<td>Within-subject (L2 vs. L1)</td>
</tr>
<tr>
<td>Materials</td>
<td>Contextualized reading task and semi-spontaneous task</td>
<td>Contextualized reading task</td>
<td>Contextualized reading task</td>
</tr>
<tr>
<td>Dependent variable(s)</td>
<td>Distance from onset of stressed syllable until peak (in milliseconds)</td>
<td>tonal height of boundary tone (in ERBs)</td>
<td>(I), (FL), (L), (H<em>i/Hi), (H), H</em>i-FL, H*-FL, and H*i-H (in Hertz)</td>
</tr>
</tbody>
</table>

Note. For research question 1, the data from only the intermediate and advanced participants were analyzed for the semi-spontaneous task. All participants were retained for analyses for the contextualized reading task.

Results

Research Question 1: Prenuclear Peak Alignment. The first research question inquired if the realization of Spanish prenuclear F0 peak alignment changes as L2 learners advance in proficiency. The answer to this question was explored both via participants’ F0 peak alignment on the reading task and on the semi-spontaneous task.
**Reading Task.** The descriptive statistics for participants’ F0 peak alignment on the reading task are presented below in Table 2. These values represent the distance between the onset of the prenuclear syllable and the peak of F0.

*Table 2. Descriptive statistics for participants’ F0 peak alignment on the reading task, in classroom context.*

<table>
<thead>
<tr>
<th></th>
<th>Beginner (n = 18)</th>
<th>Intermediate (n = 21)</th>
<th>Advanced (n = 24)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>225.47</td>
<td>224.98</td>
<td>236.78</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>10.54</td>
<td>17.11</td>
<td>13.47</td>
</tr>
<tr>
<td>Median</td>
<td>225.14</td>
<td>225.29</td>
<td>236.97</td>
</tr>
<tr>
<td>Min</td>
<td>212.00</td>
<td>215.00</td>
<td>204.14</td>
</tr>
<tr>
<td>Max</td>
<td>250.75</td>
<td>240.29</td>
<td>257.13</td>
</tr>
</tbody>
</table>

*Note. All values are presented in milliseconds.*

The descriptive statistics indicated that advanced learners seemed to present later f0 peaks (\(M = 236.78, SD = 13.47\)) than beginner (\(M = 225.47, SD = 10.54\)) and intermediate (\(M = 224.98, SD = 17.11\)) learners. Inferential statistics were conducted to determine if this difference was statistically significant.

First, the data were examined to determine if they met the assumptions for parametric tests. A visual inspection of the data indicated that not all of the categories of the independent variable appeared to be normally distributed. Shapiro-Wilk tests indicated that the distribution of the data were normal for the beginner (\(W(18) = .94, p = .256\)) and intermediate (\(W(21) = .96, p = .520\)) groups, but non-normal for the advanced group (\(W(24) = .90, p = .018\)). Consequently, nonparametric tests were conducted.
A Kruskal-Wallis test indicated that there was a main effect of group for peak alignment, 
\( H(2) = 14.82, p = .001 \). Mann-Whitney tests were used to follow up on this finding. A
Bonferroni correction was applied so that all effects were reported at a \( p = .0167 \) level of
significance. The peak alignment of the advanced group was statistically later than the
intermediate \( (U = 96, z = -3.55, p < .001, r = -0.53) \), and beginner learners \( (U = 100, z = -2.95, p \)
= .003, \( r = -0.46) \), but there was no statistical difference between the intermediate and beginner
learners \( (U = 179, z = -0.28, p = .778, r = -0.05) \). The effect sizes indicated that the differences
between the advanced and intermediate/beginner learners were medium-sized.

**Semi-Spontaneous Task.** The descriptive statistics for participants’ peak alignment on
the semi-spontaneous task are presented below in Table 3. Data from beginning learners was
excluded from this analysis because their intonational units were most times comprised of only
one word. Therefore, alignment in those cases would not be classified as prenuclear.

Table 3. Descriptive statistics for participants’ peak alignment on the semi-spontaneous task, 
classroom context.

<table>
<thead>
<tr>
<th></th>
<th>Intermediate ((n = 21))</th>
<th>Advanced ((n = 24))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>217.89</td>
<td>225.20</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>18.48</td>
<td>19.32</td>
</tr>
<tr>
<td>Median</td>
<td>216.00</td>
<td>226.40</td>
</tr>
<tr>
<td>Min</td>
<td>198.25</td>
<td>205.00</td>
</tr>
<tr>
<td>Max</td>
<td>231.25</td>
<td>246.00</td>
</tr>
</tbody>
</table>

Note: All values are presented in milliseconds.
The descriptive statistics indicate that the advanced learners seemed to have produced later peaks ($M = 225.20$, $SD = 19.32$) than the intermediate learners ($M = 217.89$, $SD = 18.48$). Inferential statistics were conducted to determine if this difference was statistically significant.

First, the data were examined to determine if they met the assumptions for parametric tests. A visual inspection of the data indicated the categories of the independent variable appeared to be normally distributed. Shapiro-Wilk tests indicated that the data for the semi-spontaneous task met the assumptions for parametric tests for the intermediate ($W(21) = 0.96, p = .460$) and advanced ($W(18) = 0.94, p = .162$) learner groups. An independent $t$-test indicated that the difference between intermediate and advanced learners was statistically significant: advanced learners realized their peaks later than the intermediate learners $t(43) = -2.74, p = .009, d = 0.82$. The effect size indicated that the differences between the advanced and intermediate learners was medium-sized.

**Research Question 2: Sentence-final Boundary Tones.** The second research question inquired whether the height of final boundary tones change as L2 learners advance in proficiency. The descriptive statistics for participants’ sentence-final boundary tones are presented below in Table 4.
Table 4. Descriptive statistics for participants’ sentence-final boundary tones, in classroom context.

<table>
<thead>
<tr>
<th></th>
<th>Beginner</th>
<th>Intermediate</th>
<th>Advanced</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(n = 18)</td>
<td>(n = 21)</td>
<td>(n = 24)</td>
</tr>
<tr>
<td>Mean</td>
<td>4.91</td>
<td>5.38</td>
<td>5.37</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>1.30</td>
<td>.92</td>
<td>.98</td>
</tr>
<tr>
<td>Median</td>
<td>5.37</td>
<td>5.35</td>
<td>5.65</td>
</tr>
<tr>
<td>Min</td>
<td>2.99</td>
<td>3.12</td>
<td>2.65</td>
</tr>
<tr>
<td>Max</td>
<td>6.70</td>
<td>6.73</td>
<td>6.89</td>
</tr>
</tbody>
</table>

Note. All values are presented in Equivalent Rectangular Bandwidth (ERB) units.

The descriptive statistics indicate that the intermediate \((M = 5.38, SD = 0.92)\) and advanced \((M = 5.37, SD = 0.98)\) learners seemed to have realized slightly higher sentence-final boundary tones than the beginner learners \((M = 4.91, SD = 1.30)\). Inferential statistics were conducted to determine if this difference was statistically significant.

First, the data were examined to determine if they met the assumptions for parametric tests. A visual inspection of the data indicated that not all of the categories of the independent variable were normally distributed. Shapiro-Wilks tests confirmed that the data were not normally distributed for the beginner \((W(18) = .88, p = .031)\), intermediate \((W(21) = .90, p = .031)\), and advanced \((W(24) = .89, p = .013)\) groups. Levene’s test also indicated that the assumption of homogeneity of variance was violated, \(F(2,60) = 4.53, p = .015\). Consequently, nonparametric tests were conducted. A Kruskal-Wallis test indicated that there was no statistical difference in terms of height in the realization of sentence-final boundary tones for the beginner, intermediate and advanced learners, \(H(2) = .83, p = .659, d = .28\).
Research Question 3: Pitch Range. The third research question inquired whether advanced L2 learners of Spanish evidence differences in pitch range between L2 Spanish and L1 English. Only women were retained for these analyses. Pitch range was explored via pitch range measures and absolute landmark measures. A total of 10 pairwise comparisons were conducted to explore the changes in pitch range for all different pitch range measures and absolute landmarks. To control for multiple pairwise comparisons, a Bonferroni correction was applied so that all effects were reported at a $p = .005$ level of significance.

Pitch Range Measures. The descriptive statistics for participants’ pitch range in Spanish and English are presented below in Table 5.

Table 5. The descriptive statistics for participants’ pitch range (Span measures) in Spanish and English, in classroom context.

<table>
<thead>
<tr>
<th></th>
<th>Spanish (n = 21)</th>
<th>English (n = 21)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>H*i- L</td>
<td>H*i- FL</td>
</tr>
<tr>
<td>Mean</td>
<td>59.95</td>
<td>61.58</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>18.45</td>
<td>28.67</td>
</tr>
<tr>
<td>Median</td>
<td>62.37</td>
<td>67.27</td>
</tr>
<tr>
<td>Min</td>
<td>29.41</td>
<td>-9.95</td>
</tr>
<tr>
<td>Max</td>
<td>90.44</td>
<td>129.35</td>
</tr>
</tbody>
</table>

Note. All values are presented in Hertz.

Inferential statistics were performed to determine whether any of the differences in participants’ Spanish and English pitch ranges were statistical.

Initial Prominent Peaks Minus Lows ($H^*i$-L). First, the sampling distribution of the differences between Spanish and English initial prominent peaks minus lows scores were
examined to determine if they were normally distributed. A visual inspection of the data indicated that the differences appeared to be normally distributed. A Shapiro-Wilk test confirmed the normal distribution of the data ($W(21) = .96, p = .564$). A dependent $t$-test indicated that the pitch range difference for initial prominent peaks minus lows was not statistically different in Spanish ($M = 59.95, SD = 18.45$) and English ($M = 61.99, SD = 27.61$), $t(20) = -.65, p = .523, d = -0.12$.

*Initial Prominent Peaks Minus Final Lows (H*i-FL).* First, the sampling distribution of the differences between Spanish and English initial prominent peaks minus final lows scores were examined to determine if they were normally distributed. A visual inspection of the data indicated that the differences appeared to be normally distributed. A Shapiro-Wilk test confirmed the normal distribution of the data ($W(21) = .98, p = .859$). A dependent $t$-test indicated that the pitch range difference for initial prominent peaks minus final lows was not statistically different in Spanish ($M = 61.58, SD = 28.67$) and English ($M = 61.10, SD = 34.17$), $t(20) = .08, p = .941, d = 0.02$.

*Noninitial Prominent Peaks Minus Lows (H*-L).* First, the sampling distribution of the differences between Spanish and English noninitial prominent peaks minus lows scores were examined to determine if they were normally distributed. A visual inspection of the data indicated that the differences may not be normally distributed, and a Shapiro-Wilk test confirmed the non-normal distribution of the data ($W(21) = .90, p = .034$). Consequently, a nonparametric test was conducted. A Wilcoxon Signed-Rank test indicated that English ($Mdn = 28.08$) had a statistically different range for noninitial prominent peaks minus lows from Spanish ($Mdn = 24.70$), $z = -3.04, p = .002, r = -0.47$. The effect size indicated that the difference between Spanish and English noninitial prominent peaks minus lows scores was medium-sized.
Noninitial Prominent Peaks Minus Final Lows ($H^*-FL$). First, the sampling distribution of the differences between Spanish and English noninitial prominent peaks minus final lows scores were examined to determine if they were normally distributed. A visual inspection of the data indicated that the differences appeared to be normally distributed. A Shapiro-Wilk test confirmed the normal distribution of the data ($W(21) = .97, p = .647$). A dependent $t$-test indicated that the pitch range difference for noninitial prominent peaks minus final lows was not statistically different in Spanish ($M = 27.16, SD = 19.02$) and English ($M = 33.51, SD = 22.79$), $t(20) = -1.31, p = .205, d = -0.30$.

