“Are you asking me or telling me?” Perception and Production of Y/N Questions and Statements in L2 Spanish

by

Olivia Margherita Marasco

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy

Department of Spanish and Portuguese
University of Toronto

© Copyright by Olivia Margherita Marasco 2020

ProQuest Number: 28092106

All rights reserved

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

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

ProQuest 28092106

Published by ProQuest LLC (2020). Copyright of the Dissertation is held by the Author.

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

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

“Are you asking me or telling me?” Perception and Production of Y/N Questions and Statements in L2 Spanish

Olivia Margherita Marasco

Doctor of Philosophy

Department of Spanish and Portuguese
University of Toronto

2020

Abstract

This study focuses on the perception and production of the initial boundary tone and pre-nuclear peak in Y/N questions and statements in L2 Spanish. Studies on different varieties of Spanish have reported that the initial boundary tone and the pre-nuclear peak are noticeably higher in Y/N questions with respect to comparable statements (Prieto & Roseano, 2010; Sosa, 1999). The current study sought to determine whether L2 Spanish speakers could accurately perceive and produce these utterance-initial intonation cues. Furthermore, this study sought to understand the nature of the relationship between L2 perception and production in intonation. Twenty-two advanced L2 Spanish speakers (L1 English) and 9 native Spanish speaking controls (NS) participated in the study, which consisted of two tasks. The perception task was a gating experiment where participants listened to utterances in increments (gates) and were asked to identify what they heard as a question or a statement. Accuracy and reaction times (RT) were recorded. The production task consisted of a narrative that was designed to elicit a Y/N question or a statement. In the narrative, one utterance was missing. Participants were asked to provide the missing utterance. Initial boundary tone and pre-nuclear peak height were measured in Hz and the differences were transformed into semitones. In perception, the findings revealed that there was

no significant difference between the L2 group and the NS group in terms of accuracy or RT. In production, the initial boundary tone findings revealed no difference between Y/N questions and statements for either group. For the pre-nuclear peak, the native speaker group realized a difference that was greater than 1.5 semitones (a difference which is perceptible to the human ear: Nooteboom, 1997) while the L2 group was just below the 1.5 semitone threshold. Finally, in order to measure the relationship between perception and production, correlations were run on perception and production results. None of the correlations were statistically significant. Individual analyses revealed interesting trends. While variability in L2 speakers was expected, the variability in the native speaker data revealed patterns that were unexpected when compared to the literature on Spanish intonation. This finding suggests the possibility of multiple ways to signal the Y/N question-statement distinction in Spanish. Such a possibility opens the way for more research into the inventory of Spanish intonation as well as the relationship between perception and production in L2 intonation.

Acknowledgments

First, I must thank my supervisor, Prof. Laura Colantoni. I am grateful for your insight, support, patience, sense of humour, and availability over the years. Even as Chair of the department, you always made time to see me and discuss my thesis. My progress would not have been possible without your support. To Prof. Ana Teresa Pérez-Leroux and Prof. Jeffrey Steele: thank you for your patience, your comments, and your insight over the years. I have learned so much from you. Your dedication to your craft and to your students is inspiring. Thank you to my external committee members Prof. Ineke Mennen and Prof. Jessamyn Schertz. I appreciate all the insight and feedback during the defence. Thank you to all five of you for agreeing to a virtual defence! What an experience!

The Department of Spanish and Portuguese would be lost without Blanca Talesnik, Rosinda Raposo, Sara Solis, and Paula Triana. You provide the graduate students in the department with endless support, wisdom, and guidance while we attempt to navigate the bureaucracy of graduate school. Thank you!!

Those who come before us show us the way forward! Thank you to Natalia Mazzaro, Danielle Thomas, Anna Limanni, Irina Marinescu, Yasaman Rafat, Yadira Álvarez-López, and Tanya Battersby for your insight, support, guidance, and friendly ear.

Much of my progress in the experimental portion of this study was possible thanks to the original intonation group: Malina Radu, Gaby Klassen, Matt Patience, and Olga Tararova and our French Department “cousins”: Simona Sunara and Meï-Lan Mammode. Thank you for going through the many versions of my tasks and providing helpful feedback. Thank you for discussing statistics and listening to my presentations multiple times. Thank you also for the fantastic memories while traveling to all sorts of exotic destinations for conferences!

Thank you to Nelson Darío González and Angélica Zamora Ríos for kindly lending your voices to the final version of the tasks. And thank you to all my participants for taking part in this study. Without you, there would be no thesis to defend!

Thank you to the many linguistic and literature colleagues from the “cueva”. We have shared many good times and few rough times. I cannot name you all here, but you know who you are, and I thank you all!

A very special thank you goes to Melanie Lopes, Diana Minster, Joanne Markle LaMontagne, Ailén Cruz, and Aga Bijos. We have many great memories together, but you were also there for me when life got in the way of my research. I hope I have been able to support you the same way you have supported me.

I would also like to thank Prof. Ellen Anderson and Prof. Michol Hoffmann for inspiring me to pursue graduate studies and for introducing me to the wonderful world of linguistics!

As a graduate student, I had the great fortune of teaching a variety of classes. I want to thank my students for providing me with so much insight. I have learned how to teach but I have also learned how to learn. I can also confidently affirm that patience is, in fact, a virtue!

I would be remiss if I did not mention the Salsa Community of Toronto. I owe a lot to all the dancers I have met over the years and all the events in which I have participated. Thank you for helping me keep my sanity through the years!

To anyone else who I have not mentioned yet: former roommates, friends from various circles and from lives prior to graduate school. There are so many people to name: I have not forgotten about any of you! Thank you!

Thank you to my parents for making the ultimate immigrant sacrifice in order to provide a better future for their children! Thank you to my brother for turning out to be a pretty cool guy and a friend like no other.

And last but not least, thank you to David. You only experienced the last leg of this journey, but you have been supportive, kind, and excited about all of it and for that I am truly grateful! I am lucky to have you in my life!

¡Mil gracias a todos!

Table of Contents

List of Tables

List of Figures

List of Appendices

Chapter 1
Introduction

1.1 Introduction

Y/N questions and comparable broad focus declarative statements in Spanish have the same surface syntactic structure. Consider the following example:

(1) Tienen helado de chocolate.
'They have chocolate ice cream.'

(2) ¿Tienen helado de chocolate?
'Do they have chocolate ice cream?'

The only way to distinguish utterance (1) from utterance (2) is by means of their intonation patterns. This dissertation looks at the perception and production of the intonation of Y/N questions and statements by L2 Spanish speakers, native speakers of English, who learned Spanish as adults. The intonation patterns that signal the Y/N question-statement distinction in Spanish are presented in Figure 1.1 and described in the paragraph below.

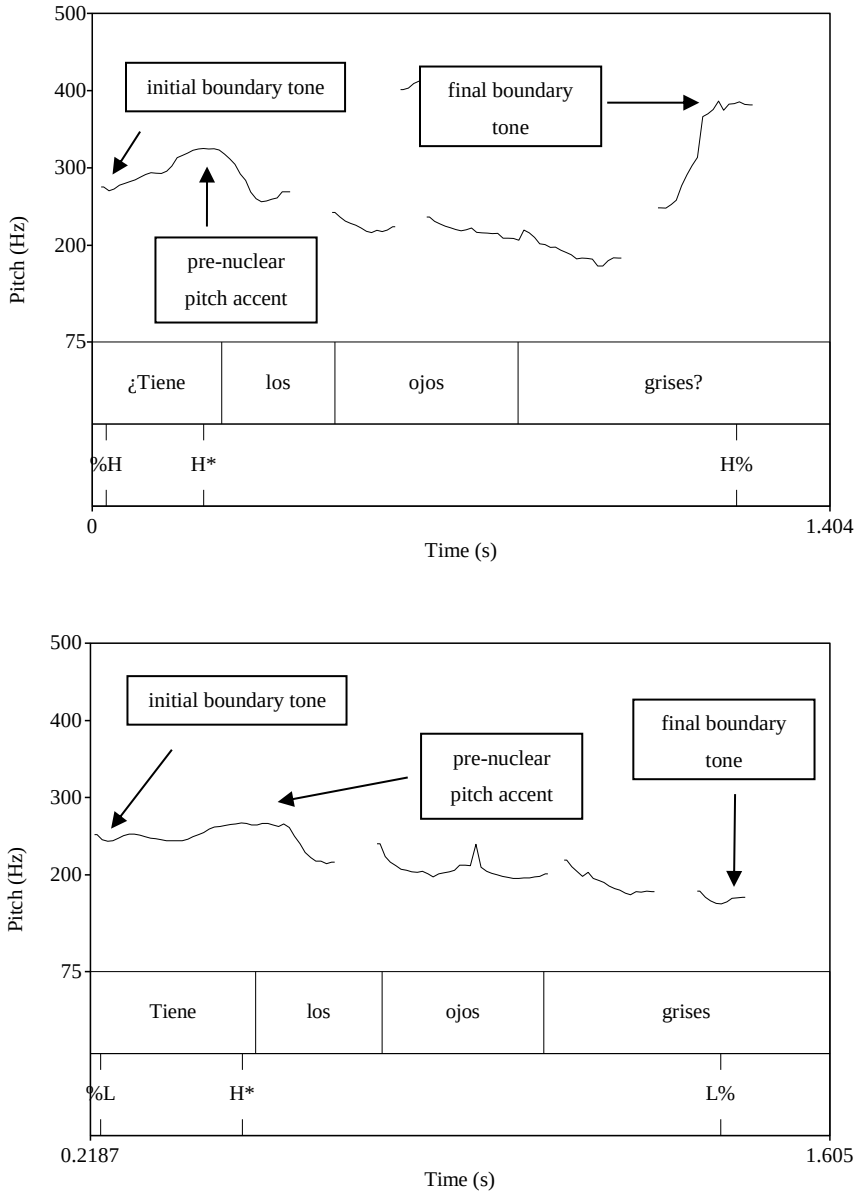

Figure 1.1. ¿Tiene los ojos grises? 'Does s/he have grey eyes?' (top) and *Tiene los ojos grises* 'S/he has grey eyes' (bottom) as spoken by a female native Spanish speaker. Initial boundary tone, pre-nuclear peak, and final boundary tone of the intonation contour are indicated.

Spanish Y/N questions have a higher initial boundary tone and pre-nuclear peak with respect to comparable statements (Canellada & Madsen, 1987; Face, 2007; Navarro Tomás, 1944; Prieto & Roseano, 2010; Sosa, 1999) (Figure 1.1). These height differences have not been found to distinguish English Y/N questions from comparable statements but studies on British (Ladd & Morton, 1997) and American (Chen, 2003) English have found that a height difference in the pre-

nuclear peak signals the difference between a *normal* and an *emphatic* utterance. Many varieties of English and Spanish are described as having a rising final boundary tone for Y/N questions and a final falling boundary tone for statements (Bartels, 1999; Bolinger, 1978; Face, 2007; Prieto & Roseano, 2010; Sosa, 1999). Additional research on the intonation of English however suggests that 1) the final rising boundary tone in English Y/N questions is not a necessary feature of this utterance type (e.g., Geluykens, 1988) and 2) a final rising boundary tone in English is also used for other utterance types (e.g., high rising terminals as described in Warren, 2016) that are not Y/N questions. Research on Spanish intonation has also reported that Caribbean varieties of Spanish will consistently signal a Y/N question with a final fall instead of a final rise (Prieto & Roseano, 2010; Sosa, 1999). This research suggests that the final boundary tone may be unreliable for the distinction between Y/N questions and statements, particularly for L2 Spanish speakers who are native speakers of English. Therefore, English speakers acquiring Spanish must learn that a higher initial boundary tone and pre-nuclear peak distinguish Y/N questions from statements in the target language, unlike in their L1. These intonational differences between the two languages are the starting point for the current study. The following section presents the research questions and an overview of the current study.

1.2 Research Questions and Overview of Current Study

This study is guided by 3 research questions:

1) Are L2 learners of Spanish able to accurately perceive the relatively higher initial boundary tone and pre-nuclear peak of Spanish Y/N questions with respect to comparable statements?

2) Are L2 learners of Spanish able to accurately produce the relatively higher initial boundary tone and pre-nuclear peak of Spanish Y/N questions with respect to comparable statements?

3) What is the relationship between perception and production of initial boundary tone and pre-nuclear peak in L2 Spanish speakers?

These research questions were investigated through a perception and a production task which involved 9 native Spanish speakers and 22 advanced L2 Spanish speakers (L1 English). The perception task was a gating experiment where the participants heard an utterance in gradual increments and were asked whether they heard a question or a statement. Presenting the utterance in increments allowed us to isolate the initial boundary tone and pre-nuclear peak and test its perception while ensuring that no other part of the utterance would be a cue for the listener about the utterance type. The production task required the participants to listen to a narrative in which one utterance (Y/N question or statement) was missing. Based on the context of the narrative, the participant was asked to provide the missing utterance. The narratives in the production task were designed to elicit a Y/N question or a statement in a guided way. This resulted in comparable utterances across speakers. A thorough description of the two tasks, as well as the stimuli, and detailed participant profiles will be presented in Chapter 3.

1.3 Contributions

The current study has three main contributions. First, it analyzes the L2 perception and production of intonation by the same participants and seeks to gain a deeper understanding of the relationship between the two. This relationship has not been explored in great detail in L2 intonation research. As we will see in Chapter 2, only a handful of studies have explicitly set out to investigate this relationship in L2 intonation and each study suggests a different relationship: Ortega-Llebaria and Colantoni (2014) found that their L2 English speakers (L1 Spanish or Mandarin) were accurate in perception before production; Puga, Fuchs, Hudson, Setter, and Mok (2018) found their L2 English speakers (L1 German) to be more accurate in production before perception; and Zárate-Sández

(2015) found a positive correlation between perception and production of the L2 Spanish (L1 English) speakers of his study. However, the positive correlation in Zárate-Sández (2015) was only present in one of the intonational cues he analyzed; the pre-nuclear pitch accent alignment showed a significant positive correlation between perception and production, but the final boundary tone height did not. While these studies will be discussed in greater detail in section 2.9, it must be noted that they involved different L1-L2 language pairs, utterance types, and parts of the intonational contour. It is clear that more research is needed in order to gain a better understanding of the relationship between L2 perception and production of intonation.