Initial Prominent Peaks Minus Noninitial Prominent Peaks ($H^*i-H^*$). First, the sampling distribution of the differences between Spanish and English initial prominent peaks minus noninitial prominent peaks scores were examined to determine if they were normally distributed. A visual inspection of the data indicated that the differences appeared to be normally distributed. A Shapiro-Wilk test confirmed the normal distribution of the data ($W(21) = .98, p = .848$). A dependent $t$-test indicated that the pitch range difference for initial prominent peaks minus noninitial prominent peaks was not statistically different in Spanish ($M = 34.42, SD = 20.12$) and English ($M = 27.60, SD = 27.92$), $t(20) = 1.36, p = .190, d = 0.27$.

Absolute Landmarks. Absolute landmarks serve to measure the level at which speakers are realizing their peaks and valleys. In Table 6, I present the descriptive of all absolute landmarks in both English and Spanish.
Table 6. The descriptive statistics for participants’ pitch range (Level measures) in Spanish and English, in classroom context.

<table>
<thead>
<tr>
<th></th>
<th>Spanish</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>(n = 21)</td>
<td></td>
<td>(n = 21)</td>
</tr>
<tr>
<td>I</td>
<td>233.80</td>
<td>222.21</td>
</tr>
<tr>
<td>H*i</td>
<td>256.63</td>
<td>255.83</td>
</tr>
<tr>
<td>L</td>
<td>197.68</td>
<td>194.09</td>
</tr>
<tr>
<td>H</td>
<td>222.21</td>
<td>224.66</td>
</tr>
<tr>
<td>F</td>
<td>195.05</td>
<td>228.24</td>
</tr>
<tr>
<td>I</td>
<td>236.40</td>
<td>252.05</td>
</tr>
<tr>
<td>H*i</td>
<td>252.27</td>
<td>252.28</td>
</tr>
<tr>
<td>L</td>
<td>199.92</td>
<td>197.04</td>
</tr>
<tr>
<td>H</td>
<td>225.58</td>
<td>231.55</td>
</tr>
<tr>
<td>F</td>
<td>225.05</td>
<td>198.73</td>
</tr>
<tr>
<td>Mean</td>
<td>26.66</td>
<td>21.77</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>26.66</td>
<td>21.77</td>
</tr>
<tr>
<td>Median</td>
<td>178.20</td>
<td>188.03</td>
</tr>
<tr>
<td>Min</td>
<td>297.70</td>
<td>329.25</td>
</tr>
<tr>
<td>Max</td>
<td>301.85</td>
<td>297.53</td>
</tr>
</tbody>
</table>

Note. All values are presented in Hertz.

**Initial Value (I).** First the sampling distribution of the differences between Spanish and English initial level scores was examined to determine if they were normally distributed. A visual inspection of the data indicated that the differences appeared to be normally distributed. A Shapiro-Wilk test confirmed the normal distribution of the data \( W(21) = .99, p = .982 \). A dependent \( t \)-test indicated that the difference in the initial level in Spanish (\( M = 233.80, SD = 26.66 \)) and English (\( M = 224.66, SD = 21.77 \)) was not statistical, \( t(20) = 2.69, p = .014, d = 0.36 \).

**Initial Prominent Peak (H*i).** First, the sampling distribution of the differences between Spanish and English initial prominent peaks scores was examined to determine if they were normally distributed. A visual inspection of the data indicated that the differences appeared to be normally distributed. A Shapiro-Wilk test confirmed the normal distribution of the data \( W(21) = .95, p = .333 \). A dependent \( t \)-test indicated that the difference in the initial prominent peaks in Spanish (\( M = 256.63, SD = 25.99 \)) and English (\( M = 255.83, SD = 38.76 \)) was not statistical, \( t(20) = 0.16, p = .875, d = 0.02 \).
Valleys (L). First, the sampling distribution of the differences between Spanish and English low scores were examined to determine if they were normally distributed. A visual inspection of the data indicated that the differences may not be normally distributed, and a Shapiro-Wilk test confirmed that the data was approaching a non-normal distribution ($W(21) = .920, p = .087$). Even though the $p$-value for the Shapiro-Wilk test did not reach alpha of .05, a non-parametric test was conducted to err on the side of caution. A Wilcoxon Signed Rank test indicated English ($M = 194.09, SD = 16.59$) and Spanish ($M = 197.68, SD = 18.46$) lows were not statistical at the adjusted $p$-value, $z = -2.45, p = .014, r = -0.38$, but the effect size indicated that the difference between the Spanish and English lows was between small and medium.

Noninitial Peaks (H). First, the sampling distribution of the differences between Spanish and English low scores were examined to determine if they were normally distributed. A visual inspection of the data indicated that the differences may not be normally distributed, and a Shapiro-Wilk test confirmed the non-normal distribution of the data ($W(21) = .890, p = .022$). Consequently, a nonparametric test was conducted. A Wilcoxon Signed Rank test indicated that English ($M = 228.24, SD = 24.97$) had statistically higher noninitial highs than Spanish ($M = 222.21, SD = 17.96$), $z = -3.04, p = .002, r = -0.28$. The effect size indicated that the difference between the Spanish and English noninitial highs was small.

---

3 A parametric dependent $t$-test was also conducted to ensure that the test selected did not change the interpretation of results. The dependent $t$-test indicated that the difference in the lows in Spanish and English were not statistical at the adjusted $p$-value, $t(20) = 2.02, p = .057, d = 0.20$, and the effect size was nominal. Because the visual inspection of the data indicated non-normality, and because the Shapiro-Wilk test was approaching statistical significance, the interpretation of the nonparametric Wilcoxon Signed-Rank test and associated effect size was retained.
Final Level (F). First, the sampling distribution of the differences between Spanish and English final level scores were examined to determine if they were normally distributed. A visual inspection of the data indicated that the differences appeared to be normally distributed. A Shapiro-Wilk confirmed the normal distribution of the data ($W(21) = .942, p = .234$). A dependent $t$-test indicated that the difference in final level in Spanish ($M = 195.05, SD = 24.28$) and English ($M = 194.73, SD = 21.17$) was not statistical, $t(20) = 0.12, p = .905, d = 0.01$.

Summary of Results. The first research question inquired if the realization of Spanish prenuclear F0 peak alignment changes as L2 learners advance in proficiency. Peak alignment was measured with a contextualized reading task and a semi-spontaneous task. The contextualized reading task tested advanced, intermediate and beginner learners, while the semi-spontaneous task tested only advanced and intermediate learners. Results indicated that on the reading task, the advanced learners realized F0 peaks that occurred statistically later than the intermediate and the beginner learners. There was a medium effect of the difference between advanced and intermediate learners ($r = -0.53$) and advanced and beginner learners ($r = -0.46$). On the semi-spontaneous task, the advanced learners produced statistically later F0 peaks than the intermediate learners, which corresponded to a medium-sized effect ($d = .82$).

The second research question inquired whether the scaling of final boundary tones changes as L2 learners advance in proficiency. Although the descriptive statistics indicated that the intermediate ($M = 5.38, SD = 0.92$) and advanced ($M = 5.37, SD = 0.98$) learners seemed to have a slightly higher sentence-final boundary tone than the beginner learners ($M = 4.91, SD = 1.30$), the inferential statistics indicated that there was no statistical difference in the sentence-final boundary tones for the beginner, intermediate and advanced groups.
The third research question inquired if advanced L2 learners of Spanish evidence differences in pitch range between L2 Spanish and L1 English. Pitch range was explored via pitch range measures and absolute landmarks. For this question, only the data for the advanced female participants was retained for analysis. Pitch range was explored via pitch range measures and absolute landmarks. The results are summarized in Table 7.
Table 7. Summary of results of pitch span and pitch level measures.

<table>
<thead>
<tr>
<th>Pitch range measure</th>
<th>p-value</th>
<th>effect size</th>
<th>Effect size interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial Prominent Peaks Minus Lows (H*i-L)</td>
<td>p = .523</td>
<td>d = -0.12</td>
<td>no effect</td>
</tr>
<tr>
<td>Initial Prominent Peaks Minus Final Lows (H*i-FL)</td>
<td>p = .941</td>
<td>d = 0.02</td>
<td>no effect</td>
</tr>
<tr>
<td>Noninitial Prominent Peaks Minus Lows (H*-L)</td>
<td>p = .002*</td>
<td>r = -0.47</td>
<td>medium effect</td>
</tr>
<tr>
<td>Noninitial Prominent Peaks Minus Final Lows (H*-FL)</td>
<td>p = .205</td>
<td>d = -0.30</td>
<td>no effect</td>
</tr>
<tr>
<td>Initial Prominent Peaks Minus Noninitial Prominent Peaks (H<em>i-H</em>)</td>
<td>p = .190</td>
<td>d = 0.27</td>
<td>no effect</td>
</tr>
<tr>
<td>Initial Value (I)</td>
<td>p = .014</td>
<td>d = 0.36</td>
<td>no effect</td>
</tr>
<tr>
<td>Initial Prominent Peak (H*i)</td>
<td>p = .875</td>
<td>d = 0.02</td>
<td>no effect</td>
</tr>
<tr>
<td>Valleys (L)</td>
<td>p = .014</td>
<td>r = -0.38</td>
<td>small/medium effect</td>
</tr>
<tr>
<td>Noninitial Peaks (H)</td>
<td>p = .002*</td>
<td>r = -0.28</td>
<td>small effect</td>
</tr>
<tr>
<td>Final Level (F)</td>
<td>p = .905</td>
<td>d = 0.01</td>
<td>no effect</td>
</tr>
</tbody>
</table>

Note. All effects sizes are interpreted according to Plosky and Oswald’s (2014) recommendations. * = p is significant at the .005 level.
The inferential statistics revealed that two of the pitch range measures were statistically significant at the adjusted alpha level of .005. One was noninitial prominent peaks minus lows, with English scoring higher than Spanish. The effect size was medium. The other statistical result was noninitial prominent peaks, with English scoring higher than Spanish. This was a small effect. It should be noted that there was also a small/medium effect for a difference in valleys in English and Spanish, with English scoring lower than Spanish, but the test results did not reach statistical significance at the corrected alpha level.
Chapter 4. Study Abroad Context

Method

The SA program where this investigation was conducted took place in Barcelona, Spain, in the summer of 2017. It was characterized by its highly selective criteria for participation and for its commitment to developing learners’ linguistic and cultural abilities. To be able to participate in this program, the L2 Spanish learners needed to have previously completed two advanced placement Spanish courses and have a minimum GPA of 2.7. Once the selection process was completed, the L2 Spanish learners had to sign a language pledge before the program started. In this pledge, they committed to speaking only in Spanish for the duration of the program. While in Barcelona, learners had 12 hours of class per week, which were divided into three content courses exclusively taught in Spanish. Additionally, students had to complete homework assignments, which were expected to take the learners about 10 hours per week to complete. Besides classes and homework, participants had to partake in two other activities: field trips to historical and cultural sites, and weekly conversation exchanges with native speakers.

All the L2 Spanish learners who participated in this program stayed in rooms, by themselves, during their sojourn abroad. The rooms where they all stayed were located in the same apartment complex, within walking distance from the university where they took their classes. Learners shared some facilities in this building complex, like the kitchens, so they had to interact with each other even when they were not in class or in an activity organized by the program.

In the following sections, I present the methods section, the analysis section, and the results section for the study that I conducted during this SA program in Barcelona. A description of the participant pool is followed by the recruitment protocol, the materials, and the procedure.
This is followed by the coding schema. Whenever the methodology employed for the SA study is the same as the classroom study, it will be stated. In the analysis section that follows, there is detailed information about the statistical analyses that were implemented, as well as the rationale for their use. Finally, statistical analyses for all the variables under study are presented in this order: a) F0 peak alignment, b) final boundary tone, and c) pitch range.

**Participants.** Twenty Spanish learners volunteered to partake in this SA study. To be retained for analysis, participants had to comply with the following criteria a) be a native speaker of English, b) be enrolled in the study abroad program in Barcelona, Spain, organized by Georgetown University, and c) not be a heritage speaker of any language. Furthermore, for a participant to be retained for analysis, she or he had to have at least 80% of analyzable sentences in the contextualized reading task. The data from 1 participant had to be discarded for not complying with one of these criteria (Spanish heritage speaker, \( n = 1 \)). The final number of participants whose data was retained for analysis was 19.

For the study of pitch range, there was an additional criterion for inclusion: only female learners were retained for analysis. This was done in order to avoid gender effects on F0 range, a distortion that could have been particularly acute in the present study given the disparity in the distribution of gender within the advanced-SA group (female, \( n = 16 \); male, \( n = 3 \)).

**Recruitment.** L2 Spanish participants were informed of the possibility to participate in an experimental study about L2 acquisition through an email from the SA director or one of her research assistants. Participants were asked to contact me if they wished to participate.

**Materials.** For this investigation in a study abroad context, I used the same materials as in the classroom-based context (See Method in Chapter 3), except the semi-spontaneous task. I excluded this task because the objective of this investigation in a SA context was to determine if
there were differences as a consequence of SA, not task-type. I included an additional language contact profile especially designed for the SA study. The language contact profile was added to obtain data pertaining to learners’ language background and language use, both before and during their stay abroad. A contextualized reading task, which was detailed in the Method section in Chapter 3, was utilized to collect the controlled utterances that are used in this study.

**Language Contact Profile.** The language contact profile implemented in this study is that of Freed, Dewey, Segalowitz and Halter (2004). Specially designed for SA contexts, this instrument is divided into two versions, which were implemented as a pre-test and post-test. The pre-test section is further divided into two parts: one gathered general background information regarding participants’ demographics (e.g., gender, age, country of birth) and language-learning experience (e.g., Spanish classes, languages spoken at home), and another part focused on language use prior to the SA program: who participants spoke with the most (e.g., professors, peers, strangers, etc.), what other activities they engaged in that could have helped them further develop their Spanish (e.g., reading, going to the movies), and so forth.

The post-test version helped build a language contact profile of the participants during the SA program. This part of the language contact profile therefore focused exclusively on the sojourn abroad, gathering information about who participants spoke Spanish with the most and to what extent, in what situations they used Spanish and how often they did so, etc. Finally, this instrument compiles information about L1 use (English, in this case), asking participants to also detail what activities they did using the L1 and how much of their time these activities took.

**Procedure.** Data collection occurred twice, once within the first week of their SA experience (pre-test, in Barcelona) and again during the last week of the stay abroad (post-test, in Barcelona). Data collection for the pre-test and post-test occurred in a private room at the same
facility where the L2 Spanish participants’ dormitories were located. During the pretest, L2 Spanish participants first signed a consent form, and then completed the language contact profile, and a contextualized reading task. In the posttest, L2 Spanish participants completed the second part of the language contact profile, and a contextualized reading task.

The data collection lasted approximately 50 minutes total for the pre-test, and 50 minutes for the post-test: 20 minutes for the background questionnaire and language contact profile, 30 minutes for the contextualized reading task. For the contextualized reading task, the L2 Spanish participants’ responses were recorded with a head-worn Shure SM10A-CN Cardioid Dynamic microphone. All data were collected by the researcher or a research assistant. Participants were offered extra credit in the class in which they had the lowest grade. The questions in the language contact profile were digitized and presented to participants on SurveyMonkey.

**Coding.** The L2 Spanish participants’ recordings for the contextualized reading task were analyzed using Praat (Boersma & Weenink, 2011). The 10 target declarative sentences in Spanish were identified in the audio files of each participant’s recording of the contextualized reading task, both in the pretest and posttest. This process resulted in 380 sentences. Out of these 380 sentences, 17 had to be eliminated because of poor audio quality or because reading was unnatural. For a participant to be retained for analysis, she or he had to have at least 80% of analyzable sentences in the contextualized reading task. No participant was dropped from analysis because of this.

For the transcription of tonal events, the coding was made following the specifications of the notation system Tones and Break Indices (ToBI: Beckman & Ayers-Elam 1997), particularly its version for Spanish (Estebas-Vilaplana & Prieto, 2008), and followed the same procedures explained in Chapter 3 for Study 1.
For the analysis of pitch range, I also followed the same procedures described in Chapter 3, and I identified the following linguistic measures: Initial value (I) Final value (FL), Local Minima or Valleys (L), Prominent Peaks (H), and Initial Prominent Peak (H*i). About 50-55 data points were recorded per participant per session (pretest-posttest), for a total of 2,052 data points. Regions where there was creaky voice or interruptions in the speech flow were regarded as unanalyzable, and hence not included.

**Analyses.** A between-subjects design was implemented to answer research questions 1, 2 and 3. For all analyses, I first checked if the data met the assumptions for parametric tests. This was done by first visually inspecting the data and then conducting Shapiro-Wilk tests on the difference between participants’ pre and post study abroad scores. If the assumption of normality was not violated, then I conducted dependent *t*-tests. If the assumption of normality was violated, I conducted Wilcoxon Signed Rank tests. Because a series of 10 pairwise comparisons were needed to answer research question 3, a Bonferroni correction was implemented for that research question to control for multiple pairwise comparisons. All effects for research question 3 are reported at a *p* = .005 level of significance.

Cohen’s *d* was calculated for all parametric tests with the calculators from Lenhard and Lenhard (2016). Because Cohen’s *d* assumes a normal distribution of the data, it was not calculated for non-parametric tests. Instead, the following effect size formula was used for non-parametric tests: \( r = \frac{z}{\sqrt{N}} \) (Rosenthal, 1994). Interpretations of effect sizes were adopted from Plonsky & Oswald (2014). For mean differences within groups, *d* values of .60 were considered small, 1.00 medium and 1.40 large. For correlation coefficients, *rs* around .25 were considered small, .40 medium and .60 large.
**Summary of Method.** A summary of the methods that were utilized to answer the three research questions for the SA context is presented in Table 8.

*Table 8. Summary of the methods utilized for the study abroad study*

<table>
<thead>
<tr>
<th>Research question</th>
<th>Parameter</th>
<th>Participants</th>
<th>Experimental design</th>
<th>Materials</th>
<th>Dependent variable(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research Question 1</td>
<td>Peak alignment</td>
<td>Advanced + SA (n = 19)</td>
<td>Pre/post-design</td>
<td>Contextualized reading task</td>
<td>Distance from onset of stressed syllable until peak (in milliseconds)</td>
</tr>
<tr>
<td>Research Question 2</td>
<td>Final boundary tone</td>
<td>Advanced + SA (n = 19)</td>
<td>Pre/post-design</td>
<td>Contextualized reading task</td>
<td>Tonal height of boundary tone (in ERBs)</td>
</tr>
<tr>
<td>Research Question 3</td>
<td>Pitch range</td>
<td>Only female Advanced + SA retained for analyses (n = 16)</td>
<td>Pre/post-design</td>
<td>Contextualized reading task</td>
<td>(I), (FL), (L), (H<em>i/Hi), (H), H</em>i-FL, H*-FL, and H*i-H</td>
</tr>
</tbody>
</table>

**Results**

**Research Question 1: Peak Alignment.** The first research question inquired whether advanced L2 learners of Spanish evidence changes in the realization of prenuclear peak alignment after a short-term 5-week study abroad program. The descriptive statistics are presented in Table 9.
Table 9. Descriptive statistics for peak range alignment in study abroad (pre-test and post-test)

<table>
<thead>
<tr>
<th></th>
<th>Pre Study Abroad (n = 19)</th>
<th>Post Study Abroad (n = 19)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>234.74</td>
<td>232.23</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>11.27</td>
<td>10.98</td>
</tr>
<tr>
<td>Median</td>
<td>233.29</td>
<td>230.88</td>
</tr>
<tr>
<td>Min</td>
<td>215.88</td>
<td>213.71</td>
</tr>
<tr>
<td>Max</td>
<td>265.88</td>
<td>266.75</td>
</tr>
</tbody>
</table>

Note. All values are presented in milliseconds.

The descriptive statistics indicated that there was not a large difference between the pre-test ($M = 234.74, SD = 11.27$) and post-test ($M = 232.23, SD = 10.98$) values in terms of F0 peak alignment. Inferential statistics were conducted to confirm that the difference was not statistical.

First, the sampling distribution of the differences between the F0 peak alignment values of pre and post study abroad were examined to determine if they were normally distributed. A visual inspection of the data indicated that the differences appeared to be normally distributed. A Shapiro-Wilk test confirmed the normal distribution of the data ($W(19) = .95, p = .325$). A dependent $t$-test indicated that the pre and post F0 values of peak alignment for these advanced learners were not statistically different, $t(18) = 1.13, p = .272, d = 0.23$.

**Research Question 2: Sentence-final Boundary Tones.** The second research question inquired whether advanced L2 learners of Spanish evidence changes in the realization of final boundary tones after a short-term 5-week study abroad program.
Table 10. Descriptive statistics for final boundary tones in study abroad (pre-test and post-test).

<table>
<thead>
<tr>
<th></th>
<th>Pre-Study Abroad (n = 19)</th>
<th>Post-Study Abroad (n = 19)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>5.06</td>
<td>5.08</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>0.99</td>
<td>0.98</td>
</tr>
<tr>
<td>Median</td>
<td>5.38</td>
<td>5.30</td>
</tr>
<tr>
<td>Min</td>
<td>3.23</td>
<td>3.22</td>
</tr>
<tr>
<td>Max</td>
<td>6.76</td>
<td>7.07</td>
</tr>
</tbody>
</table>

Note. All values are presented in Equivalent Rectangular Bandwidth (ERB) units.

The descriptive statistics did not indicate a large difference between the pre-test ($M = 5.06, SD = 0.99$) and the post-test ($M = 5.08, SD = 0.98$) sentence-final boundary tones. This finding was explored further with inferential statistics.

First, the sampling distribution of the differences in the realization of the pre-test and post-test sentence-final boundary tones were examined to determine if they were normally distributed. A visual inspection of the data indicated that the differences appeared to be normally distributed. A Shapiro-Wilk test confirmed the normal distribution of the data ($W(19) = .98, p = .979$). A dependent $t$-test indicated that the pre-test and post-test phrase-final boundary tones for these learners were not statistically different, $t(18) = -0.33, p = .745, d = -0.03$.

**Research Question 3: Pitch Range.** The third research question inquired whether L2 advanced learners of Spanish evidence changes in pitch range after a short-term 5-week study abroad program. Only women were retained for these analyses. Pitch range was explored via pitch range measures and absolute landmarks. A total of 10 pairwise comparisons were conducted to explore the changes in pitch range for all the different measures and absolute
landmarks. To control for multiple pairwise comparisons, a Bonferroni correction was applied so that all effects are reported at a $p = .005$ level of significance.

**Pitch Range Measures.** The descriptive statistics for the pitch range measures are presented in Table 11.

*Table 11. Descriptive statistics for pitch range (Span measures) in the study abroad (pre-test and post-test)*

<table>
<thead>
<tr>
<th></th>
<th>Pre study abroad (n = 16)</th>
<th>Post study abroad (n = 16)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>H*i-L</td>
<td>H*i-</td>
</tr>
<tr>
<td></td>
<td>FL</td>
<td>FL</td>
</tr>
<tr>
<td>Mean</td>
<td>64.00</td>
<td>76.71</td>
</tr>
<tr>
<td>Median</td>
<td>63.51</td>
<td>69.06</td>
</tr>
<tr>
<td>Min</td>
<td>29.61</td>
<td>37.73</td>
</tr>
<tr>
<td>Max</td>
<td>123.23</td>
<td>137.28</td>
</tr>
</tbody>
</table>

Note. All values are presented in Hertz.

Learners performed similarly on the pitch range measures before and after studying abroad. Inferential statistics were conducted to determine if any differences between participants’ pitch ranges before and after studying abroad were statistical. The inferential statistics for each pitch range measure will be explored below in turn.