The current study also contributes a more fine-grained analysis of the intonation pattern of Y/N questions and statements in Spanish. Prieto and Roseano (2019), in their overview of issues and challenges in Spanish intonational research, identify the lack of a *full* description of Spanish intonation as one of the current issues in the field. As we will see in Chapter 2, the description of Spanish intonation clearly suggests that the height difference between Y/N questions and statements is the defining distinguishing feature, however, there is no numerical value, minimum threshold, or detailed description of how to quantify this height difference. Without this reference point, investigation of this phenomenon becomes challenging, particularly when we seek to determine how native-like L2 speakers are or can be. In the current study, we worked towards determining what this threshold may be. We chose to transform pitch differences into semitones (as recommended by Gut, 2013 and Nolan, 2003). We then set the discriminability threshold at 1.5 semitones as suggested by Nooteboom (1997). By identifying this gap in the literature and attempting to provide empirical findings about it, we contribute to a more fine-grained analysis of Spanish intonation.

Finally, this dissertation contributes methodologically to L2 intonation research. Some studies have found noticeable differences between the intonation patterns of scripted experimental settings

and those of spontaneous speech (e.g., Face, 2003; Hedberg & Sosa, 2002) suggesting that methodological decisions, in particular the type of stimuli used, are not trivial in research, particularly if we seek to apply our findings to the everyday use of the language. There has been a clear shift in the last 15 years towards empirical methods that seek to characterize intonation in context as opposed to descriptions of the intonational systems based primarily on the reading of isolated sentences of sometimes doubtful semantics (Kohler, 2006; Prieto & Roseano, 2019). The concern with "highly stylized language material" (as Kohler, 2006 characterizes it) is that the findings cannot be generalized to everyday speech. The methodological choices in the current study were intended to obtain findings that are as generalizable as possible. For the perception task, the acoustic parameters of the original sound files were not manipulated. In production, a task that would elicit relatively more spontaneous production was chosen. These tasks are discussed in greater detail in Chapter 3.

1.4 Thesis Structure

In order to answer the three research questions presented in section 1.2, this dissertation proceeds as follows. In Chapter 2, I will introduce the auto-segmental metrical (AM) model of intonation with a more detailed description of Spanish and English Y/N questions and comparable statements within the model. I will then review and evaluate speech learning models as well as the relevant previous research in L2 intonation in order to formulate predictions for the research questions of this study. Specifically, I will present research on the L2 perception of intonation, the L2 production of intonation, and the relationship between L2 perception and production of intonation. In Chapter 3, I will begin by restating the research questions and providing hypotheses based on the literature reviewed in Chapter 2. In Chapter 3, I will also provide a detailed description of the methodology including participant profiles and the stimuli for both the perception and the

production tasks of this study. The results will be presented in Chapter 4 and they will be discussed in Chapter 5 in light of the predictions discussed in Chapter 3. Chapter 5 will then conclude with a discussion of the contributions of this study to the field of Spanish intonation as well as possible avenues for future research.

Chapter 2
Background

2.1 Introduction

In order to answer the research questions of this study, we must compare sentence prosody of Y/N questions and comparable statements in English and in Spanish. This chapter will start with a description of the auto-segmental metrical (AM) model, the model used to describe the main characteristics of the intonation contours of Y/N questions and statements in English and Spanish. Since this study focuses on the initial portion of the intonation contour (specifically initial boundary tone and pre-nuclear peak) in both languages, a section will be dedicated to the importance of utterance-initial cues in intonation. In order to make predictions, the second part of this chapter will discuss studies on the L2 perception and production of intonation as well as those studies that have explicitly considered the relationship between the two with respect to L2 intonation.

2.2 The Auto-Segmental Metrical (AM) Model

The most widespread model for representing intonation is the autosegmental-metrical (AM) model developed from Pierrehumbert's (1980) analysis of English intonation. The phonological primitives of this model are high (H) and low (L) tones. The intonation contours, also called configurations, derive from the combination of these tones. In the AM model, the connection between the intonation contour and the segmentals (i.e. text-tune association) occurs through the anchoring of the H or L tones to metrically strong syllables (i.e. the relatively more prominent syllable in a given word; Ladd, 2008, Chapter 2). The tone that is anchored to a metrically strong syllable is accompanied by an asterisk in the notation and is called a pitch accent. These pitch

accents can be monotonal (e.g., H*) or bitonal (e.g., H*+L). A bitonal notation is used when the pitch accent shows a rapid F0 movement with both the low and high tone of the bitonal pitch accent being associated with the stressed syllable. This differs from a monotonal pitch accent where only one tone is associated with a stressed syllable. In the case of multiple pitch accents (e.g., sentences (3) b) and (3) c) below), the final one is called the nuclear pitch accent; all the others are pre-nuclear pitch accents. The AM model also uses tones to characterize the edge of an utterance (beginning or end of a sentence). Edge tones can be divided into phrase accents and boundary tones. Phrase accents are "free-standing unstarred tones that occur after the last pitch accent" (i.e. the final stressed syllable; Ladd, 2008, p. 88) and are notated H- or L-. Boundary tones are marked with a '%' and occur at all phrase boundaries (at the beginning and at the end of an utterance).[1] Boundary tones are not anchored to metrically strong syllables. Examples of pitch accents and boundary tones are provided in (3). The metrically strong syllables to which the pitch accents are anchored are indicated in bold.

(3)

a) *Cali**for**nia*

H* L-L% (Pierrehumbert, 1980, p. 38)

b) ***E**dinburg is the capital of **Scot**land.*

H*L L* L-L% (Ladd, 2008, p. 107)

[1] The notation for the **final** boundary tone is consistently the high or low tone followed by '%' (e.g., L%). The **initial** boundary tone, however, has shown variation among researchers in the order of the notation with no change in meaning. A high initial boundary tone, for example, can be %H or H% depending on the author's preference. There is no change in meaning here.

c) *Did she **have** her **ba**by yet?*

H* L*H H% (Hedberg et al., 2004)

In the AM model, an intonation contour consists of a string of elements occurring at well-defined points in the utterance: pitch accents at metrically strong syllables and boundary tones at intonation phrase boundaries. An intonation contour minimally consists of a pitch accent followed by a phrase accent and a boundary tone (as seen in (3) a)).[2] All combinations of pitch accents, phrase accents, and boundary tones are possible in theory, but it is not the case that all possibilities are attested in every language. The relevant combinations for English and Spanish declarative and Y/N questions studied in this dissertation will be discussed in the following paragraphs.

2.3 Statements and Questions

As mentioned in Chapter 1, this dissertation examines broad focus declarative statements (referred to henceforth as 'statements') and yes-no (Y/N) questions in English and Spanish. A statement (4) involves the presentation of information by the speaker without any particular emphasis on one part of the utterance. A Y/N question (5) involves a speaker's attempt to elicit information that s/he does not possess and it requests a yes/no response from the addressee (Geluykens, 1987; Gunlogson, 2003).

(4) John looks out the window, sees the rain, and says: *It's raining.*

[2] This description and these examples are in reference to Mainstream American English. Some variants of the AM model, for example the Spanish one, do not support the existence of phrase accents. Spanish will be discussed in more detail in the following paragraphs.

(5) Mary is getting ready for work but is unsure about the weather. She heard that it could be a rainy day, so she asks John who is sitting near the window: *Is it raining?*

Having described the structures of interest, we now turn to the specifics of English and Spanish Y/N question and statement intonation.

2.3.1 The Intonation of Y/N Questions and Statements in English and Spanish

The difference between Y/N questions and statements in Spanish is signalled phonologically by differences in intonation pattern (e.g., Face, 2007; Prieto & Roseano, 2010; Sosa, 1999). Consider the following example:

(6) Rompió la mesa del comedor.
'He broke the dining room table'

(7) ¿Rompió la mesa del comedor?
'Did he break the dining room table?'

In both Spanish examples, the words and their order is exactly the same. In spoken Spanish, the intonation patterns of (6) and (7) will differ. In English, Y/N questions and statements also have some intonational differences but crucially the syntax also signals the difference between the two utterance types.[3] The specific details of the intonation contours of Y/N questions and statements in English and Spanish will be discussed in the following sections. Particular focus will be placed

[3] Non-inverted questions in English, e.g., *It's raining?,* are not considered here. Whether non-inverted questions and canonical Y/N questions have the same status is debatable. Gunlogson (2002) argues that they do not. According to her, non-inverted questions, or "declarative questions", are used in restricted contexts and they cannot be used out-of-the-blue as can canonical Y/N questions. See Saindon, Trehub, Schellenberg, and van Lieshout (2017) and Patience, Marasco, Colantoni, Klassen, Radu, and Tararova (2018) for preliminary research on the intonation patterns of declarative questions in Canadian English.

on the initial boundary tone, the pre-nuclear peak, and the final boundary tone (Figure 2.1). The initial boundary tone is an edge tone that corresponds with the very beginning of the intonational contour. Some research (Navarro Tomás, 1944; Canellada & Madsen, 1987) suggests that Spanish Y/N questions start at a higher pitch than comparable statements. No such observation has been made for English. The pre-nuclear peak is the first rise-fall associated with the first stressed syllable of the utterance. Numerous Spanish sources (Navarro Tomás, 1944; Canellada & Madsen, 1987; Sosa, 1999; Face, 2007; Prieto & Roseano, 2010) have observed this peak in the Y/N question to be noticeably higher than in comparable statements. Studies have not found this height difference in English Y/N questions and comparable statements but studies on British (Ladd & Morton, 1997) and American (Chen, 2003) English have found that a height difference in the pre-nuclear peak signals the difference between a *normal* and an *emphatic* utterance. The final boundary tone is the edge tone that describes the final portion of the intonational contour. It is often described as rising for questions and falling for statements (Bolinger, 1978) in many languages, not only English and Spanish.

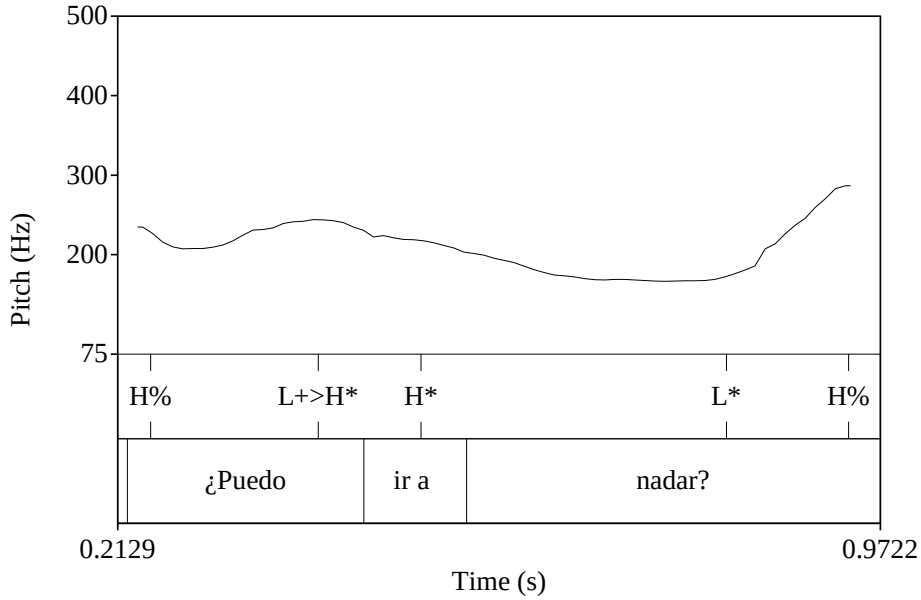

Figure 2.1. ¿Puedo ir a nadar? 'May I go swimming?' spoken by a female native Spanish speaker. Initial boundary tone (H%), pre-nuclear pitch accent peaks (L+>H* and H*), nuclear pitch accent (L*) and final boundary tone (H%) are indicated.

2.3.1.1 English Statements and Y/N Questions

This section presents a tonal inventory of the relevant elements of the intonational contour of English[4] Y/N questions and statements as mentioned above. We begin with the initial boundary tone. Beckman, Hirschberg, and Shattuck-Hufnagel (2005) find a high initial boundary tone (H%) to be very uncommon ("marginal", p. 23) in American English and Bartels (1999) explains that a high initial boundary tone is uncommon and can be perceived as insistence or impatience in American English (p. 125). The (first) pre-nuclear peak in English statements can be monotonal (H*: Ladd, 2008) or bitonal (L*+H: Bartels, 1999). The same peak in English Y/N questions has been found to be both monotonal (L*) or bitonal (L*+H) (Bartels, 1999). No noticeable or

[4] The references to the intonation system of English are for American English. To date, there is no comprehensive overview of the intonation of Canadian English, nor is it entirely clear whether or not it differs from the intonation patterns documented for American English.

significant height difference has been reported between the first pitch accent of a statement and the first pitch accent of a Y/N question in English. The final boundary tone has received considerable attention. As we saw in the previous section, most research suggests a final fall (L%) for statements and a final rise (H%) for Y/N questions in English. However, Bartels (1999) explains that the portion of the contour that corresponds to the interval between the last stressed syllable and the end of the utterance conveys the question/statement distinction. This part of the contour, as the reader may recall, is called a phrase accent. It is the phrase accent that has the "decisive role" (p. 36) in conveying a rise or a fall (respectively H- or L-). Therefore, a final fall in English can be L-L% or L-H%[5] whereas a final rise can be H-H% or H-L%. Table 2.1 provides a summary of the main patterns just described.

Table 2.1.

Summary of the intonational characteristics of English statements and Y/N questions

	statements	*Y/N questions*
initial boundary tone height	no mention	H% possible but not typical of American English
pre-nuclear pitch accent	H* or L*+H	L* or L*+H (no height difference compared to statements)
final boundary tone	falling: L-L% or L-H%	rising: H-H% or H-L%

[5] Ladd (2008) suggests that LH% is typical for British English questions but American English speakers may perceive it as condescending or as an echo-question.