*Initial Prominent Peaks Minus Lows (H*i-L).* First, the sampling distribution of the differences between the pre-test and the post-test initial prominent peaks minus lows scores were examined to determine if they were normally distributed. A visual inspection of the data
indicated that the differences may not be normally distributed. A Shapiro-Wilk test, however, indicated that the distribution of the data was normal \((W(16) = .930, p = .242)\). A dependent \(t\)-test\(^4\) indicated that the pitch range difference for initial prominent peaks minus lows was not statistically different before \((M = 64.00, SD = 25.19)\) and after \((M = 64.35, SD = 25.80)\) study abroad, \(t(15) = -.08, p = .935, d = -0.01\).

**Initial Prominent Peaks Minus Final Lows (H*i-FL).** First, the sampling distribution of the differences between the pre-test and the post-test initial prominent peaks minus final lows scores were examined to determine if they were normally distributed. A visual inspection of the data indicated that the differences appeared to be normally distributed. A Shapiro-Wilks confirmed the normal distribution of the data \((W(16) = .95, p = .428)\). A dependent \(t\)-test indicated that the pitch range difference for initial prominent peaks minus final lows was not statistically different before \((M = 76.71, SD = 30.68)\) and after \((M = 75.87, SD = 34.21)\) study abroad, \(t(15) = .175, p = .863, d = .03\).

**Noninitial Prominent Peaks Minus Lows (H*-L).** First, the sampling distribution of the differences between the pre-test and the post-test noninitial prominent peaks minus lows scores were examined to determine if they were normally distributed. A visual inspection of the data indicated that the differences appeared to be normally distributed. A Shapiro-Wilk test confirmed the normal distribution of the data \((W(16) = .983, p = .984)\). A dependent \(t\)-test indicated that the pitch range difference for noninitial prominent peaks minus lows was not statistically different

\(^4\) Because the visual inspection indicated that the differences may not be normally distributed, a non-parametric Wilcoxon Signed-Rank test was conducted in addition to the dependent \(t\)-test. The Wilcoxon signed-rank test indicated that the pre and post study abroad groups were not statistically different \((z = -0.05, p = .959, r = -0.01)\), thus maintaining the interpretation of the results of the \(t\)-test.
before \((M = 25.89, SD = 11.39)\) and after \((M = 21.92, SD = 14.42)\) study abroad, \(t(15) = 1.12, p = .281, d = .31\).

**Noninitial Prominent Peaks Minus Final Lows (H*-FL)**. First, the sampling distribution of the differences between the pre-test and the post-test noninitial prominent peaks minus final lows scores were examined to determine if they were normally distributed. A visual inspection of the data indicated that the differences appeared to be normally distributed. A Shapiro-Wilk test confirmed the normal distribution of the data \((W(16) = .95, p = .457)\). A dependent \(t\)-test indicated that the pitch range difference for noninitial prominent peaks minus final lows was not statistically different before \((M = 38.69, SD = 18.15)\) and after \((M = 33.44, SD = 19.88)\) study abroad, \(t(15) = 1.26, p = .227, d = 0.28\).

**Initial Prominent Peaks Minus Noninitial Prominent Peaks (H*i-H*)**. First, the sampling distribution of the differences between the pre-test and the post-test initial prominent peaks minus noninitial prominent peaks scores were examined to determine if they were normally distributed. A visual inspection of the data indicated that the differences appeared to be normally distributed. A Shapiro-Wilk test confirmed the normal distribution of the data \((W(16) = .961, p = .688)\). A dependent \(t\)-test indicated that the pitch range difference for initial prominent peaks minus noninitial prominent peaks was not statistically different before \((M = 38.03, SD = 14.83)\) and after \((M = 42.43, SD = 18.79)\) study abroad, \(t(15) = -1.66, p = .117, d = -0.25\).

**Absolute Landmarks**. The descriptive statistics for the absolute landmarks are presented in Table 12.
Table 12. Descriptive statistics for pitch range (Level measures) in the study abroad (pre-test and post-test)

<table>
<thead>
<tr>
<th></th>
<th>Pre study abroad</th>
<th>Post Study abroad</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I</td>
<td>H*i</td>
</tr>
<tr>
<td>Mean</td>
<td>224.53</td>
<td>258.07</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>25.11</td>
<td>34.54</td>
</tr>
<tr>
<td>Median</td>
<td>222.95</td>
<td>246.87</td>
</tr>
<tr>
<td>Min</td>
<td>185.70</td>
<td>203.43</td>
</tr>
<tr>
<td>Max</td>
<td>271.63</td>
<td>316.67</td>
</tr>
</tbody>
</table>

Note. All values are presented in Hertz.

Learners performed similarly on the absolute landmark measures before and after studying abroad. Inferential statistics were conducted to determine if any differences between participants’ absolute landmark measures before and after studying abroad were statistical. The inferential statistics for each absolute landmark measure will be explored below in turn.

Initial Value (I). First, the sampling distribution of the differences between the pre-test and the post-test initial level scores were examined to determine if they were normally distributed. A visual inspection of the data indicated that the differences may not be normally distributed. However, a Shapiro-Wilk test confirmed the normal distribution of the data \(W(16) = .93, p = .383\). A dependent \(t\)-test\(^5\) indicated that the difference in the initial level was not statistically different before \((M = 224.53, SD = 25.11)\) and after \((M = 225.28, SD = 22.25)\) study abroad, \(t(15) = -1.76, p = .863, d = -0.31\).

\(^5\) Because the visual inspection indicated that the differences may not be normally distributed, a non-parametric Wilcoxon signed-rank test was conducted in addition to the dependent \(t\)-test. The Wilcoxon signed-rank test indicated that the pre and post study abroad groups were not statistically different \((z = -0.47, p = .642, r = -0.08)\), thus maintaining the interpretation of the results of the \(t\)-test.
Initial Prominent Peak ($H^i$). First, the sampling distribution of the differences between the pre-test and the post-test initial prominent peaks scores were examined to determine if they were normally distributed. A visual inspection of the data indicated that the differences appeared to be normally distributed. A Shapiro-Wilk test confirmed the normal distribution of the data ($W(16) = .961, p = .688$). A dependent $t$-test indicated that the difference in the initial prominent peaks was not statistically different before ($M = 258.07, SD = 34.54$) and after ($M = 258.33, SD = 38.74$) study abroad, $t(15) = -.055, p = .956, d = -0.01$.

Lows or valleys ($L$). First, the sampling distribution of the differences between the pre-test and the post-test low scores were examined to determine if they were normally distributed. A visual inspection of the data indicated that the differences appeared to be normally distributed. A Shapiro-Wilk test confirmed the normal distribution of the data ($W(16) = .968, p = .812$). A dependent $t$-test indicated that the valleys were not statistically different before ($M = 194.07, SD = 20.90$) and after ($M = 193.99, SD = 18.05$) study abroad, $t(15) = 0.02, p = .982, d = 0.00$.

Noninitial Prominent Peaks ($H$). First, the sampling distribution of the differences between the pre-test and the post-test low scores were examined to determine if they were normally distributed. A visual inspection of the data indicated that the differences appeared to be normally distributed. A Shapiro-Wilk test confirmed the normal distribution of the data ($W(16) = .98, p = .917$). A dependent $t$-test indicated that the difference in the noninitial prominent peaks was not statistically different before ($M = 220.04, SD = 27.61$) and after ($M = 215.90, SD = 25.63$) study abroad, $t(15) = 1.29, p = .218, d = 0.15$.

Final Value ($F$). First, the sampling distribution of the differences between the pre-test and the post-test final level scores were examined to determine if they were normally distributed. A visual inspection of the data indicated that the differences appeared to be normally distributed.
A Shapiro-Wilk test confirmed the normal distribution of the data ($W(16) = .971, p = .857$). A dependent $t$-test indicated that the difference in the final level was not statistically different before ($M = 181.35, SD = 29.22$) and after ($M = 182.47, SD = 27.86$) study abroad, $t(15) = -.26, p = .801, d = -0.04$.

**Summary of Results.** The first research question inquired whether advanced L2 learners of Spanish evidence changes in the realization of prenuclear peak alignment after a short-term 5-week study abroad program. The results indicated that there were no statistical changes in the realization of prenuclear peak alignment after a 5-week study abroad program.

The second research question inquired whether advanced L2 learners of Spanish evidence changes in the realization of final boundary tones after a short-term 5-week study abroad program. The results indicated that the L2 learners did not evidence changes in the realization of boundary tones after a 5-week study abroad program.

The third research question inquired whether advanced L2 learners of Spanish evidence changes in pitch range after a short-term 5-week study abroad program. For this question, only the data for the female participants were retained for analysis. Pitch range was explored via pitch range measures and absolute landmarks. The results indicated that the participants did not evidence changes in pitch range after a 5-week study abroad program.
Chapter 5. Discussion and Conclusions

Overview

The main and overarching objective of this dissertation has been to provide theoretical analyses and interpretations of L2 Spanish intonation phenomena. The discussion that follows is not limited to my results, as it is also based on results reported in previous research, as well as the theoretical underpinnings of the LILt and, when possible, other models or theories of phonological acquisition. Furthermore, the discussion in this chapter brings together results from both classroom context and SA context research so as to provide a unified theoretical account of L2 Spanish intonation phenomena. This chapter is organized as follows: first, I present a discussion of each of the variables under study (prenuclear peak alignment, final boundary tones, and pitch range); second, I consider what the limitations of the current study are, as well as possible areas of future research; and third, I present the final conclusions of this dissertation.

Discussion

Peak Alignment. The results of this dissertation indicate that there are proficiency-based differences in the temporal alignment of F0 peaks along the segmental string. The tonal alignment of the prenuclear pitch accent (L*<)H* onto the text was comparatively, and significantly, delayed in learners of advanced proficiency. No significant differences were observed between beginner and intermediate learners. Lower proficiency learners regularly aligned their F0 peaks soon after the onset of the first posttonic syllable, whereas advanced learners aligned theirs significantly later, albeit also within the confines of the first posttonic syllable. The difference between intermediate and advanced learners held true for the controlled reading task as well as for the semi-spontaneous task. Thornberry (2014) and Zárate-Sández (2015), who used controlled and semi-spontaneous tasks as I did, found no significant effects
attributable to task type. The results in the present study confirmed that regardless of task type, advanced learners differentiated themselves from the other less proficient groups. Figure 30 below depicts a graphic representation of peak alignment behavior for all the proficiency groups under study in this dissertation:

![Figure 30](image)

**Figure 30.** Representation of the results of F0 peak alignment behavior by group.
The dotted lines represent the (B)eginning and (I)ntermediate learners, whereas the continuous line represent the (A)dvanced learners.

The study abroad group, whose prenuclear F0 peaks were already considerably delayed at the program’s onset, evidenced no significant differences from pre-test to post-test. That is, after being immersed in Barcelona for a 5-week short-term study abroad program, these advanced learners did not change the way in which they aligned their F0 peaks in prenuclear position. Finding marginal or no results in phonological development in study abroad, however, comes as no surprise, as some researchers have concluded the same (Bongiovanni, Long, Solon, & Willis, 2015; Díaz Campos, 2014), particularly in terms of intonational development in short-term programs (e.g., Henricksen et al., 2010). Furthermore, the exceptions that have been found in some previous studies (e.g., Henricksen et al. 2010; Méndez Seijas, 2018; Craft, 2015) could only be appreciated at the individual level.
To confirm if at least some learners in the SA group had improved, but their results were undetectable from the macro, group-wide perspective, I took a more detailed look at the data from the language profile (Freed et al., 2004) and compared it with the individualized results for peak alignment. This helped me determine if speaking more to native speakers of Spanish during the stay abroad or interacting less in English with classmates, among many other factors, could potentially correlate with enhanced individualized improvement. However, a preliminary exploration of the data indicated that none of these factors seemed to have a tangible impact. That is, learners who stood out because their F0 peak alignment occurred markedly later had already developed this feature by the program’s onset. In other words, time spent speaking to native speakers, watching movies or listening to music in Spanish, to name just a few variables, did not seem to be correlated with any changes in tonal behavior in prenuclear position. As Trofimovich and Baker’s (2007) investigation had suggested, there are prosodic properties that will only develop, if at all, after prolonged stays abroad. The results of this dissertation and other studies with analogous learner populations and contexts (e.g., Thornberry, 2014; Trimble, 2013a) appear to indicate that tonal alignment is probably one of the aspects that are relatively resistant to acquisition in early stages of development.

As discussed at length in the literature review, tonal alignment properties in English and Spanish have been extensively studied. Spanish, on the one hand, exhibits rising intonation in prenuclear position, and peak alignment is typically realized in posttonic syllables (Face, 1999; Sosa, 1999). In English, on the other hand, prenuclear F0 peaks tend to be more closely aligned with stressed syllables (Pierrehumbert, 1980). The degree of peak rightward displacement in both English and Spanish is conditioned by a series of factors that include, but are not limited to, pressure from right-hand prosodic context (see Silverman & Pierrehumbert, 1990 for English,
and Prieto et al., 1995 for Spanish), tonal crowding (Arvaniti, Ladd, & Mennen, 2006; Prieto & Torreira, 2007), and syllable structure (Prieto et al, 2005). In light of these cross-linguistic differences and similarities, the results in the present study suggest that advanced L2 Spanish learners’ prenuclear realizations are moving towards Spanish-like alignment, albeit perhaps not yet reaching the target timing that characterizes native speakers. Findings from the production study in Zárate-Sández (2015), therefore, appear to be confirmed here: low proficiency learners and more advanced learners of Spanish are in an “alignment spectrum” (p. 139). Zárate-Sández, who did include an L1 Spanish comparison group, further concluded that advanced learners where somewhere in the middle between low proficiency learners and native Spanish speakers.

These observed differences in temporal implementation pose great challenges for researchers in terms of their phonological representation. Although differences in tonal alignment have proven to distinguish phonological categories in some languages (e.g., Dilley, 2007 in English), the varying degrees of alignment at the prenuclear position in broad-focus declaratives in Spanish do not trigger semantic distinctions. Zárate-Sández (2015) is the only researcher to have looked into this phenomenon from a perception standpoint for a similar learner population. He found that advanced L2 learners seemed to perceive late peaks differently from less proficient learners as well as from native speakers.

Despite the fact that production data, and even perception data, appear to indicate at least some degree of target-like linguistic development as proficiency increases in the L2, adequately representing a phonological system only through the lens of production and perception data is a daunting task. The challenge of confidently ascertaining the true nature of a phonological unit is particularly salient in the case at hand because the differences in tonal behavior for prenuclear alignment patterns in English and Spanish have been represented by means of two distinct tonal
targets in previous research: Spanish has been described as having either an underlying H* pitch accent or an L*<H pitch accent, whereas English has been said to have an underlying H* pitch accent.

This disagreement over how pitch accents are represented most likely stems from the fact that the very notion of category is not straightforward in intonational phonology. Meanings in intonation are expressed by using both gradient and categorical distinctions on a more regular basis than in other linguistic domains (Gussenhoven, 1999). Furthermore, how speakers and listeners interpret the gradient-categorical relationship in intonation is not yet thoroughly understood. Grice et al. (2017), for instance, found that certain distinctions in meaning were being achieved differently across participants in their study. Some speakers produced displaced peaks to convey a specific meaning (e.g., focus), while other speakers appeared to rely on other strategies (e.g., a higher F0 peak). The authors concluded that more than consistency in how meaning is created across speakers (all speakers use X to express Y), what seemed to matter most was how individual speakers used specific phonetic parameters in a gradient scale to convey specific categorical meaning. In other words, speakers use the available phonetic space in a consistent fashion, modulating parameters (e.g., F0 peak alignment, tonal height) always in the same direction (e.g., later or higher peaks always signal the same meaning). Listeners tune-in to this consistency in phonetic implementation to establish gradient-categorical distinctions at the phonological level.

To further illustrate the challenges in categorizing intonational units, Spanish F0 peak alignment provides a clear example: perception studies in Spanish have indicated that the most typical intonation contour in prenuclear position is frequently perceived as the instantiation of a high tone (H*) rather than a rising one (L*<H), both by native speakers of Spanish (Face &
Prieto, 2007) as well as non-native speakers whose L1 is English (Zárate-Sández, 2015). If we extrapolate this disagreement in perception to how we as researchers interpret production data, there may be significant discrepancies in our interpretations and conclusions. For instance, the discrepancy in prenuclear pitch accents in the Spanish and English inventories could cause us to ponder whether the late peaks observed in L2 Spanish advanced learners should be interpreted as the creation of a new tonal category (L*<H) that did not form part of their L1 inventory, or whether these learners have instead made phonetic adjustments to the realization of an intonational category that was already available for transfer from their L1 system (H*). These two interpretations are so disparate that they cannot be simultaneously veridical.

Cross-linguistic differences in tonal alignment can be examined within the framework of the L2 Intonation Learning Theory from various linguistic dimensions. The LILt offers a series of dimensions that can be evaluated in terms of transfer and is therefore compatible with frameworks of phonological transfer like the Speech Learning Model (Flege, 1995). Before the LILt was proposed, it was unclear in what respect intonational features could be similar or dissimilar cross-linguistically, and thus easier or more difficult to acquire. The inability to analyze intonational units from various perspectives left no option for intonation researchers but to conceive of pitch accents or boundary tones as indivisible. In other words, pitch accents and boundary tones were analyzed unidimensionally. For instance, if both English and Spanish were to have an underlying H* pitch accent in prenuclear position, researchers could only conclude that this was a case of cross-linguistic similarity. The alternative interpretation, Spanish having an underlying L*<H and English an H*, would in turn be interpreted as a case of cross-linguistic dissimilarity.
The model proposed by Mennen (2015) allows for more nuanced cross-linguistic distinctions to be made, enabling a reinterpretation of the analyses in the previous paragraph. In the case of a shared cross-linguistic category (H*), for instance, we can now analyze this tone from two perspectives, that is, two dimensions: we can observe both its behavior in the realizational dimension (the phonetic mapping of the tone and the actual timing of its alignment) as well as how it functions within the systemic dimension (its tonal composition, its distribution, etc.). Given that learners’ deviances in the target form are related to alignment (realizational dimension), we must conclude that despite being similar at the systemic level, English and Spanish differ at the realizational level.

The previous analysis was based on the assumption that there is only one phonological category (H*) shared by both languages. There are two logical and likely consequences of this similarity at the systemic dimension and the dissimilarity at the realizational dimension. Flege (1995, 2007) theorizes that when a learner perceives an acoustic distinction between an L1 sound and an L2 sound, the linguistic system reacts in one of two ways: it either assimilates or dissimilates the L2 sound into or out of the L1 category.

If both languages have the same underlying category, it would not be unlikely for the system to incur phonetic assimilation: learners, even if they are able to acoustically differentiate relatively earlier and later peaks, would end up merging both realizations into the same underlying intonational category. If, or when, they begin to perceive timing differences in the input, learners would treat such variation as instances of allotonic variation. As proficiency increases, though, it is possible that more advanced learners begin to perceive timing differences in tone alignment (Zárate-Sández, 2015) and consequently begin to also phonetically implement these differences. Although this seems like a plausible explanation, one would expect this
realizational difference to be more resistant to acquisition, particularly because English and Spanish would be cross-linguistically similar at the systemic level, and the phonetic difference would be occupying the same phonological space.

The most common and theoretically sound phonological representation of Spanish prenuclear pitch accents, however, is not H*, but rather L*<H (Hualde & Prieto, 2015). Within the LILt, then, English H* and Spanish L*<H would incur a difference at the systemic dimension, which naturally results in a difference at the realizational dimension as well. Mennen (2015) states that in intonation, meaning (semantic dimension) is a crucial factor in predicting relative ease of difficulty. As it pertains to meaning at this position (prenuclear), in neither language do broad-focus declaratives make semantic distinctions, which could therefore add an extra level of complexity for learners. A cross-linguistic analysis of English and Spanish intonation at prenuclear position can be summarized below in Figure 31.

<table>
<thead>
<tr>
<th>Prenuclear Peak Alignment</th>
<th>Cross-linguistic Comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>Systemic dimension (different)</td>
<td>L*&lt;H in Spanish vs. H* in English</td>
</tr>
<tr>
<td>Realizational dimension (different)</td>
<td>Delayed F0 peak in Spanish vs. aligned F0 peak in English</td>
</tr>
<tr>
<td>Semantic dimension (not involved in contrast)</td>
<td>No phonological contrasts in this position</td>
</tr>
</tbody>
</table>

*Figure 31.* A cross-linguistic analysis of English and Spanish intonation at prenuclear position. In the left column, I present the linguistic dimensions most likely affected by prenuclear peak alignment, and whether or not there is cross-linguistic similarity or dissimilarity. In the right column, I present the specific features in each language.
If Spanish and English differ at the systemic level, the deviances in learners’ production (and perception) would result from two perceptually similar contours that are, however, phonetic mappings of two distinct phonological categories at the representational level. One possible prediction of the SLM for similar yet distinct sounds [contours], one which does not exist in the L1 at this position (English does not have L*<H in prenuclear position), is category merger, at least in early stages of development. That is, the instantiations of L*<H would be interpreted as allotones of an existing L1 tonal category (H*). This could explain why beginning and intermediate learners realize early peaks when they are speaking Spanish: input delivers a contour that they perceive as similar to or even exactly the same as their L1’s for that position, therefore, they realize prenuclear contours in an L1 fashion. As proficiency increases, however, learners will have listened to more instances of the Spanish contour and will have begun to form a new category by means of category dissimilation. That is, learners of more advanced proficiency will differentiate the two contours both acoustically and representationally, and in doing so create a new intonational category (L*<H) for their L2.

As mentioned before, within the SLM, phonetic similarity creates adverse conditions for acquisition to occur, while dissimilarity creates linguistic conditions that are more amenable to acquisition. The results for the variable peak alignment in this dissertation appear to back this prediction, at least to a certain extent: peak alignment is acquired by learners who have achieved an advanced proficiency level in Spanish, and who can perhaps discern at least some of the realizational differences between the L1 and the L2 alignment patterns. Since the learners in this group had never been immersed in a Spanish-speaking country, it can be argued that their contact with the L2 was rather limited. In spite of this, they were able to acquire late alignment patterns. This level of attainment, albeit perhaps not completely native-like, could mean that the advanced
learners created a new tonal category for the L2. One question remains unanswered, though: why this parameter is not even easier to acquire (i.e., at even lower proficiency levels of Spanish). The answer may be that there are other factors playing a role: the semantic dimension is not affected at all by this realizational difference, and peak displacement, although possible in English in some positions, is very infrequent (Dainora, 2002). Future studies should further test this interpretation while more effectively controlling for the effects that the semantic and frequency dimensions may have.

It appears that the SLM could explain the results obtained in this dissertation for the variable F0 peak alignment, at least to some degree. Given the SLM’s (Flege, 1995) second hypothesis: “A new phonetic category can be established for an L2 sound that differs phonetically from the closest L1 sound if bilinguals discern at least some of the phonetic differences between the L1 and the L2 sound” (p. 239), slight adjustments to this hypothesis based on the LILt would suffice to adapt it to intonation: A new category (a difference at the systemic level) can be established if learners can discern at least some of the phonetic difference between the L1 and the L2 contour. Following Mennen (2015), the semantic dimension is a more relevant factor for intonation than for segments. Therefore, another hypothesis that could be added for intonation would be: A systemic distinction can be more easily established if the discernible phonetic difference between the L1 and the L2 contour affects the semantic dimension. The fact that the early/late distinction in the comparison English/Spanish does not affect the semantic dimension would probably explain why late peak alignment in Spanish is so difficult to acquire for L2 Spanish learners (L1 English).