2.3.1.2 Spanish Statements and Y/N Questions

This section presents a tonal inventory of the relevant elements of the intonational contour of Spanish Y/N questions and statements. In Spanish statements, the initial boundary tone is a low tone (L%) (Canellada & Madsen, 1987; Face, 2007; Navarro Tomás, 1944; Prieto & Roseano, 2010; Sosa, 1999). There is some debate, however, concerning the initial boundary tone height in Y/N questions. Navarro Tomás (1944) and Canellada and Madsen (1987) describe it as being noticeably higher than a comparable statement. Face (2007), however, finds no difference between statement and Y/N question initial boundary tone and Prieto and Roseano (2010) make no mention of this height difference when distinguishing between Y/N questions and statements.[6] Pre-nuclear pitch accents in Spanish statements and Y/N questions are typically high and can be either monotonal (H*) or bitonal (L+H*). The height of this pre-nuclear pitch accent is what contrasts statements and Y/N questions. In fact, pre-nuclear peak height in questions is noticeably higher than comparable statements (Canellada & Madsen, 1987; Face, 2007; Navarro Tomás, 1944; Prieto & Roseano, 2010; Sosa, 1999).[7] The peak of pre-nuclear pitch accents in both statements and Y/N questions in Spanish is generally realized on the post-accentual syllable (Face, 2007; Prieto & Roseano, 2010; Sosa, 1999). Finally, the sources concur on the presence of a falling final boundary tone (L%) for statements versus a rising final boundary tone (H%) for Y/N questions for the majority of the varieties of Spanish. Some varieties of Spanish, however, do not follow this pattern as concerns the final boundary tone; they will be discussed in the following section. Phrase accent (L- or H-) is not mentioned in any of these sources. A phrase accent (H- or L-) is unnecessary

[6] Of the ten Spanish varieties surveyed in Prieto and Roseano (2010), only Venezuelan Andean Spanish observes a higher initial boundary tone (%H). This tone, however, is found only in uncertainty statements (e.g. 'She may not like the present I have bought her').

[7] It is important to note here that the literature does not quantify this difference in Hz or any other measurement. The following chapters discuss how to overcome this.

because Spanish word stress is generally ultimate, penultimate or ante-penultimate. As a result, there will be at most two unstressed syllables between the last stressed syllable, which carries the nuclear pitch accent, and the final boundary tone. Either type of phrase accent here would not create a meaningful contrast in Spanish (Sosa, 1999, p. 95). Table 2.2 provides a summary of the main patterns just described.

Table 2.2.

Summary of the intonational characteristics of Spanish statements and Y/N questions

	statements	*Y/N questions*
initial boundary tone height	low: L%	high: H% (debated)
pre-nuclear pitch accent	H* or L+H*	H* or L+H* (noticeably higher than comparable statements)
final boundary tone	falling: L%	rising: H%

2.4 The Importance of Utterance-Initial Cues

A discussion on the importance of utterance-initial intonation cues is warranted here because a common view is that the final portion of the intonation contour is sufficient to communicate the utterance type. This view would make the rest of the contour seem unimportant or irrelevant. A number of studies have found that listeners, in the absence of any other cues, will rely on a final rise to identify Y/N questions and a final fall to identify statements (Face, 2007; Grabe et al., 2003). While 70 % of the world's languages displays this type of pattern (Bolinger, 1978), there are a number of languages that do not use a final rise to signal a Y/N question and a final fall to signal a statement (e.g., Chickasaw: Gordon, 1999; Akan: Kügler, 2017; other African languages: Rialland, 2009). Furthermore, Spanish and English, which generally make use of the most common

patterns for Y/N questions and statements (as seen in Tables 2.1 and 2.2) can also vary in their use of the final boundary tone making it less reliable in terms of determining the utterance type. High rising terminals in English and dialectal variation in Spanish serve to illustrate this point.

In English, high rising terminals (HRT), also known as 'valley girl speak' or 'uptalk', are characterized by the use of a final rise in a declarative sentence. This phenomenon has been observed in a number of English speaking countries including the United States (Ching, 1982), Canada (Shokeir, 2008), Australia (Guy et al., 1986), and New Zealand (Warren, 2005).

In Caribbean varieties of Spanish, Y/N questions often end in a final fall instead of the expected rise (Prieto & Roseano, 2010; Sosa, 1999). If languages *exclusively* signaled a Y/N question with a final rise and a statement with a final fall, situations like those above would be cause for great confusion and miscommunication. Research on other languages supports the idea that the final rise is not always a reliable cue for utterance type. Data from German and Greek are presented below.

Petrone and Niebuhr (2014) studied the role of the pre-nuclear area of the intonation contour in Standard Northern German (henceforth German) Y/N questions and statements. German can signal Y/N questions syntactically but, unlike English, it can also signal a Y/N question prosodically while using statement syntax. In German, the pattern H* L*+H L-L% (where H* represents the pre-nuclear peak, L*+H the nuclear peak, and L-L% the edge tone) is used for both Y/N questions and statements. The difference between the two types of utterances lies in the steepness of the fall after the pre-nuclear peak (H*). German statements have a much steeper fall than Y/N questions because the H* in statements peaks occurs 40ms later than it does in Y/N questions. This acoustic difference prompted the authors to hypothesize that German listeners would be cued to sentence type by the pre-nuclear peak alone. Two experiments were designed to test this hypothesis. The first experiment was a gating experiment. Participants heard utterances in three increments: after the rise and fall of pre-nuclear peak, right before the nucleus, and the entire

utterance. They were asked to rate what they heard on three scales: 1) astonished-matter of fact, 2) uncertain-certain, and 3) questioning-knowing. Fifteen native German listeners were successful in distinguishing between a Y/N question and a statement when they only had the pre-nuclear information available (after the first gate). This finding suggests that the "same" phonological pitch accent (H*) is realized differently in the two types of utterances. Based on the phonetic analysis of the stimuli, Petrone and Niebuhr (2014) suspected that the steepness of the fall after the pre-nuclear peak could signal the sentence type. Experiment 2 was designed to test this hypothesis. Stimuli were modified for steepness, alignment, and shape. Steepness was modified by shifting the end of the F0 fall of the pre-nuclear peak in four steps of 50 ms each. Alignment was modified by shifting the peak of the pre-nuclear pitch accent 80 ms to the right. Shape was modified by introducing an additional F0 point in the middle of the falling movement and that was then shifted 25 Hz upwards to create a concave shape and 25 Hz downwards to create a convex shape. Both statement and question bases were used. Of the three cues that were modified, the authors found steepness of the fall of the pre-nuclear peak to be the strongest cue for the participants. Specifically, a steep fall after the pre-nuclear peak was a strong cue for a statement while a shallow fall was a strong cue for a Y/N question. These findings suggest that the question-statement distinction is not exclusively realized by the intonational nucleus. In fact, in this case, the intonational nucleus in Y/N questions and statements does not distinguish the two utterance types at all.

A similar study on Greek Y/N questions and statements was conducted by Baltazani, Kainada, Lengeris, and Nikolaidis (2015). Previous studies on Greek describe the difference between Y/N questions and statements in pre-nuclear pitch accents to be the following: Y/N questions show downstep (lowering of pitch at specific pitch accent throughout the utterance), have higher H tones, and earlier alignment (location of the pitch accent peak with respect to the syllable with which it is associated) compared to their statement counterparts. This study tested two hypotheses: 1)

listeners would discriminate between string identical statements and questions if the nuclear pitch accent and final boundary tone were removed; and 2) differences between string-identical statements and questions would be seen in the phonetic details of the pre-nuclear pitch accents. Recordings of ten pairs of identical statements and Y/N questions were used in 2 perception tasks for a total of 240 stimuli (10 questions and 10 statements from 12 different speakers) with the nuclear pitch accent and final boundary tone removed. In an AX discrimination task, listeners were presented with pairs of stimuli with the nuclear pitch accent and final boundary tone removed. They were told that the utterances were incomplete and asked to decide whether the first member of the pair was the same as the second. In the identification task, listeners were asked if they were hearing a statement or a Y/N question. Overall, when presented with the pre-nuclear area of the utterance, listeners were successful in distinguishing between statements and Y/N questions in Greek, in support of hypothesis 1. For hypothesis 2, an acoustic analysis of alignment, slope, and down-step was performed on a subset of 60 correctly perceived utterances. The authors found that L*+H aligned earlier in questions than in statements, that the slope was significantly steeper in questions than in statements, and that down-step was present in questions but not in statements. In sum, there are systematic acoustic differences between Y/N questions and statements in the pre-nuclear area in Greek. These differences were salient enough for listeners to distinguish between utterance type without hearing the nuclear part of the contour. These two studies serve to underline the fact that, even in languages where typically a final rise is associated with Y/N questions and a final fall is associated with statements, differences in the pre-nuclear area are fundamental when differentiating utterance type.

2.5 Interim Summary

Before discussing the previous research in L2 intonation, we briefly summarize the information that has been presented thus far. The intonation model in the current study is the AM model, the primitives of which are a high (H) and low (L) tone. Together they form pitch accents (H*, L*, L+H*, etc.), which are associated with stressed syllables, and initial and final boundary tones, (L% and H%) which are associated with the edge (beginning or end) of the intonation phrase. Of particular interest here are the differences between the initial boundary tone and the pre-nuclear peak (first pitch accent in an utterance with multiple pitch accents) of English and Spanish Y/N questions and comparable statements. Crucially, a height difference between Y/N questions and statements has been found in Spanish but not in English. Research on other languages has also shown a growing interest in utterance-initial intonation cues because researchers have found them to be important and sufficient for distinguishing between Y/N questions and statements. We now turn to the discussion of speech learning models, previous research in the L2 perception of intonation, the L2 production of intonation, and the relationship between the two.

2.6 L2 Speech Models

Before reviewing the relevant empirical research in L2 intonation, it is important to discuss some influential L2 speech learning models. In particular, the Speech Learning Model (SLM: Flege, 1995) and the Perceptual Assimilation Model applied to second language speech sounds (PAM-L2: Best & Tyler, 2007) are discussed here. Both the SLM and the PAM-L2 propose L1-based cross linguistic influence on L2 speech, however, neither model explicitly considers L2 intonation but rather focuses on segmentals. While some principles from these models discussed in the following paragraphs are applicable beyond the specific scope of L2 segmental acquisition, there is a clear need for a model that focuses primarily on L2 intonation. To this end, Mennen (2015)

has proposed a model called the L2 Intonation Learning Theory (LILt) which proposes four different dimensions researchers should consider when investigating L2 intonation. In the following paragraphs, we will elaborate on the SLM, PAM-L2, and LILt.

The SLM proposes that adults maintain the ability to accurately perceive the full range of sounds of the world's languages and to create new phonetic categories throughout their lifetime. The successful acquisition of L2 sounds, however, depends on the perceived cross-linguistic similarity between the L1 and L2 sounds. Specifically, L2 sounds can be perceived as "similar" or "new" when compared to an L1 sound. If the listener perceives the L2 sound to be similar to an instance of a particular L1 sound, the model predicts that the L2 sound will be more difficult to acquire. In turn, if the listener perceives the L2 sound to be different from any of the L1 sounds and thus new, the result will be the creation of a new category. The categories in this model are position sensitive allophones rather than abstract phonemes. The acoustic properties of a sound in context have a great effect on the categorization of "similar" or "new". Other factors that can affect the success of new category creation include L2 language experience (i.e., input) and age of acquisition of the L2 (the earlier the more successful). With respect to production, Flege (1995) suggests that non target-like L2 production results from inaccurate perception of the L2 sound as well as insufficient experience with the target language.

The PAM-L2 makes predictions based on the perception of pairs of non-native contrasts. In this model, categories are not stored as mental representations. Instead, the listener perceives the articulatory gestures of each sound. Accurate discrimination will be the result of how the learner equates (or assimilates) each sound in the contrast to the L1 sound category. A pair of non-native sounds can be categorized in different ways with respect to a speaker's L1 sounds and this will determine whether or not the learners can successfully discriminate these sounds. If the two non-native sounds are assimilated to two different L1 categories, the discrimination is predicted to be

good, whereas, if the two non-native sounds are assimilated to a single L1 category, discrimination is predicted to be poor. Furthermore, if the two non-native sounds are assimilated to the same L1 category but one is considered a "better" exemplar than the other, discrimination could vary depending on the difference perceived by the L2 learner. In the PAM-L2, the effect of experience and the age of onset of acquisition also play a role in the successful assimilation. The contribution of age to perceptual learning occurs primarily through interactions with length of residence, relative usage of the L1 and L2, and relative quantity and quality of input from native L2 speakers. These two models have been tested primarily via the study of the acquisition of vowels (e.g., Flege et al., 1999; Levy, 2009) and consonants (e.g., Flege et al., 1995; Rose, 2010). There have been a few studies, however, that illustrate how some of the principles from these L2 segmental models are applicable to suprasegmental features. Trofimovich and Baker (2006), for example, investigated the effect of experience (i.e., years of residence in the United States) on the L2 acquisition of 5 suprasegmentals by L1 Korean speakers: stress timing, speech rate, pause frequency, pause duration, and peak alignment. With respect to stress timing, English is labelled as a stress-timed language which means its rhythm is characterized by alternations in the relative degree of stress of neighbouring syllables, with stressed syllables being significantly longer than unstressed ones and most vowels in unstressed syllables reducing to schwa. Korean, in turn, is labelled a syllable-timed language which means that it does not exhibit alternations in degree of stress and has syllables that are approximately the same in duration. Speech rate in this study was calculated by dividing the number of uttered syllables in an utterance by the total duration (in seconds) of the utterance. Previous research (as cited by Trofimovich and Baker, 2006) has found that L2 speakers exhibit slower speech rates than native speakers of the same language. Pauses were defined as any break in the speech stream longer than 100 ms in length. Pause frequency was calculated by averaging the number of pauses for each participant across the six sentences; pause

duration was computed by averaging pause durations for each participant across the six sentences. Learners also commonly pause more in their L2 than in their L1. Peak alignment, the location of the highest peak with respect to the stressed syllable, shows some different patterns between English and Korean. English signals syllable prominence with pitch peaks while Korean marks phrase boundaries with them. The two languages, however, do not seem to diverge a great deal in the location of the pitch accent, as both languages place the peak on the last word in a phrase (aligned with the onset of the syllable in English and with its offset in Korean). The L2 speakers in Trofimovich and Baker (2006) were adult learners of English and had been residing in the US for either 3 months, 3 years, or 10 years. Through a delayed repetition task, Trofimovich and Baker obtained utterances from 40 participants (10 per residence group as well as 10 native English controls) for a total of 240 declarative sentences (6 utterances X 10 speakers per group X 4 groups). The authors found that the L2 group that had been living in the United States the longest (10 years) also produced stress timing patterns that was closer to the native English-speaking controls suggesting that L2 experience affected this particular suprasegmental. Speech rate, pause duration, and pause frequency were found to correlate more with age of first exposure to English (i.e., the earlier the first exposure to English, the more native-like their speech rate, pause duration, and pause frequency). Finally, and perhaps most interestingly, none of the L2 groups were native-like in their peak alignment. In fact, despite differences in length of residence and age of first exposure, all three L2 groups did not differ significantly from each other but they were all significantly different from the native English-speaking controls. The results suggest that suprasegmental learning depends on the particular suprasegmental studied. Trofimovich and Baker (2006) proposed that what is readily learnable from L2 experience are only those L2 suprasegmentals that are relatively perceptually distinct from L1 suprasegmentals, not those that are perceptually similar across learners' L1 and L2. One can conclude, based on this suggestion, that the authors believe

that peak alignment in the two languages is the least perceptually distinct feature while stress timing is the most perceptually distinct. While this study did not set out to explicitly test Flege's (1995) SLM, the proposal shares Flege's idea that perception will have an effect on production. Specifically, if the L2 speaker perceives the L2 feature to be different (or new) compared to the L1 feature, a new category will be formed. If a new category is formed for the L2 feature, this will be evident in production.