Finally, as additional support to this slight modification to the SLM hypothesis, a further look at the data from a more individualized perspective reveals that some of the advanced
learners continued to realize early peaks. Because the contrast L<H*-H* in prenuclear position in broad-focus declaratives does not affect meaning (semantic dimension), it is possible that some learners may never fully establish a new L2 category, and therefore, never perceive or produce late F0 peaks in Spanish. That is, “[c]ategory formation for an L2 sound may be blocked by the mechanism of equivalence classification. When this happens, a single phonetic category will be used to process perceptually linked L1 and 12 sounds (diaphones)[diatones].” (Flege, 1995: p. 239).

**Final Boundary Tone.** The results in this dissertation indicate that there are no statistically significant proficiency-related differences in how learners phonetically map final boundary tones in L2 Spanish. In the study abroad group, no changes were found from pre-test to post-test, either. Finding no group differences in phrase-final intonation in L2 Spanish (L1 English), however, is not an unexpected outcome. As was stated in Chapter 2, Spanish and English are believed to use the same underlying boundary tone in this position (Estebas-Vilaplana & Prieto, 2008; Beckman & Pierrehumbert, 1986): L%. The results in this dissertation might seem somewhat at odds with Zárate-Sández’s (2015) findings. Zárate-Sández, unlike me, found an inverse relation between tonal height and proficiency: the more advanced a learner is, the lower her phonetic instantiations of L% are.

Previous researchers of phrase-final intonation in L2 Spanish (e.g., Henriksen et al., 2010, Thornberry, 2014) have found that High Rising Terminal (HRT) was a pervasive phenomenon, particularly in lower proficiency learners. Henriksen et al. indicated nonetheless that the rate of final high rises appears to diminish as proficiency increased. This tendency was explained as a transparent instance of transfer from the L1 (English), a language in which HRT is an extended occurrence (Warren, 2016). In this dissertation, however, there were no clear
patterns of HRT in the L2 data that could be unquestionably justified as greater or lesser degrees of transfer. In other words, it was not the case that beginning learners were more susceptible to realizing final rises than more advanced learners, or that learners in the SA group were more prone to this type of final contour at the program’s onset, but less likely by the program’s end. Although results in this dissertation show that final rising intonation does occur, albeit infrequently, this phenomenon was not circumscribed to the interlanguage of any particular proficiency group. A more accurate conclusion to draw from the data would be that HRT is used in L2 Spanish regardless of proficiency level.

Discussions regarding HRT in American English, particularly in young people from Southern California and Massachusetts, are frequent in both academic contexts (Armstrong, Piccinini, & Ritchart, 2015; Fletcher, Grabe, & Warren, 2005; Ritchart & Arvaniti, 2014) as well as in popular culture (e.g., news outlets posited that Elizabeth Holmes, the infamous CEO of Theranos, avoided using final rising intonation to sound more interesting). An overlooked element in this discussion of final rising intonation in broad-focus declaratives, however, is that this phenomenon is not exclusive to English, as it has also been reported to occur in Spanish in the United States (Vergara, 2015). This similarity in phrase-final intonation of broad-focus declaratives in English and Spanish is actually expected, given the bidirectional nature of language contact. If rising intonation in declaratives has extended from English to also become a characteristic of Spanish in the United States, and is hence possibly used by instructors in Spanish classrooms across the country, we might want to reconsider the constant showcasing of HRT as an example of transfer, and instead present it as an example of language contact. As Flege (1995) points out: “Phonetic categories established in childhood for L1 sounds evolve over
the life span to reflect the properties of all L1 or L2 phones identified as a realization of each category” (p. 239).

Despite minor differences in the interpretations of final intonation contours in L2 Spanish, particularly as it pertains to rising intonation, there is overwhelming consensus in the field that learners of all proficiencies map their underlying L% boundary tones onto relatively low frequencies of the tonal space (Méndez Seijas, 2018; Thornberry, 2014; Trimble, 2013a, 2013b, Zárate-Sández, 2015). This mapping should be further analyzed from two perspectives, a phonological perspective and a phonetic perspective.

From a phonological perspective, there is arguably one underlying tonal category that is similar for both languages: L%. From this point of view, instances of HRT ought to be interpreted as allotones of L%. To my knowledge, phrase-final rising contours in broad-focus declaratives in English (or Spanish) have not been interpreted as instantiations of a different underlying boundary tone. As explained in Chapter 2, within the framework of the AM model, the phonological space is most commonly divided into two (L and H) tonal categories or levels. Differences in phonetic mappings within the confines of these levels are regarded as phonetic, and hence irrelevant for phonological analyses. The intra-level variations, referred to as variations in scaling, have been an area in which the AM model and the notation system ToBI have struggled the most (see Dilley & Breen, 2018 for a discussion). The shortcomings of the AM model to interpret scaling phenomena are partly due to a) the fact that the linguistic functions of scaling are not yet fully understood, and also b) the model’s inability to more appropriately describe syntagmatic aspects of intonation. Scaling, a phenomenon initially considered to be purely phonetic (Ladd, 1999), is not so clearly seen as a gradient feature of intonational phonology anymore, at least not in all languages. For instance, recent research
suggests that some languages use scaling modifications to elicit phonological contrasts (see Méndez Seijas, 2010 for Venezuelan Spanish; Savino & Grice, 2011 for Bari Italian; Vanrell, 2006, 2011 for Majorcan Catalan), that is, categorical distinctions.

From a phonetic perspective, it has been reported that the mapping of Spanish L% occurs at relatively lower frequencies when implemented by Spanish native speakers than by L2 Spanish learners, a difference that becomes more acute the lower a learner’s proficiency level is (Zárate-Sández, 2015). As it pertains to the learner groups examined in this dissertation, and contrary to previous findings, there was no inverse relationship between proficiency and final F0 height. To further corroborate that participants did not change their phrase-final intonation in L2 Spanish, a comparison between L1 English and L2 Spanish detailed in the Pitch Range section confirmed that there were indeed no cross-linguistic height differences whatsoever.

Through the lens of the LILt, the initial cross-linguistic analysis of final boundary tones would be that Spanish and English are similar at the systemic level (both have an underlying L boundary tone), but different realizationally in how this shared underlying tone is phonetically mapped onto the segmental string: bilingual Spanish native speakers implement final boundary tones at lower frequencies than L2 Spanish learners do (L1 English). A cross-linguistic analysis of English and Spanish intonation at phrase-final position is summarized in Figure 32.
<table>
<thead>
<tr>
<th>Final Tonal Boundary Tone</th>
<th>Cross/linguistic Comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>Systemic dimension (similar)</td>
<td>Same underlying boundary tone: L%</td>
</tr>
<tr>
<td>Realizational dimension (different)</td>
<td>Higher phonetic mapping of the boundary tone by L1 Spanish speakers than L2 Spanish learners of any proficiency</td>
</tr>
<tr>
<td>Semantic dimension (similar)</td>
<td>Same meaning associated with phase-final low intonation</td>
</tr>
</tbody>
</table>

*Figure 32.* Cross-linguistic analysis of English and Spanish intonation at phrase-final position.

In the left column, I present the linguistic dimensions most likely affected by the differences in the scaling of final boundary tones, and whether or not there is cross-linguistic similarity or dissimilarity. In the right column, I present the specific features in each language.

In this scenario, the SLM’s postulate “bilinguals strive to maintain contrasts between L1 and L2 phonetic categories, which exist in a common phonological space” (Flege, 1995, p. 239) might look like a reasonable starting point to interpret the results. The language in this postulate refers to phonetic distinctions, that is, realizational distinctions. This phonetic perspective is clear when the author states that “sounds are related to one another at a position-sensitive allophonic level, rather than at a more abstract phonemic level” (p. 239). However, we could also wonder if this cross-linguistic difference also pertains, to some extent, to the systemic dimension.

A traditional AM analysis of phrase-final boundary tones for the current cross-linguistic comparison would force us to assume that the difference between English and Spanish is in the degree of scaling, and hence, a matter pertaining to the phonetics of intonation. Having only two pitch levels dramatically limits the number of available phonological contrasts, thus restricting other possible avenues of formal analysis. A valid question here would be if the challenge is not necessarily that the SLM is insufficiently equipped to explain phonological contrasts discernable
in our data, but rather if the current capabilities of the AM framework and ToBI are constraining our capacity to interpret these data.

A recently proposed version of the AM model might provide an opportune alternative to how we interpret differences in scaling, among other phenomena: The enhanced Autosegmental-Metrical (AM plus, or simply AM+) theory and the Rhythm and Pitch (RaP) Prosodic System (Dilley & Breen, 2018). The AM+ looks to improve the traditional AM framework by strengthening the way in which paradigmatic and syntagmatic aspects of intonation are represented. In the AM+, phonological representations are based on the syntagmatic features [+/- same], [+/higher], which provide very specific realizational information: T [-same, +high], for instance, would be a tone that is different from and higher than the preceding tone. They also propose a feature to codify tonal distance [+/- small]. For instance, in the context of two adjacent tones Hn/ Tn-1, the first of which is a high tone, a choice for the second tone between Tn-1 [-same, -high, +small] and Tn-1 [-same, -high, -small] would generate a slightly lower realization of L in the first case, and a substantially lower realization of L in the second. The possible permutations of three levels (high, medium, and low), along with the features [+/-same] [+/-high] [+/-small] create a phonological system of five levels, instead of the two of the traditional AM model. Table 13 below shows all the possible tone-features combinations that would result in fine levels:
Table 13. Final level “paradigmatic” tone specifications derived from a set of syntagmatic features (from Dilley & Breen, 2018, with a minor correction).

<table>
<thead>
<tr>
<th>Context</th>
<th>Lexical phonological specification</th>
<th>Phonetic interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>T(_{\text{EH}})/r</td>
<td>[-same, +higher, -small]</td>
<td>Substantially higher than the mean pitch</td>
</tr>
<tr>
<td>T(_{\text{H}})/r</td>
<td>[-same, +higher, +small]</td>
<td>Slightly higher than the mean pitch</td>
</tr>
<tr>
<td>T(_{\text{M}})/r</td>
<td>[+same]</td>
<td>Equal to the mean pitch</td>
</tr>
<tr>
<td>T(_{\text{L}})/r</td>
<td>[-same, -higher, +small]</td>
<td>Slightly lower than the mean pitch</td>
</tr>
<tr>
<td>T(_{\text{EL}})/r</td>
<td>[-same, -higher, -small]</td>
<td>Substantially lower than the mean pitch</td>
</tr>
</tbody>
</table>

It could be argued, then, that the tonal syntagmatic relationships at phrase-final position in these two languages are different, causing the phonetic mapping of L\% to be lower in Spanish than in English. That is, the final boundary tone in these languages would be post-lexically indexed in two distinct language-specific ways: T\(_{n}/L_{n-1}\) [+small] in English vs. T\(_{n}/L_{n-1}\) [-small] in Spanish. If we interpret these two differently indexed tones as equivalent to two independent “categories”, or at least equivalent to two different post-lexical phonological rules, we could then say that English and Spanish use two different tones in the same phonological space. If no L2 contrasts happen within the confines of this shared phonological space, then the semantic dimension would be unaffected. In this context, I would again concur with Flege (1995) in that “[c]ategory formation for an L2 sound may be blocked by the mechanism of equivalence

---

classification. When this happens, a single phonetic category will be used to process perceptually linked L1 and L2 sounds (diaphones)[diatones]” (p. 239).

One practical consequence of equivalence classification is that the cross-linguistic difference in how phrase-final Ls are mapped would most likely not be perceived by learners at lower levels of proficiency. Rather, only time and substantial amounts of input would allow for long-term memory representations to form a new, target-like tonal category. This interpretation is consistent with results from previous research in L2 Spanish. This interpretation, for instance, would explain why Zárate-Sández (2015) found an inverse relation between level of Spanish proficiency and low(er) phonetic mappings of L%.

The LILt, within the limits of the more traditional AM approach that I have adopted for this dissertation, would lead to a slightly different conclusion: the cross-linguistic difference would only be realizational. Couched within the LILt, and based on the results of this dissertation, it appears as though it is more difficult to acquire a realizational L2 aspect when the L1 and the L2 are similar at the systemic dimension. This interpretation would explain why the results in this dissertation showed no development whatsoever towards phrase-final target tonal behavior, even at advanced levels of proficiency. It would also explain how the acquisition of final boundary tones is different from and more difficult than F0 peak alignment.

These results suggest that a difference that only transpires at the realizational level is perhaps more difficult to be discerned and acquired than one that affects both the systemic and realizational dimensions. This is even more difficult when the realizational difference does not affect the semantic dimension. For this reason, acquiring target-like height in phrase-final intonation should be (and results indicate that is) a more arduous task for L2 Spanish learners who are native speakers of English.
**Pitch Range.** Pitch range has generally been treated as a global intonational parameter and has mostly been reported as such. For instance, Hanley et al. (1966) stated that Japanese and Latin American Spanish have an overall wider pitch range than American English, and Eady (1982) found that Mandarin and Taiwanese have similar pitch ranges. L2 intonation has also caught the attention of researchers looking into pitch range phenomena. Braun (1994), for instance, reported that L2 German speakers of L1 Turkish had a wider pitch range in German than did German native speakers. For Spanish as an L2, Trimble (2013a) reported that L2 Spanish speakers (L1 English) widened their pitch range after spending a semester abroad in Venezuela. Trimble confirmed, albeit indirectly, Hanley et al.’s (1966) observation about the cross-linguistic difference between Spanish and English.

Mennen et al. (2014), however, found that cross-linguistic differences in pitch range are not always global, as they can also occur locally. That is, following Flege’s (1995) hypothesis that L1 and L2 sounds are related at a position-sensitive level, Mennen and colleagues studied pitch level and pitch span at different points along the F0 contour. They found that L2 English-L1 German speakers presented significant differences in tonal distance for the following three pitch span measures: a) Initial Prominent Peaks Minus Final Lows, b) Noninitial Prominent Peaks Minus Lows, and c) Initial Prominent Peaks Minus Noninitial Prominent Peaks. Their results for absolute landmarks showed no significant group differences in pitch level for Initial Prominent Peaks, Noninitial Prominent Peaks, and Final Lows.

As presented in the results section, advanced learners of Spanish in the classroom context evidenced some pitch range differentiation between their L1 (English) and L2, both in terms of pitch level as well as pitch span. With regards to pitch level, the statistical analyses\(^7\) that were

---

\(^7\) Please note that a Bonferroni adjustment was conducted, and the new alpha was set at \(p = .005\).
performed indicated that a) Noninitial Peaks were statistically different ($p = .002$) and had a small effect size ($r = -0.28$), b) Low landmarks approached significance ($p = .014$) and had a small effect size ($r = -0.38$), and c) all the other landmarks were neither statistical nor had any effect sizes. The implementation of a Bonferroni correction may be a factor as to why Low landmarks did not reach significance; therefore, the small effect size might be a better indicator that these landmarks underwent at least some degree of change. In more tangible terms, the results reported above indicate that Noninitial Peak landmarks were realized at lower frequencies in L2 Spanish than in L1 English, and Low landmarks that moved in the opposite direction, to higher frequencies in the L2. Figure 33 below depicts how valleys were realized at a higher F0 range, and Peaks were instead realized at a lower F0 range.

Figure 33. Direction of the difference in pitch range between English and Spanish. The red dots represent the valleys and the blue dots represent the noninitial peaks. The arrows signal the direction of the difference between L1 English and L2 Spanish.

In terms of pitch span, I followed Mennen et al.’s (2012, 2014) design by implementing a series of linguistic measures to calculate the tonal distance between specific landmarks along the F0 contour. The statistical analyses indicated that there was only one significant difference in pitch range between speakers’ L2 Spanish and L1 English broad-focus declaratives: the tonal
distance between Noninitial Prominent Peaks and Lows (H*-L) was statistically different ($p = .002$), and evidenced a medium effect size ($r = -.47$). As a point of comparison, it is relevant to note that none of the other linguistic measures were statistical or evidenced even a small effect size. The statistical significance and medium effect size of the formula H*-L allow for a reliable interpretation of the results as they pertain to this tonal distance measure: broad-focus declaratives in L2 Spanish have a narrower pitch span than in L1 English. This difference is not global; rather, it appears to be position-sensitive. The observed narrowing of pitch range in L2 Spanish did not affect any measures that involved Initial Level, Initial Prominent Peak, or Final Level. As I had already indicated, landmarks realized at these last three positions within IPs were mapped at similar $f_0$ heights regardless of the language. These results of pitch level and pitch span might constitute the first time that this position-sensitive nature of pitch range has been observed. Figure 34 below depicts the difference in range between English and Spanish.

![Figure 34](image)

**Figure 34.** Difference in pitch range between L1 English and L2 Spanish.

The red line represents L2 Spanish, and the blue line represents L1 English. The red line traces a narrower pitch range than English in the middle of the intonational phrase.

Regarding level and span measures where no differences were detected, no further conclusions can be drawn from the current data. To better understand pitch range phenomena,
particularly in those undifferentiated positions or measures, future studies will benefit from including a Spanish native speaker comparison group. This, unfortunately, was outside of the purview of the current study. When studying Spanish learners in the United States, the input learners receive from instructors, colleagues or Spanish-speaking friends is representative of not one variety, but rather of many different varieties of Spanish, many of which surely encode intonational meaning in different ways. A comparison group would be more appropriate in study abroad context, for instance, where a greater degree of homogeneity among native speakers would warrant more reliable comparisons between L1 and L2 data. My initial project for this dissertation included this idea, but it had to be discarded for logistical reasons.

Because I did not have a native speaker control group, it remains unclear if there were no discrepancies in the values of some landmarks and measures because they were already target-like to begin with, or because phonetic contrasts in those particular positions are more difficult to develop at lower proficiency levels.

With regards to the findings from previous research detailed in Chapter 2, both generally regarding pitch range phenomena in L2 acquisition, and specifically regarding pitch range in L2 Spanish, some of the results of this dissertation corroborate and some contradict previous research. The points of agreement and disagreement, however, require a nuanced analysis because of the degree of complexity of this topic, as well as because subtleties in the results might force some type of theoretical reanalysis within a model I have adopted.

As it pertains to points of agreement, the results of this dissertation lend credence to at least some of the theoretical assumptions in the SLM (Flege, 1995) and LILt (Mennen, 2015): pitch range appears to be, indeed, a parameter that functions at a position-sensitive level. Because of this, it is possible to explain why advanced learners in the current study made local
modifications to some landmarks in specific positions within the phrase, but not in others. Moreover, as Mennen et al. (2012) have pointed out, corroborating the position-sensitive nature of pitch range tangentially asserts the appropriateness of using linguistic measures (e.g., accentual peaks) rather than long term distributional measures (LTD) (e.g., a speakers’ mean F0) to study this phenomenon. Linguistic measures are also a more logical tool in our case, because landmarks, not LTD measures, are thought to be linked to abstract phonological units (Ladd, 1999; Mennen et al., 2014).

Contrary to some of the findings of previous research in L2 Spanish intonation (Méndez Seijas, 2018; Trimble, 2013a), the data in this dissertation show that, on the one hand, L2 learners in the classroom context narrowed, rather than expanded, their tonal range; and on the other hand, the study abroad group evidenced no changes whatsoever from pre-test to post-test. The apparent discrepancies in these results, however, must be interpreted very carefully. Trimble (2013a), for example, studied a group of learners in a long-term study abroad program. Given the weight that length of stay and quality and quantity of input have on L2 acquisition, it is in fact unsurprising that the learners in the short-term study abroad did not change their pitch range in any observable way. What is more, given the length of stay of the participants in Trimble’s study, his post-test results might be offering a glimpse of very advanced learners that are more than a few steps ahead in the acquisition process than the advanced groups that participated in the studies that I have reported in this dissertation.

As it pertains to Méndez-Seijas (2018), it should be noted that his study used spontaneous speech data from only four learners. His results are, therefore, not supposed to be generalizable, but rather only treated as a sample of possible progress after a short-term study abroad program. It is not unlikely that a more fine-grained analysis of our data would also find one or two
participants whose pitch range approached target-likeness in some parameter. The objective of the studies included in this dissertation, however, was to examine group changes rather than individual changes.

Also, unlike previous research (e.g., Mennen et al., 2014; Trimble, 2013a), the modifications of pitch range reported in this dissertation moved away, not towards, native-like norms. Trimble (2013a), for instance, found that his participants widened their global pitch range so as to approximate target-like levels of Venezuela Spanish. A similar conclusion was reached by Mennen et al. (2014), whose participants implemented changes in order to produce native-like intonation. The narrowing of the pitch range observed in my data would be in line with results from Kelm (2015), who suggested that regardless of cross-linguistic differences, learners first narrow their pitch range. This observation by Kern appears to be indirectly confirmed by Backman (1972), whose L1 Spanish participants’ pitch range in L2 English was narrower than that of L1 English speakers. Perhaps the first step in the development of pitch range in an L2 is to narrow it regardless of L1-L2 differences and, only at later stages, make modifications so as to approach target-like forms. This reported narrowing of pitch range could hardly be interpreted as a case of transfer.

Within the LILt, pitch range phenomena would be treated as a purely realizational parameter. Being a more global parameter, it would be different from peak alignment and final boundary tones. The results of pitch range reinforce our hypothesis that realizational differences that do not affect the systemic or semantic dimensions appear to be very resistant to acquisition, at least in relatively early stages of development. Some preliminary confirmation of this can be found not only in this dissertation, but also in Mennen et al. (2014) and Trimble (2013a), whose advanced participants improved significantly after spending a significant amount of time abroad.
After prolonged stays abroad, the LILt and SLM could predict that learners begin to implement changes, first probably at a global level and then at a more local level (Mennen et al., 2014). The data in this dissertation, which for this research question only focused on advanced learners, unfortunately cannot be used to determine if global changes precede local changes. Further experimental endeavors will have to answer this question.

All in all, it seems like the LILt, in conjunction with (but not limited to) a model like the SLM, offers a theoretical base on which to make solid interpretations, hypotheses, and predictions of some cross-linguistic differences in pitch range. However, if more studies find a similar narrowing of pitch range in learners of lower proficiencies, a tonal behavior that in the case of L2 Spanish-L1 English cannot be explained through transfer, it might be necessary to look elsewhere for a framework that allows, for instance, for universal constraints to also play a role in the acquisition of phonetic and phonological parameters of intonation.

It would not be the first time in the history of the field of SLA that transfer by itself cannot explain everything, as early models such as the Contrastive Analysis (Lado, 1957) soon discovered. Since then, a debate still exists about the extent to which other aspects (e.g., markedness, social factors, Universal Grammar, transfer) affect L2 acquisition, and also if they operate simultaneously.

Within models of generative grammar, for instance, it is assumed by some that the initial state in L1 acquisition is Universal Grammar (UG) unaltered by the linguistic environment. In L2 acquisition, there is no consensus as to what the weight of UG is, or even if it has any influence on the L2 system at all. The debate pertaining to UG in SLA has revolved around the issue of “accessibility” to UG. Some of the theoretical models developed throughout the years have supported full access to UG, but no transfer whatsoever (Full Access Hypothesis: Epstein,
Flynn, & Martohardjono, 1996, 1998), or have assumed that both transfer and UG are involved (Full Access Full Transfer Hypothesis: Swartz & Sprouse, 1994, 1996; see White, 1989, for a thorough discussion). Still, the most widely accepted opinion is that both transfer and UG may play a role in L2 acquisition (Major, 2008). Pitch range may be just another instance in which we realize the need for models that explain transfer without disregarding the role of other aspects (e.g., cognitive, social, environmental, UG) that are also relevant in SLA research. One such model that makes all these connections possible, called Dynamic Systems Theory (De Bot, Lowie, & Verspoor, 2007) is (re-)gaining traction in the field of SLA, and recent research on prosody and intonation has begun to be couched within it (e.g., Mücke, 2018; Roessig, Mücke, & Grice, 2019). Application of the Dynamic Systems Theory to L2 Spanish intonation data, however, will have to wait until another graduate student writes her dissertation.