While the findings in Trofimovich and Baker (2006) represent an important step in the research on L2 suprasegmental acquisition, the question of how predictions in L2 suprasegmental acquisition can be made still remains. An important difference between segments and intonation is that in order to successfully acquire intonation, L2 speakers must acquire the language-specific intonation-meaning relationship. Mennen's model, by proposing multiple dimensions along which to investigate L2 intonation, provides a useful framework for L2 intonation research. In order to make predictions about the relative difficulty of production and perception in L2 intonation, the LILt model proposes four dimensions along which similarities and differences between L1 and L2 can be analyzed: systemic, realizational, semantic, and frequency. The **systemic** dimension refers to the inventory of structural elements (pitch accents, boundary tones, etc.). This means not only the possibility of certain elements in specific parts of the intonation contour but also how they are allowed to combine within a particular contour. The initial boundary tones %L and %H are found in the inventory of both English and Spanish. With respect to the pre-nuclear peak, monotonal H* as well as bitonal L*+H and L+H* are present in both languages. The **realizational** dimension is concerned with the alignment, scaling, and slope of the elements identified in the systemic dimension. In the current study we are investigating height differences and, in both languages, with respect to the initial boundary tone, an %H will be higher in pitch than an %L. With respect to the pre-nuclear peak, both English and Spanish have the option of realizing the same pitch accent

at different pitch heights as well. The **semantic** dimension focuses on the way the structural elements convey meaning. With respect to the initial boundary tone, a higher one in Spanish signals a Y/N question whereas a higher one in American English is an indication of insistence or impatience. The pre-nuclear peak height difference in Spanish signals a Y/N question-statement distinction while in English the height difference has been found to distinguish normal from emphatic utterances. The **frequency** dimension is concerned with the frequency and distribution of the intonation primitives. A higher initial boundary tone in American English is described as "marginal" or uncommon while descriptions of Spanish intonational systems suggest that a higher initial boundary tone in Y/N questions is typical of the intonation contour of Y/N questions. Similarly, a higher pre-nuclear peak in Y/N questions is expected as part of the intonation contour in Spanish while the normal versus emphatic distinction in English is likely less common. In sum, Mennen's model allows for a clear cross-linguistic comparison across multiple dimensions which in turn allows researchers to make predictions about what type of difficulties (phonological, phonetic, semantic) L2 learners may have when acquiring L2 intonation. It should be noted, however, that Mennen's model is not meant to take the place of previous L2 speech models but rather to supplement them with tools to specifically characterize L2 intonation. Since the LILt is still a relatively new model, there are a few aspects that need to be explored further with empirical research. First, while the LILt is similar to both the SLM and the PAM-L2 in proposing that production difficulties may be perceptually motivated, it encounters some difficulty in determining perceptual difficulty in intonation without reference to the semantic dimension of intonation. The LILt also shares with the SLM and the PAM-L2 the proposal that age of learning is an important predictor of success. However, not many L2 intonation studies have tested this assumption and it may be the case that not all aspects of intonation are subject to the same influence of age of learning (see Trofimovich and Baker, 2006 above for example).

The discussion of these models serves as the starting point in the study of the L2 acquisition of intonation. They provide useful assumptions that allow researchers to identify potential difficulties for the L2 learner. However, since the acquisition of L2 intonation involves a language-specific intonation-meaning relationship, both the SLM and the PAM-L2 cannot be used to make specific predictions. It is not surprising then that Mennen's (2015) model introduces a semantic dimension along which the L2 learner of intonation may show difficulties. However, Mennen's model cannot make predictions about L2 intonation outcomes because, given that intonation may signal multiple functions, it is difficult to establish if certain differences are gradient or categorical. Furthermore, intonation interacts with other prosodic parameters such as tempo, duration, pauses, and voice quality making it difficult to establish cross-language similarity or difference. In sum, these models serve to establish some basic assumptions about L2 learning as well as potential areas of difficulty in L2 intonation. They can also provide some explanation of the results that we will see in Chapter 4. However, given what has been discussed above, they cannot be used exclusively to make predictions about the outcomes of the current study. To this end, we must also consider previous empirical research in L2 intonation.

We now turn to the final portion of this chapter to discuss studies in the L2 perception, L2 production, and relationship between L2 perception and production of intonation.

2.7 L2 Perception of Intonation

The following paragraphs discuss studies on the perception of L2 intonation that are relevant in predicting what the outcome of the current study may be. Each study is presented in detail and the section concludes with a summary of the relevant findings. The studies presented here consider different populations and different tasks, but each study contains a finding that contributes to the formulation of hypotheses for the current study. This section begins with a study (Grabe et al.,

2003) that compares the results of a perception task that involves only pure tones with one that involves sentences that contain the same patterns as the pure tones highlighting how L2 speakers can be successful in acoustic discrimination tasks but their success rate is affected by the introduction of language. The following study (Liu & Rodriguez, 2012) then illustrates how L2 speakers can distinguish between Y/N questions and statements with high accuracy rates. However, as discussed in section 2.3, Spanish makes use of the higher pitch in IBT and PNP to signal Y/N questions compared to statements which English does not do. This section, therefore, presents two studies (Radu et al., 2018; Trimble, 2013) that consider how L2 speakers perceive a feature that is not signaled in the same way in the L1. Finally, two studies (Gili Fivela, 2012; Radu et al., 2018) that make use of reaction times (RT) to measure difficulty in L2 intonation are discussed.

Grabe, Rosner, García-Albea, and Zhou (2003) tested native speakers of English, Mandarin, and Spanish[8] with 11 pure tone contours (no segmental information) that were modeled on well-attested English contours which were divided into two groups based on the direction of the final pitch movement: falling (HL) or rising (LH). Typologically, English and Spanish are more similar to each other and both are considerably different from Mandarin, a tone language. Tone languages use intonation for lexical distinctions as well as suprasegmental ones, while English and Spanish, both intonation languages, only use intonation suprasegmentally. Participants in this study listened to pairs of stimuli and were asked to rate the degree of similarity or difference (1 = very similar, 10 = very different) between each pair. The speakers from the three language groups behaved very similarly in their categorization of these contours. Their responses were plotted on a multi-

[8] The Mandarin speakers and Spanish speakers were tested in China and Spain respectively. They were not expected to be English speakers however, most of the Spaniards had rudimentary knowledge of English and the Mandarin speakers were able to read some English but they were not fluent speakers.

dimensional scale and, despite having very different L1s, their responses were statistically indistinguishable. A different group of L1 English, L1 Spanish, and L1 Mandarin speakers then repeated this task with the same contours, this time containing segmental information. Once segmental content was included, these speakers showed L1-influenced categorization patterns. Specifically, Mandarin speaker responses patterned differently compared to Spanish and English speakers. The authors suggested that the L1 of the Mandarin group affected the way these speakers heard and categorized the contours differently than the Spanish or English listeners. Since the only difference between the stimuli of the two tasks was the absence versus presence of segmental information, the authors concluded that the difficulty for L2 speakers when it comes to intonation, is not acoustic, but related to linguistic meaning. The authors suggest that the universal auditory mechanism that allows speakers to process the basic distinction between rising and falling contours, interacts with a given native language modifying this mechanism as a result of the interaction with a specific L1.

Liu and Rodriguez (2012) used the categorical perception paradigm to investigate the effect of the L1 on the perception of English Y/N questions and statements by native and L2 English (L1 Chinese[9]) speakers. Their study focused on the final portion of the utterance where a rise is typically associated with a question and a fall is typically associated with a statement in English (as we saw in section 3). Both identification (*question or statement?*) and discrimination (*same or different?*) tasks were administered to the groups who showed a typical S-curve in the identification task. In a graph where the Hertz values of the statement-question continuum are plotted on the x-axis and the percentage of question identification is plotted on the y-axis, as the

[9] The authors described these participants as native Chinese speakers but did not specify if they spoke Mandarin, Cantonese, or another Chinese language.

Hz value increases so does the percentage of question identification. The line that is produced by uniting these points looks like an "s". At approximately 250 Hz, listeners started to identify utterances as questions. While both groups showed the typical S-curve pattern, the slope of the Chinese native speakers' S-curve differed from that of the L1 English speakers' slope in being steeper suggesting an earlier category boundary (in Hz). This is to say that Chinese speakers started identifying questions in English at a lower frequency and the category change from statement to question was more abrupt compared to native English speakers. This difference was statistically significant, suggesting that language background plays a role in L2 perception of intonation. Once the identification task had identified where the category boundary existed for the speakers, the discrimination task results, when plotted on the same graph, were expected to show a clear peak in the same area. A distinct peak would have provided additional support for the existence of the category boundary. The discrimination task in this study showed very small peaks for both groups suggesting a weak category boundary. Within-category discrimination (e.g., for two utterances with the *question* pattern) was higher for English speakers than Chinese speakers whereas between category discrimination (question versus statement) was similar for both speaker groups. The authors suggest that speakers may be able to categorically interpret F0 contours (as suggested by the existence of the S-curve) but not categorically perceive them (as suggested by the modest peaks of the discrimination task). In other words, when given two options, speakers are able to choose one of the two, but when asked if two utterances are the same or different, they have greater difficulty deciding.

Trimble (2013) tested the perception of Spanish broad focus statements and Y/N questions by L2 speakers whose L1 was English. The author tested the perception of two different varieties of Spanish: Toledo (Spain) and Mérida (Venezuela). As discussed in section 2.3.1.2, many varieties of Spanish make use of a final rise in the intonation contour to signal a Y/N question. This is true

of the Toledo variety as well. However, the Mérida variety of Spanish makes use of a final fall when signalling a Y/N question, making it unlike the other varieties of Spanish but also unlike English. Trimble (2013) tested 43 L2 speakers who were divided into two proficiency groups: second semester students and third/fourth-year students. The author assumed that the third/fourth year students would be at a higher proficiency level than the second semester students.[10] The third/fourth-year student group was further divided into those who had or had not spent time abroad in a Spanish speaking area (specifically Mérida or Toledo). The goals of this study were: a) to determine whether learners of Spanish could perceive the most common intonational cues that are used to distinguish broad focus statements and Y/N questions; b) to determine whether advanced learners are more accurate in their perception than beginner learners; c) to determine whether advanced learners are affected by input. Specifically, if advanced learners are exposed to a variety of Spanish that signals questions differently than their own L1, would they perform differently than other advanced learners who did not receive the same input? Participants listened to 20 pairs of lexically identical utterances and were asked if they heard a question or a statement. Overall, accuracy in the perception of Toledo Spanish was at 99% (99% for statements and 100% for questions) while accuracy in the perception of Mérida Spanish was at 72% (97% for statements and 44% for questions). Since questions in Mérida Spanish seemed to be difficult for L2ers, the author took a closer look at this portion of the data. He found that only the third/fourth-year students that had spent time abroad in Mérida were 67% accurate in the identification of Mérida Spanish questions. The other speaker groups (second semester, third/fourth-year students with no experience abroad, and third/fourth-year students with experience in Toledo) were only 33%, 53%, and 42% accurate in the identification of Mérida Spanish questions respectively. These differences

[10] No other measure of L2 proficiency was provided.

were statistically significant. Trimble found that overall, L2 learners had difficulty perceiving intonational cues that were not in their L1. In this study, Mérida Spanish questions which, unlike English, use a final fall in the intonation pattern to signal an interrogative utterance, were accurately perceived only 44% of the time by all L2 learners together. The advanced speakers who had spent time in Mérida, however, were accurate 67% of the time. In conclusion, only the advanced speakers who had received input that matched the stimuli were more accurate than the rest of the participants.

Radu, Colantoni, Klassen, Patience, Pérez-Leroux, and Tararova (2018) tested the perception of English declarative questions (DQ), absolute interrogative questions (AQ) (i.e., Y/N questions), and statements (S) by L1 Spanish speakers using three different tasks: an intonation-only task (participants only heard the intonation contours of utterances without the segmental content), a segmental and lexical information task (participants heard segmental and prosodic information of isolated sentences), and a contextualized task (participants heard a brief narration and were then asked to choose which of three possible utterances made the most sense to follow the narration). These authors were interested in the L2 speakers' overall accuracy including the role of context in the perception of English utterance types. The L2 speakers in this study were considered advanced speakers of English as they had been residing in the host country for approximately 9.9 years, had an average age of arrival of 14 years, and reported use of the L2 (English) in their day to day activities approximately 55% of the time. The Native English speakers outperformed the L2 speakers overall, obtaining higher accuracy scores on all tasks and utterance types (accuracy rate of 80% or more). A binomial mixed-effect model showed that L2 speakers performed better in the intonation-only and the segmental and lexical information tasks. Furthermore, of the three utterance types, the identification of declarative questions proved to be the most difficult. L2 speakers had more difficulty correctly identifying declarative questions in the contextualized task

(42% accurate compared to the intonation only-task and the segmental and lexical information task, 73% and 77% respectively). L2 speakers in this study were expected to have difficulty matching syntactic structures and intonation contours to the appropriate meaning and the results supported this prediction. The task with most context (the contextualized task involving matching the appropriate utterance to the narration) proved to be the most challenging with the lowest accuracy scores. The other two tasks that were more acoustic in nature, saw a higher accuracy rate. Declarative questions proved especially challenging for L2 speakers, especially in the contextualized task. Radu et al. (2018) also report on reaction times. They expected longer reaction times for structures and tasks that were more challenging for the L2 speakers. Overall, L2 speakers were slower than the controls (3412 ms versus 3171 ms respectively) in the intonation only and the segment and intonation tasks. Additionally, both groups reacted slightly faster to DQs than to AQs (authors do not provide times here). These differences were not statistically significant. The reaction times in the context task were noticeably longer than the other two tasks as they all clustered around 20 seconds. The authors, however, did not find any statistically significant difference between groups or utterance types.