Limitations and Future Research

All research is subject to limitations and this dissertation is no exception. One limitation is the lack of comparison groups for all variables under study. In this dissertation there was no L1 Spanish comparison group, and the L1 English comparison group was implemented only in the classroom context for pitch range phenomena. Without comparison groups, I had to rely instead on results from previous studies, many of which used very different experimental designs or had far smaller participant pools. Future researchers should include both L1 Spanish and L1 English comparison groups. For long-term study abroad research, it would be particularly useful to include an L1 Spanish comparison groups, because after being exposed mostly to one variety of Spanish, it might be easier, for instance, to more clearly discern changes in pitch level and pitch span.
Another limitation of the current study is the small amount of data that I was able to use. In other words, I based my conclusions on only a handful sentences per learner. Acknowledging that variability is prevalent in learners’ interlanguage, future research would benefit from more targeted approaches than mine. For instance, it might be more helpful to study just one variable (e.g., peak alignment) and collect and analyze much more data per learner to better understand its acquisition. This more targeted approach would also make it easier to implement and compare production and perception data, thus being able to see phenomena from two different but complementary perspectives.

Despite its shortcomings, this dissertation constitutes the first approximation to the study of L2 Spanish intonation using a theoretically-grounded model designed exclusively to analyze intonation phenomena: The L2 Intonational Learning Theory. As this dissertation represents the first large-scale foray into this area, there was oftentimes no similar previous research to rely on and learn from, and for this reason my conclusions had to be somewhat speculative at times. Future research should build upon this effort I have made, implement specific changes to improve the design, and then further test if the areas of difficulty that I have identified hold true for the same and other intonational phenomena, in Spanish and in other languages.

Within the LILt, I focused primarily on two (out of four) dimensions where cross-linguistic interactions can be identified: the realizational dimension and the systemic dimensions. Future research would benefit from also exploring deviances as a consequence of transfer in other dimensions. For instance, it would be enriching to the field of L2 intonation to better understand how deviances in the semantic dimension can help or hinder the development of new intonation categories. Also, the role of frequency is largely unexplored in intonation research in general, let alone in L2 intonation.
Finally, I based most of my analysis on the SLM, a model of transfer. As was evident in the discussion, transfer does not seem to explain all of the results of pitch range phenomena in this dissertation. This should serve as a cautionary tale: we need to explore other models that better integrate transfer and other factors, be they cognitive, social or even theoretical. An interesting model, the Dynamic Systems Theory, appears to offer enough flexibility as to allow different theoretical models to coexist within its domain. One benefit of the LILt, which makes it compatible with almost any model, is that it offers a series of parameters on which to base comparisons without being too rigid as to what phonological theories of acquisition can be used to explain the data.

Conclusions

In this section, I explore one by one the research questions that guided the studies within this dissertation. Below is a brief summary of the main findings for each research question.

**Study 1. Second Language Phonological Acquisition in Classroom-based Context.**

1. Prenuclear pitch alignment: Does the realization of Spanish prenuclear peak alignment change as L2 learners advance in proficiency?

The results indicated that F0 peak alignment does change as L2 learners advance in proficiency. Learners of beginning and intermediate proficiency realized early peaks that are more characteristic of the L1 (English), whereas advanced learners realized peaks statistically later, albeit still within the confines of the first posttonic syllable. These results appear to indicate that L2-like peak alignment timing is a parameter only acquired when a learner has achieved advanced proficiency.

2. Sentence-final boundary tones: Does the scaling of final boundary tones change as L2 learners advance in proficiency?
The results indicated that there was no statistical difference between groups as it pertains to the scaling of final boundary tones. Therefore, at least with regards to the range of proficiency levels included in this dissertation, there was no evidence of acquisition that could be explained as a result of greater language proficiency.

3. Pitch range: Do advanced L2 learners of Spanish evidence differences in pitch range between L2 Spanish and L1 English?

The results indicated that there were differences in pitch level and pitch span between speakers of L1 English and L2 Spanish. The differences in pitch level affected noninitial prominent peaks and lows. No changes in F0 level were evidenced at the beginning or end of intonational phrases. In terms of span, there was a significant difference in the pitch span measured by the formula noninitial prominent peaks minus lows. This means that advanced Spanish learners narrowed the pitch range of most utterance-internal tones, leaving phrasal edges at L1 levels.

**Study 2. Second Language Phonological Acquisition in Study Abroad.**

1. Prenuclear pitch alignment: Do advanced L2 learners of Spanish evidence changes in the realization of prenuclear peak alignment after a short-term 5-week study abroad program?

The results indicated that there was no statistical difference between the pre-test and post-test as it pertains to the alignment of prenuclear pitch accents. That is, the same degree of alignment was reported before and after the 5-week short-term study abroad program.

2. Sentence-final boundary tones: Do advanced L2 learners of Spanish evidence changes in the realization of final boundary tones after a short-term 5-week study abroad program?
The results indicated that there was no statistical difference between the pre-test and post-test as it pertains to the scaling of final boundary tones. That is, the scaling of final boundary tones remained at the same F0 height after the 5-week short-term study abroad program.

3. Pitch range: Do L2 advanced learners of Spanish evidence changes in pitch range after a short-term 5-week study abroad program?

The results indicated that there was no statistical difference between the pre-test and post-test as it pertains to pitch level and pitch span measures. That is, learners’ pitch range did not change after a 5-week short-term study abroad program.
Appendix A: Background Questionnaire

Background Information

1. Study ID Number: 2017-0716

2. Date of birth:

3. Sex:

4. Please indicate your year in school
   (please underline your answer)
   
   1. Freshmen
   2. Sophomore
   3. Junior
   4. Senior
   5. Graduate Student (Masters level)
   6. Doctorate

5. Do you have normal or corrected to normal vision or hearing? (please underline your answer)

   Yes   No

   If not, please explain.

Language background and experience

1. What is your native or first language? (The language you first spoke) If there is more than one, please indicate all languages you consider to be your native language.
2. Have you taken linguistics or language analysis courses in high school or college?

3. If so, please list the courses below.

4. Are you a Linguistics Major or Minor? (If yes, please specific)

5. Do you know any other languages? Please indicate with an X which languages you know, what your current proficiency level is, and in what context you studied this language in the spaces below.

Language:

<table>
<thead>
<tr>
<th>Level</th>
<th>Context</th>
<th>(months studied)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poor</td>
<td>Language school</td>
<td></td>
</tr>
<tr>
<td>Basic</td>
<td>Elementary school</td>
<td></td>
</tr>
<tr>
<td>Intermediate</td>
<td>Middle school</td>
<td></td>
</tr>
<tr>
<td>Advanced</td>
<td>High school</td>
<td></td>
</tr>
<tr>
<td>Fluent</td>
<td>College</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Private tutor</td>
<td></td>
</tr>
<tr>
<td></td>
<td>with family</td>
<td></td>
</tr>
<tr>
<td></td>
<td>computer software</td>
<td></td>
</tr>
<tr>
<td></td>
<td>other</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(please describe)</td>
<td></td>
</tr>
</tbody>
</table>
### Language

<table>
<thead>
<tr>
<th>Level</th>
<th>Context</th>
<th>(months studied)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poor</td>
<td>Language school</td>
<td></td>
</tr>
<tr>
<td>Basic</td>
<td>Elementary school</td>
<td></td>
</tr>
<tr>
<td>Intermediate</td>
<td>Middle school</td>
<td></td>
</tr>
<tr>
<td>Advanced</td>
<td>High school</td>
<td></td>
</tr>
<tr>
<td>Fluent</td>
<td>College</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Private tutor</td>
<td></td>
</tr>
<tr>
<td></td>
<td>with family</td>
<td></td>
</tr>
<tr>
<td></td>
<td>computer software</td>
<td></td>
</tr>
<tr>
<td></td>
<td>other</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(please describe)</td>
<td></td>
</tr>
</tbody>
</table>

### Language

<table>
<thead>
<tr>
<th>Level</th>
<th>Context</th>
<th>(months studied)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poor</td>
<td>Language school</td>
<td></td>
</tr>
<tr>
<td>Basic</td>
<td>Elementary school</td>
<td></td>
</tr>
<tr>
<td>Intermediate</td>
<td>Middle school</td>
<td></td>
</tr>
<tr>
<td>Advanced</td>
<td>High school</td>
<td></td>
</tr>
<tr>
<td>Fluent</td>
<td>College</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Private tutor</td>
<td></td>
</tr>
<tr>
<td></td>
<td>with family</td>
<td></td>
</tr>
<tr>
<td></td>
<td>computer software</td>
<td></td>
</tr>
<tr>
<td></td>
<td>other</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(please describe)</td>
<td></td>
</tr>
</tbody>
</table>
Appendix B: Target Declarative Sentences in Spanish (Peak Alignment)

<table>
<thead>
<tr>
<th>Target Sentences</th>
<th>Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manolo me manda flores</td>
<td>Manolo sends me flowers</td>
</tr>
<tr>
<td>Él bebe la leche</td>
<td>He drinks the milk</td>
</tr>
<tr>
<td>Mi hermana camina al trabajo</td>
<td>My sister walks to work</td>
</tr>
<tr>
<td>El pelo castaño me gusta</td>
<td>I like brown hair</td>
</tr>
</tbody>
</table>
## Appendix C: Target Declarative Sentences in Spanish (Final Boundary Tones and Pitch Range)

<table>
<thead>
<tr>
<th>Target Sentences</th>
<th>Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>El clima era muy cálido</td>
<td>The weather was very warm</td>
</tr>
<tr>
<td>Manuela tomó una limonada</td>
<td>Manuela drank a lemonade</td>
</tr>
<tr>
<td>Fueron a ver a una amiga</td>
<td>(They) went to see a friend</td>
</tr>
<tr>
<td>Manuela leyó una novela</td>
<td>Manuela read a novel</td>
</tr>
<tr>
<td>Comieron algunas bananas</td>
<td>(They) ate some bananas</td>
</tr>
<tr>
<td>Bebieron algo en un bar</td>
<td>(They) drank something at a bar</td>
</tr>
</tbody>
</table>
Appendix D: Target Declarative Sentences in English (Pitch Range)

Target sentences in English

My brother ran eleven miles this morning

My mother made a wonderful dinner

The rain this morning flooded my bedroom

I want to buy new running shoes this summer
References


Colantoni, L. (2011). Broad-focus declarative sentences in Argentine Spanish contact and non-contact varieties. In C. Gabriel and C. Lleó (Eds.), *Intonational phrasing in Romance and


Henriksen, N., Meghan E. Armstrong, & L. García-Amaya. (2016). The intonational meaning of polar questions in Manchego Spanish spontaneous speech. In M. Armstrong,
N. Henriksen, & M. Vanrell (Eds.), *Interdisciplinary approaches to intonational grammar in Ibero-Romance: Approaches across linguistic subfields*, 181-206.


acquisition, attrition, languages in multilingual situations (177-188). Berlin: Springer Verlag.


of Hawaii at Manoa.


Simonet, M. (2009) Nuclear pitch accents in Majorcan Catalan declaratives: Phonology,


Olarrea, & E. O'Rourke (Eds.), *Handbook of Hispanic Linguistics* (pp. 729-746). Oxford:
Wiley-Blackwell.

tones by Cantonese speakers: The role of native phonological and phonetic properties.
Paper presented at the Thirteenth Australasian International Conference on Speech
Science and Technology, Melbourne, Australia.

categories: The role of phonetic properties. *Poznań Studies in Contemporary Linguistics*,
47, 133–145.

dialectología*. Madrid: Cátedra.

language phonetics* (Doctoral Dissertation). Georgetown University, Washington, D.C.