Radu et al. introduced reaction time data into their L2 perception study as a way to quantify the difficulty of the tasks for the participants. The use of RT data in intonation is fairly new and infrequent. Longer reaction times in a particular task reflect a higher degree of difficulty (Jiang, 2012; Schneider et al., 2011). Reaction time data have been successfully used in categorical perception studies in L1 intonation (Chen, 2003; Schneider et al., 2011). To my knowledge, there are no other L2 intonation studies that consider RTs in Y/N question and statement distinction.

Gili Favela (2012) tested the perception of non-focus (or broad focus) and corrective (or contrastive) focus in English by L1 Italian speakers with "medium-to-low competence and exposure to English" (p. 25). In English, the non-focus (NF) peak is lower than the corrective focus

(CF) peak. In Italian the NF peak is higher than the CF peak. Furthermore, the NF peak is realized later in the syllable compared to its CF counterpart. This means that the CF pitch accent in Italian is similar to the NF and CF in English in terms of alignment (they are all realized at approximately the same place in the syllable) and similar to the NF in English for scaling (height). Gili Favela hypothesized that the Italian speakers would equate both English NF and CF pitch accents with the Italian CF category. She also hypothesized that the speakers would have difficulty discriminating between the two English pitch accents.[11] Three native Italian speakers and three native English speakers produced the recordings that served as the stimuli in this experiment. The author chose a representative production of both CF and NF for each speaker (6 speakers x 2 productions each = 12 utterances). Some of the stimuli were manipulated by means of prosodic-transplantation: the target word of the Italian sentences was extracted from the audio signal and the English prosody was imposed on it. This resulted in stimuli with L1 segmental and prosodic characteristics (NF-IT and CF-IT) and stimuli with L1 segmental but L2 prosodic characteristics (NF-IT_ENG and CF-IT_ENG). The target word that carried the pitch accent in question was "Guendalina" from the statement "Guendalina is meeting Monica" (IT - *Guendalina vede Monica.*). This statement can be produced with a NF pitch accent in the target word when it is the answer to the question "What's up?" (IT - *Che succede?*) and it can be produced with a CF pitch accent in the target word when it is the answer to the question "Is Maria meeting Monica?" (IT - *Maria vede Monica?*).

Participants (11 Italian native speakers) were given identification, goodness-rating, and discrimination-oddity tests. Reaction times were also measured as the author hypothesized that the

[11] The exact number of stimuli and the accuracy scores of this study are not reported in detail in Gili Favela (2012). The author refers the reader to another publication that would have focused exclusively on it. This publication appears to have been in preparation at the time but unfortunately it was never published. The main purpose of reporting on this study, however, is the contribution of reaction time data in L2 intonation.

cognitive load would be higher in processing the L2 stimuli and that would result in longer RTs. In identification and goodness-rating, participants were given a context (one of the two questions above) and were asked to identify the correct response or to determine whether what they heard was appropriate. In the oddity tests, participants were given a question with three possible answers and they had to point out which, if any, could not be used to answer the question. Identification scores were quite high overall. The L2 prosody stimuli received lower identification and goodness ratings overall, especially for English NF. Participants also had trouble discriminating English NF and CF (in line with the predictions). Finally, participants were successful in the oddity test, especially when they had to point out the CF item out of NF items. Reaction times overall were longer for stimuli in which L2 prosody was copied compared to stimuli with original L1 prosody. This difference was statistically significant. Furthermore, RTs were longer for NF stimuli with English prosody in comparison to NF (and CF) stimuli that showed Italian prosody, and they were also longer for CF stimuli with English prosody in comparison to CF stimuli with Italian prosody. The author concludes that RTs support the results on lower identification scores and goodness ratings and suggest a greater cognitive load in the processing of NF stimuli with English prosody. However, we do not have the accuracy score of those tests.

In sum, L2 learners are very successful with tasks designed to test their ability to discriminate tones and pitch patterns of varying heights in the absence of segmental information (Grabe, Rosner, García-Albea, & Zhou, 2003). They are also able to discriminate between Y/N questions and statements with relatively high accuracy rates (Liu & Rodriguez, 2012). Structures that are not present in the L1, however, may prove to be challenging in the L2 (Radu et al., 2018; Trimble, 2013). Tasks that involve linguistic meaning (e.g., segmental information accompanying the contour or the addition of a small narration prior to the perception task) result in additional difficulty for the L2 learners (Grabe et al., 2003; Radu et al., 2018). Reaction time data may provide

information about the difficulty of some of the linguistic structures under study, particularly when the accuracy rates are well above chance (Gili Fivela, 2012; Radu et al., 2018).

2.8 L2 Production of Intonation

The following paragraphs discuss studies on the production of L2 intonation that are relevant in predicting what the outcome of the current study may be. Each study is presented in detail and the section concludes with a summary of the relevant findings. The studies presented here consider different populations and different tasks, but each study contains a finding that works towards the formulation of hypotheses for the current study. All the production studies in this section investigate an L2 that is not English. Two of the studies focus on the L2 production of Y/N question intonation (Henriksen et al., 2010; Santiago Vargas & Delais-Roussarie, 2012) while the third study (Mennen, 2004) presents findings on the L2 phonetic realization of the pre-nuclear pitch accent. This section begins with the analysis of the L2 French (L1 Spanish) production of the final boundary tone in French Y/N questions (Santiago Vargas & Delais-Roussarie, 2012). The target language (French) has the option of realizing the final boundary tone in Y/N questions as either a rise or a fall but the findings show that L2 speakers do not make use of this variability and continue to use the only pattern that signals Y/N questions in their L1 (Mexican Spanish): final rising boundary tone. The second study (Henriksen et al., 2010) focuses on the pre- and post-study abroad production of Y/N questions and statements of L2 Spanish speakers. This study seeks to find a developmental sequence in L2 intonation by measuring the most consistent patterns used by each speaker before and after the study-abroad program. Despite the seemingly homogenous group of L2 speakers, Henriksen et al. find inter-speaker variability. Finally, the last study (Mennen, 2004) of this section is concerned with the L2 realization of the pre-nuclear peak in Greek. This study

finds that even very advanced L2 speakers are not always able to reach native speaker values when it comes to the alignment of the pre-nuclear peak in the L2.

Santiago Vargas and Delais-Roussarie (2012) tested the production of final contours and prosodic phasing of L2 French speakers (L1 Mexican Spanish) with a French proficiency of A2 or B1 within the Common European Framework of Reference for Languages.[12] Y/N questions are generally described as having a final rise in both French and Spanish. In French, however, speakers can also use a level or falling final contour in Y/N questions that have subject-verb inversion or make use of the *est-ce que* interrogative particle. The authors investigated the learners' production of three types of French Y/N questions (declarative, subject-verb inversion, particle *est-ce que*). Through a reading task and interactive oral production (participants were interviewed and asked to talk about their projects, their experience in French classes, etc.), 301 Y/N questions were extracted from the production of Spanish-speaking French learners as well as Native French and Spanish speaker controls. The L2 French speakers in this study made exclusive use of the final rising contour in all three possible Y/N question structures of French. The authors suggest that the L2 speakers are influenced by their L1 in selecting a final contour for Y/N questions in French. These Mexican Spanish speakers exclusively use a final rise in their L1 to signal a Y/N questions and they continue to use a final rise in their L2 (French) to signal Y/N questions. While Santiago Vargas and Delais-Roussarie's study analyzes an L2 that is not English, thus making a new empirical contribution, and also considers the production of Y/N questions, a few observations about this study are warranted. The proportion of final rise to final fall in French Y/N questions is not equally distributed. In fact, the authors claim that the canonical description of Y/N questions

[12] Levels A2 and B1 correspond to beginner and intermediate levels called "basic user" and "independent user" respectively in the Common European Framework of Reference for Languages (see *The CEFR Levels*, 2020 for a detailed description of the levels).

in French is a final rise. However, the native French speaker data that they reported in their study were characterized by flat and falling contours predominately in spontaneous speech. Furthermore, the L2 speakers in this study were taking French classes at the Autonomous University of Mexico at the time of testing but there is no information given about the type of input they received. Based on the expectations of this study, the researchers seem to have assumed that their L2 speakers received exposure to all the possible contours that are used to signal Y/N questions in French. This is unclear given the learners' experience as described by the researchers.

Henriksen, Geeslin, and Willis (2010) analyzed the intonation contours of Spanish declaratives and absolute interrogatives (i.e. Y/N questions) as realized by four English-speaking study abroad students.[13] The goal of the study was to measure the changes in production before and after the study abroad program. Specifically, this study considered the differences in the realization of the initial pitch accent, the final pitch accent, and the final boundary tone for each utterance type. The study abroad program took place in Spain in a full immersion environment. Vocabulary and grammar tests prior to the study abroad program revealed that these participants had no comprehension barriers in the L2.[14] The authors described the learners as having "extensive experience with Spanish but limited real world experience with native speakers of the language" (p. 125). No other proficiency measures were provided. Using a contextualized reading task in which participants read a series of target sentences that were preceded by a discourse context, the

[13] Henriksen, Geeslin, and Willis (2010) also analyzed pronominal interrogatives (wh-questions). However, since they are not relevant here, they will not be discussed here.

[14] Despite being enrolled in university Spanish courses, these L2 speakers were given a grammar test to "gauge the formal grammatical knowledge of each participant" (p.127) and to ensure their knowledge was comparable. The grammar test consisted of a paragraph where certain elements were in multiple choice format. The participants had to choose the best response given the context. The grammar test focused primarily on subjunctive mood, copula contrast, pronoun agreement, and aspect in past tense forms. Participants were also given a vocabulary test to ensure that they were familiar with the words that they would need in order to produce the target sentences. They were tested on 11 vocabulary items. For each word, they were asked to choose a synonym from three possible options.

authors compared the learners' production of declaratives and absolute interrogatives at the beginning and at the end of their study abroad program and found noticeable changes overall. Specifically, the degree of consistency of the patterns increased following the 7-week program. The authors were interested in the most frequent intonational strategy used by each learner for each sentence type before and after the study-abroad program. Henriksen, Geeslin, and Willis sought to measure change in intonation patterns over time in order to describe the developmental trends that exist in second language intonation. While it is difficult to speak of developmental trends with so few participants in such a small amount of time, the between-speaker variation in the L2 intonation of these four speakers is of interest here. One speaker did not show any change in pre- and post-study abroad production as she did not make use of the typical rising pre-nuclear pitch accent in neither declaratives nor interrogatives. A second speaker very noticeably made use of the pre-nuclear rise after the study abroad program, especially with absolute interrogatives where the use of the pre-nuclear rise was present in 97% of her utterances (and was not present prior to going abroad). A third speaker implemented a final falling contour for declaratives in approximately 48% of the utterances, where rises were predominant prior to her stay abroad. The same speaker's absolute interrogative production, however, simply saw an increase in the expected pattern as she was already producing it (66% before versus 88% after). The fourth speaker also had predominantly final rises in declaratives before the stay abroad but produced final falls 50% of the time after the program. This same speaker's interrogative production was already target-like prior to going abroad and stayed consistent after the 7-week period. While the authors of this study were interested in developmental trends in L2 intonation, their findings show just how variable L2 production of intonation can be across a seemingly homogeneous group of L2 speakers.

While pre-nuclear peak height in L2 Y/N questions and/or statements has not been studied to my knowledge, studies have looked at pre-nuclear peak alignment in L2 intonation. Peak height and

alignment are often linked because speakers have been found to replace one with the other (Gussenhoven, 2004). Mennen's (2004) study, for example, looked at bi-directional interference in the intonation of Dutch speakers of Greek. Her study focused on the pre-nuclear pitch accent in declarative utterances which is LH* for both languages. She wanted to determine whether Dutch-speaking learners of Greek could achieve target-like production of the L2 pre-nuclear pitch accent while still realizing natively the pre-nuclear pitch accent in their L1. Dutch and Greek differ, however, in the phonetic realization of this pre-nuclear peak: in Greek, the peak is realized on the vowel of the post-accentual syllable while in Dutch it is realized on the accentual syllable. Furthermore, since Dutch has a phonological contrast between long and short vowels, this peak will occur earlier in long-vowel than in short-vowel syllables. Mennen's study involved 5 native Greek speakers, 5 native Dutch speakers, and 5 L1 Dutch-L2 Greek speakers. The bilingual speakers were highly proficient in their L2 (Greek) as they had extensive experience with the L2 (between 12 and 35 years), held a university degree in Modern Greek Language and Literature, and were teaching Greek at the university level at the time of testing. Participants read declarative sentences that were matched as closely as possible in Greek and Dutch. There were 20 declarative sentences in Greek and 40 declarative sentences in Dutch, 20 with a short vowel in the target word and 20 with a long vowel in the target word. The target words are not provided in the study, but they are described as having lexical stress on the antepenultimate syllable, followed by 2 to 5 unaccented syllables, and with predominantly sonorant sounds with a few cases of voiced obstruents. The participants were instructed to read the sentences as naturally as possible. Mennen found that the pre-nuclear peaks produced by the Dutch speakers of Greek were earlier than those

of the native Greek speakers.[15] The L2 Greek peak was produced before the offset of the target vowel. One L2 speaker, however, emerged as having a pattern comparable to native Greek. In fact, this participant's peak realization fell within the norm of Greek speakers by realizing the peak after the offset of the target vowel. These findings were all statistically significant. Mennen suggests that the reason for this difference is that this participant had been exposed to Greek at an earlier age than the other participants (15 compared to 20-25). However, the author cannot be sure that this was the only factor given that a number of factors could have been responsible for this particular speaker's success. This study illustrates the difficulty of phonetic realization of pre-nuclear peaks even in very advanced L2 speakers. Furthermore, it provides another example of the variation present in L2 speech production.

The studies discussed above suggest that the L1 has a strong influence on the production of L2 intonation. In some cases, L2 learners use what is similar in their L1 even when the L2 provides more options for the same structure (Santiago Vargas & Delais-Roussarie, 2012). In other cases, even very advanced L2 speakers have difficulty reaching native speaker values (Mennen, 2004). The L2 production of intonation also shows great variability between speakers. L2 speakers with seemingly similar profiles have shown between-speaker variability in production at different proficiency levels and in different language pairs (Henriksen et al., 2010; Mennen, 2004).

2.9 The Relationship Between the L2 Perception and Production of Intonation

There is a general consensus among researchers that there is a relationship between perception and production but the nature of this link is debated (Gambi & Pickering, 2018). In L2 segmental

[15] The same speakers (L1 Dutch-L2 Greek) were also tested in their L1 (Dutch) and the author found that while the speakers maintained a contrast between long and short vowel syllables, this difference in timing was not as large as for the native speakers of Dutch.

acquisition, accurate perception before accurate production is the most common finding (e.g., Rochet, 1995) but researchers have suggested other possible relationships such as accurate production occurring before accurate perception (e.g., Sheldon & Strange, 1982) or accurate perception and production occurring in a non-linear, somewhat independent way (e.g., Bohn & Flege, 1990) (see Llisterri, 1995 for overview). In L2 intonation, only a small number of studies have considered this relationship explicitly. These studies vary in language pairings, structures, part of contour analyzed, and tasks. The following paragraphs explore these studies in greater detail.

Some studies on L2 intonation have found that accurate perception occurs before accurate production. An earlier version of the study to be presented in Chapter 3 (Marasco, 2014) was conducted with advanced L1 English-L2 Spanish speakers. This study sought to determine whether accurate perception was a pre-requisite for accurate production. The prediction, based on the majority of L2 segmental studies, was that accurate perception would be a pre-requisite for accurate production. Five L2 Spanish speakers (L1 English), who all started learning Spanish as adults, participated in the study. The participants were advanced speakers of Spanish who were graduate students involved in teaching Spanish grammar to undergraduates at the time of testing. The perception task was a categorical perception task in which the stimuli were Spanish Y/N questions and statements (e.g., *¿Tienen helado de chocolate?* 'Do they have chocolate ice cream?'; *Tienen helado de chocolate.* 'They have chocolate ice cream.') with pre-nuclear peak or initial boundary tone modification. One native Mexican Spanish speaker recorded the stimuli and average values for the initial boundary tone and pre-nuclear peaks were calculated. The modifications were based on those average values. For example, a statement would be modified to be more question-like by raising the peak of the pitch accent or boundary tone to the average value in Hz for questions of the speaker who recorded the stimuli. Similarly, a question would be modified to be more

statement-like by lowering the peak of the pitch accent or boundary tone to the average value in Hz for statements. The final inventory of utterances used in the perception task was the following: 1) Y/N question or statement without modifications; 2) statements with raised initial boundary tone; 3) statements with raised pre-nuclear peaks; 4) Y/N questions with lowered initial boundary tones; 5) Y/N question with lowered pre-nuclear peaks. The final boundary tone was removed from all the stimuli. In the discrimination (AX) task, the participants listened to 96 pairs of utterances. While participants were 100% accurate when the two utterances were identical (32 pairs of utterances), they accurately identified *different* utterances correctly 74% of the time overall. Furthermore, pairs that had similar peak heights were identified as similar utterances more often than those that had similar boundary tones. In the identification task (*question or statement?*), participants listened to 72 utterances. The 24 original utterances (12 Q and 12 S) each had one original version, a peak modification version, and one boundary tone modification version (24 x 3 = 72). Utterances that were more statement-like (unmodified statements and statements with a raised initial boundary tone) were identified as statements 70%-90% of the time while utterances that were more question-like (unmodified questions and questions with a lowered initial boundary tone) were identified as questions between 70% and 100% of the time. Participants had greater difficulty identifying utterances with a modified pre-nuclear peak (questions with a lowered peak or statements with a raised peak) as the cues were more conflicting. Identification rate hovered around 50%.

The production task included a series of contexts narrated by a native Spanish speaker in which one key utterance was missing. Participants listened to the narration and were asked to fill in the missing utterance as if they were the person saying it. Each context was designed to elicit a Y/N question or a statement. The target-like result would have been a higher initial boundary tone and pre-nuclear peak in the question with respect to the comparable statement. Only one of the five

participants showed native-like differences in production. Since there were no participants who were accurate in production but not in perception, the findings suggest that accurate perception of intonation is acquired before accurate production by L2 learners.

Ortega-Llebaria and Colantoni (2014) investigated how the presence or absence of context could shape the L2 perception and production of intonation. In particular, tasks with more context were predicted to be characterized by greater L1 transfer effects than those with less or no context based on previous research that found that tasks focused primarily on acoustic perception resulted in less L1 transfer compared to tasks with more context or higher levels of processing. With respect to the relationship between L2 perception and production of intonation, these researchers expected to find more evidence of L1 transfer in production if target-like perception precedes target-like production in L2 intonation. Therefore, any evidence of L1 transfer in perception would also be observed in production but not the opposite. Participants (L2 English-L1 Spanish or Mandarin) took part in two perception tasks (with context and without context) and two production tasks (also with context and without context).

The no-context perception task was a forced choice identification task in which participants heard isolated utterances differing in focus placement. They were asked to match these utterances to one of three possible sentences that differed only in intonation. These three utterances were low pass filtered to remove all segmental content. The task required participants to match one of the three low pass filtered F0 contours to the F0 contour of the isolated sentence. In the context perception task, participants heard a question and three possible answers related to a story that they had previously heard. The three answers differed in focus (on the subject, verb or object) but only one was appropriate for the question asked. For example, in a context that was designed to elicit focus on the verb, the participant would hear: *Did Froggy sleep in his jar?* The options presented to the participant would be: 1) FROGGY escaped from the jar; 2) Froggy ESCAPED from his jar; 3)

Froggy escaped from his jar. The expected answer was 2). Ortega-Llebaria and Colantoni found that participants performed in a more target-like manner in the no context condition and that Mandarin speakers behaved more like native English controls than Spanish speakers.

In the no-context production task, participants were asked to imitate de-contextualized utterances differing in focus placement. In the context condition, participants were prompted to answer questions related to a story that they had previously heard. The responses required focus to be placed on the subject, verb or object and only one answer was appropriate based on the question. The structure was identical to the context perception task described above. The only difference was that the participants here had to answer the question themselves instead of choosing from three possible focus conditions. The authors found that overall L2 speakers were more target-like in the no-context condition where focus on one element of the utterance had no real meaning because the utterance was decontextualized. Based on their L1, participants showed different degrees of transfer. Specifically, Spanish speakers inserted pauses after focalized elements creating a new intonation phrase while Mandarin speakers' performance was comparable to English native speakers. Given the fact that L1 transfer was more evident in production than in perception, particularly in the case of the Spanish speakers, the authors suggest that target-like perception occurs before target-like production in the intonation of L2 speakers.

The opposite trend (accurate production before accurate perception) was found by Puga, Fuchs, Hudson, Setter, and Mok (2018) who explicitly set out to find a relationship between the L2 perception and L2 production of intonation. They tested German learners of British English on a variety of sentence types: statement, continuation, statement question, echo question, Y/N question, wh-question, closed tag, open tag, checking tag, sarcasm, and checking. The participants were advanced learners of English enrolled in a Bachelor's or Master's degree in English literature and linguistics at the time of testing. German and English are typologically similar as they are both

intonation languages with stressed-time rhythm and use pitch in a systematic way to mark different syntactic structures. There are limited differences in tone inventory and the meaning of tones between these two languages. Even though studies have shown that German and English seem to share a common tone inventory of phonological representations, the two languages differ in the way these representations are realized phonetically. For example, the alignment of the peak might be realized on different vowels or for a longer duration.

Puga et al.'s study analyzed the nuclear contour of the sentence types mentioned above. The expected patterns for English were the following: fall for statement, wh-question, and closed tag; rise for statement question, Y/N question, open tag, and checking tag; fall-rise or rise for echo question and checking; level or rise for continuation; and rise-fall or fall for sarcasm.[16] In the production task, participants were asked to read a short story consisting of 28 sentences that contained all the utterance types mentioned above. The perception task was a forced choice task where participants were presented with each of the sentences from the story with which they had become familiar in the production task. Each sentence was presented with 5 tonal variations of the nuclear portion of the utterance. Participants were asked to indicate which of the contours was the best fit. Forty percent of the participants were equally accurate in both tasks while 18% were not accurate in either task. The remaining 41% of participants were divided into those who were accurate in production but not in perception (27%) and those who were accurate in perception but not production (14%). The number of participants who were accurate in production but not in perception was almost double the number of those who were accurate in perception only. Given these findings, the authors suggest that there is no evidence of an implicational relationship between correct perception and correct production in L2 intonation.

[16] The terms rise, fall, rise-fall, fall-rise, and level belong to the British school of transcription.

Zárate-Sández (2015) tested the perception and production of the intonational contour of unmarked, non-emphatic declarative utterances in Spanish. He sought to a) understand how English learners of Spanish perceive Spanish declaratives; b) how English learners of Spanish produce Spanish declaratives; and c) assess the nature of the relationship between perception and production in the intonation patterns of L2 speakers. His target population was L2 speakers of Spanish (L1 English) of low, high, and very high proficiency. He tested both pre-nuclear peak alignment and final boundary tone. Perception was tested via an imitation task in which participants had to listen to and reproduce stimuli that had been manipulated along a continuum.[17] If the perception of the stimuli was categorical, the responses would cluster into two categories. If the perception was gradient, the responses would fall along a continuum. The pre-nuclear peak was manipulated to create 10 stimuli with peaks at different points within or just after the stressed syllable. Perception of peak alignment was categorical (i.e., responses clustered into two categories) for all groups except the low proficiency L2 and Native English speakers. L2 speakers who showed a clear difference between early and late peak alignment had understood the difference in Spanish (early peak = narrow focus). The manipulation of the final boundary tone consisted of raising or lowering it in order to create 10 stimuli spanning 90 Hz. Participants' perception of the final boundary tone was more gradient. Broad increments were perceived in similar ways by all participants. The more proficient L2 speakers patterned with the native speakers. To elicit production, speakers engaged in sentence reading and storytelling tasks. The unmarked, non-emphatic declarative utterances from these tasks would be considered native-like

[17] Zárate-Sández (2015) addresses the problematic issue of using speech production to make claims about speech perception. He suggests, however, that one claim has been put forth suggesting that imitation will be constrained by the speaker's grammar such that the speaker will only reproduce what is permitted in his or her grammar (p. 166). Considering that the perception task in this study yielded significant and categorical effects, the author is inclined to support this task for perception and encourages replication research and further testing.

if late pre-nuclear peak alignment and a falling final boundary tone were produced. In both tasks, more proficient learners aligned pre-nuclear pitch accents progressively later compared to less proficient learners. In final boundary tone production, low proficiency speakers produced final boundary tones at a higher point in their range (i.e., not falling therefore not target-like). This trend decreased and their production became more target-like as proficiency increased. Finally, Zárate-Sández looked at the relationship between perception and production in L2 intonation. He ran a correlational analysis and found only the results of pre-nuclear peak perception and production to be significant. Specifically, he found that the later the categorical shifts occurred in perception, the later the peaks were aligned in production.

The studies discussed above reveal various possible relationships between perception and production of L2 intonation. Marasco (2014) and Ortega-Llebaria and Colantoni (2014) found accurate perception before accurate production, Puga et al. (2018) found that L2 speakers could be accurate in production before reaching the same levels of accuracy in perception, and Zárate-Sández (2015) found a correlational relationship. What ties these studies together is the goal to investigate the relationship between L2 perception and production of intonation. Unfortunately, these studies vary in several ways making it very difficult to formulate any kind of generalization or hypothesis. First of all, these studies consider different structures, from simply looking at one utterance type (Zárate-Sández, 2015) to considering a variety of them (Puga et al., 2018). Furthermore, these studies look at different parts of the contour with some focusing on the beginning (Marasco, 2014), others on the end (Puga et al., 2018), and some on more than one part of the contour (Zárate-Sández, 2015). Perception tasks varied from forced choice (Marasco, 2014; Ortega-Llebaria & Colantoni, 2014; Puga et al., 2018) to imitation (Zárate-Sández, 2015) and production tasks ranged from reading tasks (Zárate-Sández, 2015), to semi-spontaneous tasks (Marasco, 2014; Ortega-Llebaria & Colantoni, 2014; Puga et al., 2018) to free conversation tasks

(Zárate-Sández, 2015). The language pairings, structures, part of contour analyzed, and tasks are very different in all these studies. These differences make it very difficult to compare these studies in a systematic way.

2.10 Summary

The latter half of this chapter began by exploring well known speech learning models: Flege's (1995) Speech Learning Model, Best and Tyler's (2007) Perceptual Assimilation Model applied to L2, and Mennen's (2015) L2 Intonation Learning Theory. The former two models were developed with the primary focus on L2 segmental acquisition and have been applied to L2 vowel (e.g., Flege et al., 1999; Levy, 2009) and consonant (e.g., Flege et al., 1995; Rose, 2010) acquisition but make some assumptions that are applicable to L2 intonation, namely 1) L2 speech production errors are perceptually motivated and 2) L1 and L2 categories exist in a common phonological space that can result in the assimilation or dissimilation of the L2 sound to the L1 category. An example of how some of the assumptions in the SLM and PAM-L2 can be applied to L2 suprasegmental acquisition come from Trofimovich and Baker (2006) who found that their L2 English (L1 Korean) participants did not show similar patterns across 5 different suprasegmentals. They suggested that some of these suprasegmentals may be more prominent than others and therefore more easily acquired. This finding is compatible with Flege's idea of what L2 speakers may perceive as similar or new with respect to the L1 sound inventory. Despite the possibility of applying some of the assumptions from the speech models designed for L2 segmentals, there is a need for a model that is primarily focused on L2 intonation. This is because the acquisition of L2 intonation involves language-specific intonation-meaning relationships. To this end, Mennen (2015) has proposed a model that allows researchers to focus on different dimensions which aid in the cross-linguistic comparison between L1 and L2 intonation identifying

possible areas of difficulty in L2 intonation. In addition to phonetic and phonological dimensions that are typical of segmental studies, Mennen's model proposes a semantic and a frequency dimension to account for these language-specific intonation-meaning relationships. As Mennen states however, this is still a relatively new model and more empirical data is needed. As a result, we discussed some L2 intonation research to provide additional information.

The studies discussed in the latter part of this chapter have discussed L2 intonation in perception, production, and the relationship between the two. In perception, L2 learners are able to successfully discriminate tones and pitch patterns of varying heights in the absence of segmental information (Grabe, Rosner, García-Albea, & Zhou, 2003). They are also successful in tasks that require them to discriminate between Y/N questions and statements (Liu & Rodriguez, 2012). Some studies have found, however, that a source of difficulty for L2 listeners may come from the presence of a structure that does not occur in the L1 (Radu et al., 2018; Trimble, 2013) or a task that involves linguistic meaning (e.g., segmental information accompanying the contour or the addition of a small narration prior to the perception task) (Grabe et al., 2003; Radu et al., 2018). Additional information about the difficulty of some of the linguistic structures may be extracted from reaction time data (Gili Fivela, 2012; Radu et al., 2018). In production, studies have shown that the L1 has a strong influence on L2 intonation. Some L2 learners have opted to use structures that are similar to their L1 even when the L2 provides more options for the same structure (Santiago Vargas & Delais-Roussarie, 2012). Some L2 learners, even at very advanced proficiency levels, continue to have difficulty reaching native speaker values (Mennen, 2004). Finally, L2 production of intonation studies have revealed between-speaker variability among speakers with seemingly similar profiles. Between-speaker variability has been found at different proficiency levels and in different language pairs (Henriksen et al., 2010; Mennen, 2004).

Studies that have set out to explicitly test the relationship between L2 perception and production of L2 intonation suggest different possibilities for this type of relationship: Marasco (2014) and Ortega-Llebaria and Colantoni (2014) found accurate perception before accurate production, Puga et al. (2018) found that L2 speakers could be accurate in production before reaching the same levels of accuracy in perception, and Zárate-Sández (2015) found a positive correlation between perception and production but not across all parts of the intonation contour. These studies vary in language pairings, structures, part of contour analyzed, and tasks, making it challenging to systematically compare them or make predictions about what the relationship between L2 perception and production in intonation may be.

Having reviewed some of the influential speech models (SLM, PAM-L2, and LILt) and having discussed the relevant L2 intonation research, Chapter 3 will restate the three research questions introduced in Chapter 1 along with their respective hypotheses that are informed by both the models and the empirical studies. The rest of the chapter will present the methodology used in the current study.

Chapter 3
Research Questions, Hypotheses, and Methodology

3.1 Research Questions and Hypotheses

The present chapter discusses the research questions, hypotheses, and methodology of the current study. The participants in this study took part in a perception task and a production task that will be described in greater detail in section 3.4. We begin this chapter with the three research questions that were introduced in Chapter 1 and, based on the research presented in Chapter 2, formulate hypotheses. Section 3.2 presents some methodological issues in intonation and provides the reader with the motivation of the choices made here. Section 3.3 introduces the participants in detail while sections 3.4.1 and 3.4.2 describe the perception and production tasks respectively. Finally, the chapter concludes with the experimental protocol and the data analysis, the results of which will be presented in Chapter 4.

Question 1: Are L2 learners of Spanish able to accurately perceive the higher initial boundary tone and pre-nuclear peak of Spanish Y/N questions with respect to comparable statements?

Hypothesis 1: The L2 learners in this study will be successful (accurate) in perceiving the higher initial boundary tone and pre-nuclear peak of Spanish Y/N questions with respect to comparable statements. Factors that could affect accuracy rates are discussed below.

If we apply Flege's (1995) SLM to the current study, it predicts that accurate L2 perception is the result of the creation of a new category for an L2 distinction (question versus statement) that is not present in the L1 inventory. Conversely, according to the PAM-L2 (Best & Tyler, 2007), if the question-statement distinction in Spanish is assimilated into what English uses this height distinction for (i.e., not to signal question versus statement), this will be considered poor

assimilation. In the cross-linguistic comparison of the Y/N question-statement distinction, Mennen's (2015) LILt predicts difficulties in the semantic and frequency dimensions for both the initial boundary tone and the pre-nuclear peak. English and Spanish differ in the semantic dimension because a higher initial boundary tone, when present, signals impatience or insistence in American English (Bartels, 1999) whereas it signals a Y/N question in Spanish (Sosa, 1999). With respect to the frequency dimension, the high initial boundary tone is uncommon in American English (M. E. Beckman et al., 2005), but is reported as part of the intonation contour of Y/N questions in Spanish. With respect to the pre-nuclear peak, a height difference has been found to distinguish normal from emphatic utterances in English (Chen, 2003; Ladd & Morton, 1997) while it has been reported as part of the intonation contour of Y/N questions in Spanish (Face, 2007; Sosa, 1999). While these models identify the potential areas of difficulty for the L2 speaker in the perception of this specific L1, the empirical studies, that were discussed in Chapter 2, provide additional information with respect to the type of results we may obtain from these L2 speakers. Liu and Rodriguez (2012) found that when L2 speakers were given a choice between Y/N questions and statements, they were highly accurate. These authors, however, expressed some concern with respect to their findings because they found lower than expected discrimination rates among their L2 speakers. They proposed that this finding could be explained by the fact that when L2 speakers are given two options, they are able to choose one of the two, but when they are asked if two utterances are the same or different, they have greater difficulty deciding. Some intonation studies have proposed the use of RTs as a way to investigate these difficulties further. One such study is Gili Favela (2012) who found that her L2 participants (L1 Italian) had difficulty between broad focus and contrastive focus pitch accents in English and this difficulty was reflected in the longer RTs. Radu et al. (2018) and Trimble (2013) found that when an intonation pattern signals something in the L2 that is not signaled in the same way in the L1, L2 speakers have greater

difficulty perceiving it accurately. In this study, the difficulty lies in the fact that the question-statement distinction is not signaled by a higher pitch in English as it is in Spanish. As a result, if RTs are an indication of the difficulty, L2 speakers in this study will be expected to have longer RTs than NS and the identification of Y/N questions is expected to require more time than statements in L2 speakers.

Question 2: Are L2 learners of Spanish able to accurately produce a higher initial boundary tone and pre-nuclear peak of Spanish Y/N questions with respect to comparable statements?

Hypothesis 2: L2 speakers are expected to produce a higher initial boundary tone and pre-nuclear peak in *their own* Y/N question production with respect to *their own* statement production. However, these differences in L2 production are not expected to reach native speaker differences. Some between-speaker variation is expected.

If the findings reveal no difference between the production of Y/N questions and statements in L2 Spanish, the SLM (Flege, 1995) would suggest that this inaccurate production is a result of the fact that the L2 speakers have not yet established a category for this (phonetic) difference. Similarly, Mennen's (2015) LILt supports the idea that production inaccuracies are perceptually motivated. The PAM-L2 (Best & Tyler, 2007) exclusively focuses on perception suggesting that their findings may serve to complement Flege's SLM. In this regard we may assume that the PAM-L2 indirectly supports the idea that inaccurate production is a result of inaccurate perception. While L1 patterns have been found to affect L2 production of intonation (Henriksen et al., 2010; Mennen, 2004; Santiago Vargas & Delais-Roussarie, 2012), L2 speakers have been able to realize intonation patterns that are not in their L1 (Mennen, 2004). Despite seemingly homogenous L2 speakers, between-speaker variability has occurred (Henriksen et al., 2010; Mennen, 2004). While Henriksen et al. (2010) do not discuss the potential source of the variability in their L2 Spanish

(L1 English) speakers, Mennen (2004) suggests that her L2 Greek (L1 Dutch) speakers' production may be affected by the age of first contact with the L2.

Question 3: What is the relationship between perception and production of the initial boundary tone and pre-nuclear peak in L2 Spanish speakers?

Hypothesis 3: Participants are expected to show a relationship between perception and production in the form of a positive correlation.

Both Flege's (1995) SLM and Mennen's (2015) LILt expect errors in production to be perceptually motivated. In this study, participants are expected to be accurate in perception as well as be able to signal the Y/N question-statement distinction in production as explained in hypotheses 1 and 2 above. More accurate perception will therefore correlate with more target-like production.

3.2 Methodological Issues in Intonation

Designing an experiment for intonation research is not without its challenges. A discussion of the motivation behind the design of the tasks follows. In the perception task, an effort was made to limit the manipulation and modification of the stimuli because speech samples with modified F0 contours still conserve information from the original contour in the harmonics. In fact, the harmonics carry so much information that the F0 need not be physically present for pitch to be perceived (Nooteboom, 1997). It follows that a task designed to test the perception of a modified stimulus, may in fact be eliciting a participant response about the original stimuli instead. Furthermore, participants are sensitive to unnatural sounding stimuli (Face, 2007). That is to say that participants' responses may be influenced by the modification more than expected and this could have effects on the results. Since both factors could complicate the findings of a study, we opted here for a perception task that did not alter the F0 contour.

Highly structured production tasks where participants are asked to repeat words or utterances in isolation are also not ideal because they may not produce generalizable findings. Hedberg and Sosa (2002) and Face (2003) found that the intonation patterns of scripted experimental settings were noticeably different from intonation patterns in spontaneous speech. Specifically, Hedberg and Sosa (2002) found unexpected (compared to the literature) patterns in the final portion of the Y/N question intonation contour while Face (2003) found noticeable differences in F0 peak alignment, downstepping, final lowering, and de-accenting. Generalizability is important because it allows researchers to determine how much of what we observe in an experiment is applicable to everyday speech learning. An alternative to a highly scripted task is a semi-spontaneous or controlled elicited production task in which participants are guided in such a way that similar structures and vocabulary are produced across participants. This type of task results in two benefits: 1) the data are comparable between participants and 2) the findings may become generalizable.

3.3 Participants

A total of 31 participants were tested in this study at the University of Toronto over the course of one summer. Nine participants were L1 Spanish speakers and 22 participants were advanced L1 English-L2 Spanish speakers.[18] The 9 native Spanish speakers (5 female and 4 male), who served as controls in this study, had a mean age of 27.4 (range: 21-35) at the time of testing. They originated from different Spanish-speaking countries including Argentina (n = 3), Venezuela (n = 2), Mexico, Uruguay, Chile, and Spain (n = 1 from each). They were all intermediate speakers of English who had received some formal education in English prior to arriving to Canada.

[18] Some L2 participants reported being heritage or bilingual speakers of languages that are prosodically unrelated to Romance languages (specifically Cantonese, Mandarin, Tamil, and Somali).

The 22 L1 English-L2 Spanish speakers (14 female and 8 male) who participated in this study had a mean age of 27.8 years (range: 20-62) at the time of testing. All the L2 speakers had begun learning Spanish as adults (mean age of initial contact: 17 years; range: 10-31) receiving on average 57 months of formal Spanish language instruction (range: 6-120) in Canada. Most L2 participants reported having had both native and non-native Spanish instructors. Most L2 participants also reported receiving native speaker input while spending time abroad in a variety of Spanish speaking countries including Mexico, Cuba, Spain, Colombia, Argentina, Guatemala, Chile, Peru, Ecuador, El Salvador, and Costa Rica. Time spent abroad ranged anywhere from 3 months to 3 years and did not include time spent on vacation in these countries. L2 participants also reported their mean use of Spanish at school, home, work, and social settings on a daily basis. On average, the L2 speakers used Spanish 20% of the time (range: 3% to 42%). The L2 participants were asked to self-evaluate their proficiency in Spanish on a scale with four possible options: beginner, intermediate, advanced, and near-native. Eleven L2 speakers (50%) rated themselves as advanced speakers, 7 L2 speakers (32%) rated themselves as near-native, and 4 L2 speakers (18%) rated themselves as intermediate. The overall proficiency self-ratings were transformed into a numerical score ranging from 1-4, such that 1 corresponds to beginner, 2 to intermediate, 3 to advanced, and 4 to near-native (Radu, 2020; Radu et al., 2018). Table 3.1 presents a summary of the participants' information (see Appendix A for details on individual participants).

Table 3.1.

Participant Profiles

	Native Speakers			L2 Speakers		
	Mean	SD	Range	Mean	SD	Range
Age	27.4	6.3	21-35	27.8	9.7	20-62
Age of first contact with SPA	-	-	-	17	4.1	10-37
Use of Spanish (%)	-	-	-	20%	14	3% - 42%
Formal Spanish Language Instruction (months)	-	-	-	57	36.9	6-120
Self assessment	-	-	-	3.1	0.7	2-4

The oral proficiency of these L2 speakers was also measured via an accentedness rating test. Participants read The North Wind and The Sun (Appendix D) at the beginning of the experiment (details on experimental protocol to follow). These recordings were then rated by five native speaker judges who were unaware of the goal of the study. Each judge was presented with the all the L2 recordings as well as the native speaker recordings. Within each set, the order of speakers was randomized. The judges were asked to rate the recordings from 1 (very strong accent: definitely non-native) to 5 (no foreign accent at all: definitely native) (as in Bongaerts et al., 2000). The judges were told that they would hear 30-second passages read by native and non-native speakers of Spanish, but they were not informed about the proportion of native and non-native speakers. The average rating for each participant is reported in Table 3.2.

Table 3.2

Native speaker judges' evaluation of oral Spanish proficiency for all participants on a scale of 1 (very strong accent) to 5 (no foreign accent at all)

Native Speakers		L2 Speakers					
Speaker	Mean	Speaker	Mean	Speaker	Mean	Speaker	Mean
NS01	5	AS01	3.6	AS14	3.6	AS24	3.4
NS02	5	AS02	3	AS15	2.4	AS25	1.8
NS03	5	AS04	2.2	AS16	2.6	AS26	4.6
NS04	4.8	AS05	3.8	AS17	2.8	AS29	1
NS05	4.5	AS06	3.4	AS18	2.8		
NS06	4.8	AS07	2.8	AS19	2		
NS07	5	AS10	4.8	AS20	1.8		
NS08	5	AS12	1.4	AS22	3		
NS09	4.8	AS13	3	AS23	1.4		
NS Group Mean: 4.9				L2 Group Mean: 2.8			

As can be seen in table 3.2, the native Spanish speaker group received a higher accentedness score than the L2 speaker group. A Mann-Whitney test confirmed that the difference between the native speaker group and the L2 group was statistically significant: *U = 194.5, p. < .001*.

While lexical/grammatical competence was not tested for the L2 speakers, we obtained three different measures that aimed to offer a more complete picture of the L2 speakers target language competence. A correlational analysis was carried out between self-ratings and percentage of use of Spanish, self-ratings and accentedness scores, and accentedness scores and percentage of use of Spanish. A Spearman correlation test revealed a positive correlation (r_s (20) = .44, p = .03) between accentedness scores and percentage of use of Spanish (Figure 3.1).

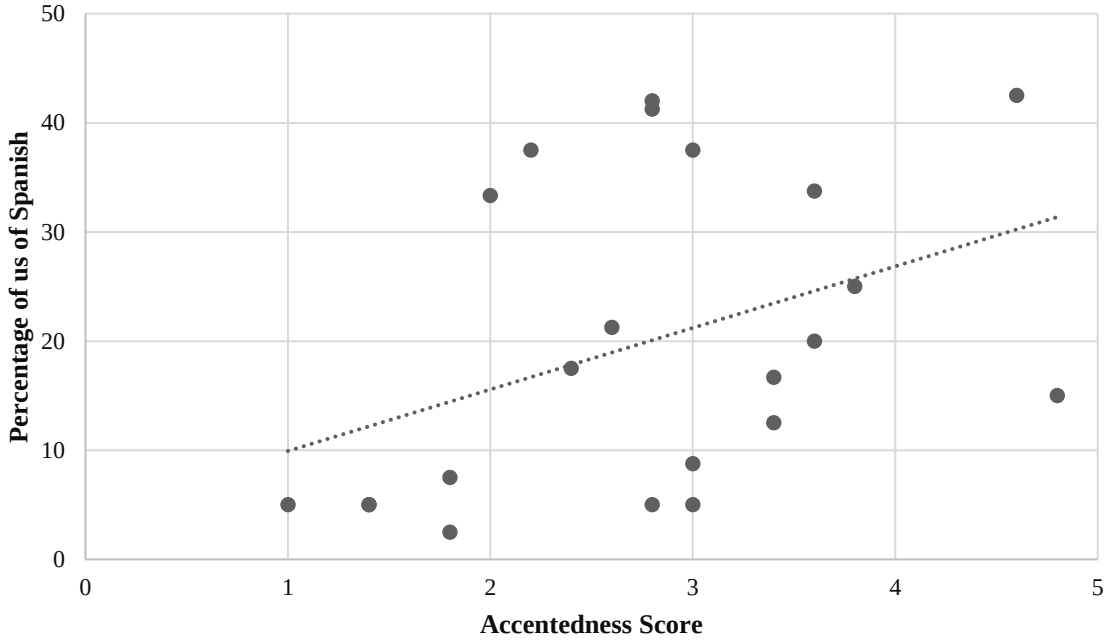

Figure 3.1. Correlation between L2 speaker accentedness scores and percentage of use of Spanish.

Having discussed the participants' profiles, we now turn to the description of the perception and production tasks in the current study.

3.4 Tasks

3.4.1 Perception Task

The perception task was a gating experiment based on the gating paradigm used by Grosjean (1980) to study word recognition processes. The gating structure has been used primarily in L1 intonation studies (Falé & Faria, 2006; Face, 2005, 2007; Petrone & Niebuhr, 2014; Saindon et al., 2017; Vion & Colas, 2006) as a way to isolate and focus on one part of the intonation contour. Twelve pairs of utterances (Y/N question and comparable statement) were selected and divided into three gates with each gate corresponding to a specific section. The gates were divided as follows: **Gate 1**: after the unstressed syllable following the first stressed syllable of the utterance in order to capture the peak of the pitch accent which is generally realized on the post-tonic syllable (Face, 2007; Prieto & Roseano, 2010; Sosa, 1999); **Gate 2**: before the nuclear pitch accent and final boundary tone which, when all other cues are contradictory or absent, signal utterance type in many varieties of Spanish and English (Face, 2007; Grabe et al., 2003); **Gate 3**: entire utterance (see Appendix B for full list of utterances and their gate divisions). Figure 3.2 illustrates the gate division of the Y/N question *¿Tienen helado de chocolate?* 'Do they have chocolate ice-cream?' spoken by a native Mexican Spanish speaker.

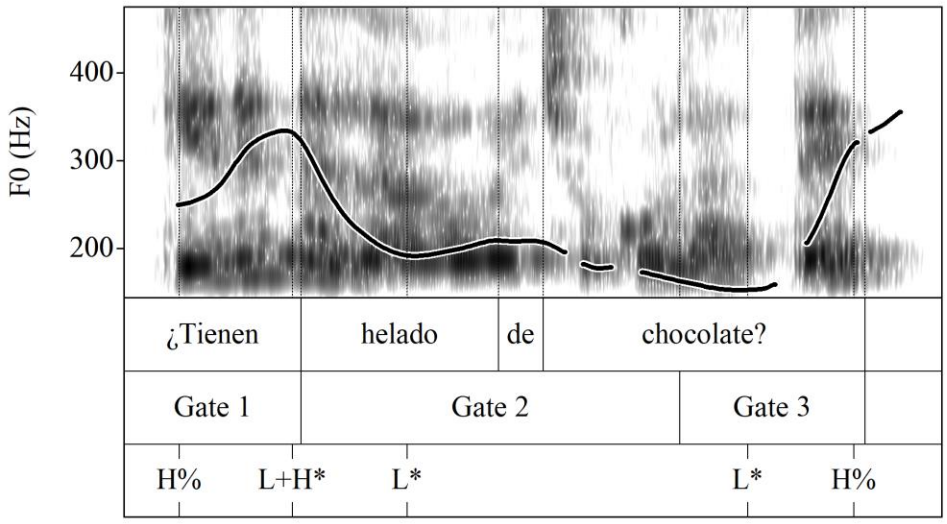

Figure 3.2. ¿Tienen helado de chocolate? 'Do they have chocolate ice-cream?' spoken by a native Mexican Spanish speaker. The three gates are indicated as well as the initial boundary tone (H%), the pre-nuclear peak (L+H*), other pitch accents (both L*), and the final boundary tone (H%).

These three gates are important points in the utterance. Gate 1 contains the height difference that contrasts Spanish Y/N questions from both their statement counterparts and English Y/N questions (see Chapter 2, section 2.3.1). Gate 2 contains everything except the nuclear pitch accent and the final boundary tone. Finally, the entire utterance (Gate 3) served as a control to determine whether the listener could distinguish statements from Y/N questions overall.

The perception task started with participants listening to a brief narrative (approximately 15 seconds) which was intended to allow them to establish the speaker's pitch range (details about perception stimuli are presented in the following paragraph). This was followed by a practice round that contained two statements and two wh-questions. The participants heard increments of the utterance and, at each increment (gate), they saw a screen that asked them to indicate whether they heard a question or a statement by pressing a specific key on the keyboard. The participants did not receive any feedback about their answers. Then, the participants performed the task with the

experimental stimuli. The gating experiment involved a total of 24 utterances (12 Y/N questions and 12 comparable statements) presented in random order. The time gap between gates was set at 200ms (following Grabe, Rosner, García-Albea, & Zhou, 2003) because anything less than 200ms may not be perceived as an interruption or a break in the speech (Nooteboom, 1997). The time gap between utterances was set at 1000ms. Reaction times were recorded at each gate. The perception task was administered through SuperLab 4.5 (2015).

3.4.1.1 Acoustic Analysis of Perception Stimuli

The perception task stimuli were produced by a female native speaker of Mexican Spanish in her early 40s with a university education who was unaware of the goal of the study. She was asked to read brief narratives that contained the target utterance 3 times (the narratives were identical to the ones used in the production task; see Appendix C for the full list). The utterance used to create the perception stimuli was taken from the second repetition. The first repetition was not usable because the reader was not familiar with the material. There was a great deal of hesitation which resulted in many pauses in her speech and many inquiries ("am I doing this correctly?"). The second repetition was smoother. The speaker was more comfortable with the task and with being recorded. The third repetition was done in the event that the quality of the second round of recording was compromised but, in the end, it was not needed.

Figure 3.3 shows the mean frequency of the initial boundary tone and the pre-nuclear peak of all the stimuli. This acoustic analysis was performed on the perception stimuli in order to ensure that the stimuli presented the same patterns that are reported in the literature (and in Chapter 2) concerning the relative height of the initial boundary tone and pre-nuclear peak.

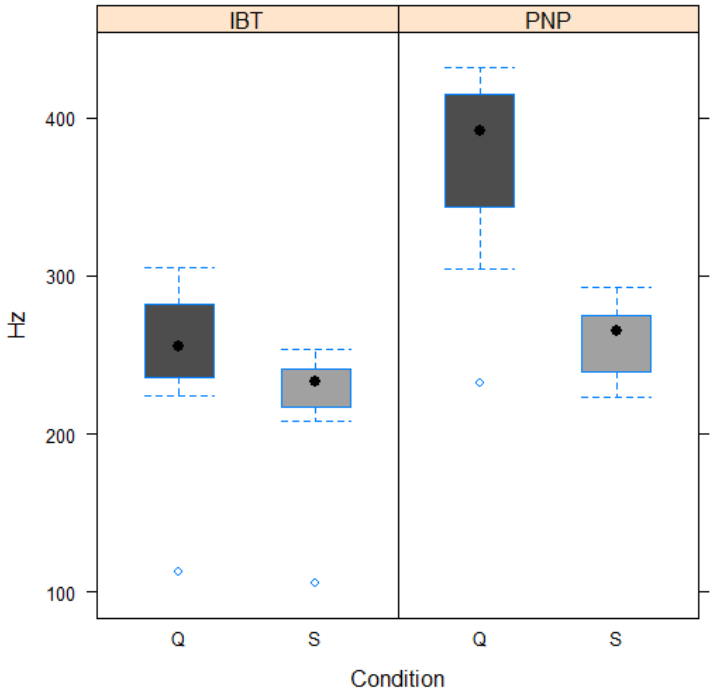

Figure 3.3. Mean frequency (Hz) of the initial boundary tone (IBT) and pre-nuclear peak (PNP) of the stimuli. For each pair, Y/N questions (Q) are on the left in dark grey and statements (S) are on the right in light grey.

For both the initial boundary tone and the pre-nuclear peak, the mean frequency was greater in Y/N questions than in statements. Specifically, the mean frequency of the initial boundary tone in questions was 250 Hz (*SD* 51) versus 222 Hz (*SD* 39) in statements. The mean frequency of the pre-nuclear peak was 373 Hz (*SD* 60) in questions versus 260 Hz (*SD* 22) in statements. The measured values are consistent with the patterns reported for Spanish Y/N questions and statements (see Chapter 2, section 2.3.1.2 for details). A Wilcoxon Signed-Rank test indicated that both the initial boundary tone and pre-nuclear peak height differences (28Hz and 113Hz respectively) were statistically significant, **IBT**: $Z = -2.19$, $p < .05$; **PNP**: $Z = -3.05$, $p < .001$. These values were also converted into semitones to establish their perceptual relevance. The difference between questions and statements in the initial boundary tone was 2.05 semitones while the difference between questions and statements in the pre-nuclear peak was 6.25 semitones. A

difference of 1.5 semitones or greater is considered reliably discriminable by the human ear (Nooteboom, 1997) which suggests that these differences in the perception stimuli were expected to be perceptible to the listeners.

3.4.2 Production Task

The production task was designed to elicit semi-spontaneous productions of comparable Y/N questions and statements in Spanish. For each item, participants listened to a scenario narrated by a male native speaker of Colombian Spanish. He was from Bogotá, in his late 30s, university educated, and unaware of the purpose of the study. The narrative was accompanied by images and one utterance during the narrative was masked by a barking dog (Cournane, 2014). The participant was asked to produce the missing utterance as if s/he were the person saying it. The prompt was an explicit question asking what the character in the story said. The narratives were structured so that the target utterance would be predictable. There were 24 scenarios in total (12 designed to elicit statements, 12 designed to elicit comparable Y/N questions; see Appendix C for the full list of scenarios). Figure 3.4 below provides an example.

The participant hears:	The participant sees:
Hace calor. Un papá llega con su hijo a la heladería. Al niño sólo le gusta el helado de chocolate así que el papá le pregunta al empleado de la heladería: 'It's a hot day. A father and son go to the ice-cream store. The son only likes chocolate ice-cream so the father asks the worker:'	
****dog barking****	Same as above
Desafortunadamente no tienen y entonces el niño dice que no quiere nada. 'Unfortunately, they don't have chocolate ice-cream and the boy says he doesn't want anything.'	
¿Qué le dice el papá al empleado? 'What does the father ask the worker?'	***¿Qué le dice el papá al empleado?***

Figure 3.4. One of the production task items designed to elicit a Y/N question.

The figure shows which images appeared at the different points of the narrative. The production task was administered through PowerPoint.

3.4.3 Experimental Protocol

All participants started the session by reading the Spanish version of the fable *The North Wind and the Sun* (Appendix D), which was then used for accentedness ratings (as we saw in section 3.3). The production task followed, and the entire session was recorded. Before the perception task, the participants were asked to fill out a language questionnaire (Appendix E). Next, they carried out

the perception task. Finally, participants were asked to speak freely about their favourite food for 30 seconds. The experiment lasted 45 minutes and participants were remunerated $15 CDN.

3.5 Data Analysis

The data obtained from the perception task was analyzed for overall accuracy and reaction times were also recorded. The realization of the initial boundary tone height and pre-nuclear peak height were measured in Hertz and the differences were transformed into semitones in order to be able to compare across speakers (Gut, 2013; Nolan, 2003). Annotation, as shown in the bottom tier of Figure 3.2, was carried out based on the inventory described in section 2.3.1.2. The results of the analysis of both tasks follow in Chapter 4.

Chapter 4
Results

4.1 Introduction

This chapter presents the results for the perception and production tasks. The analysis here focuses specifically on the perception and production of the initial boundary tone and pre-nuclear peak (information contained in gate 1). As we have seen in Chapter 2, there are differences between English and Spanish Y/N questions and comparable statements in the initial portion of the intonation contour of these two types of utterances. Specifically, Y/N questions in Spanish have a higher initial boundary tone and pre-nuclear peak compared to their statement counterparts (section 2.3.1.2). Before looking at the results, the three research questions and accompanying hypotheses that were presented in Chapter 3 are repeated below.

Question 1: Are L2 learners of Spanish able to accurately perceive the higher initial boundary tone and pre-nuclear peak of Spanish Y/N questions with respect to comparable statements?

Hypothesis 1: The L2 learners in this study will be successful (accurate) in perceiving the higher initial boundary tone and pre-nuclear peak of Spanish Y/N questions with respect to comparable statements.

Question 2: Are L2 learners of Spanish able to accurately produce a higher initial boundary tone and pre-nuclear peak of Spanish Y/N questions with respect to comparable statements?

Hypothesis 2: L2 speakers are expected to produce a higher initial boundary tone and pre-nuclear peak in *their own* Y/N question production with respect to *their own* statement production. However, these differences in L2 production are not expected to reach native speaker differences. Some between-speaker variation is expected.

Question 3: What is the relationship between perception and production of the initial boundary tone and pre-nuclear peak in L2 Spanish speakers?

Hypothesis 3: Participants are expected to show a relationship between perception and production in the form of a positive correlation.

All the statistical tests in this chapter were performed in R (R Core Team, 2017) using the lme4 package (Bates et al., 2015) for mixed models.

4.2 Perception Results

This section presents perception accuracy scores and reaction times at gate 1 for the nine native speakers and twenty-two L2 speakers who participated in the study.

4.2.1 Perception Accuracy

Accuracy is operationalized here as the correct identification of a question or a statement at gate 1. Overall, the native Spanish speakers and L2 speakers performed better than chance (50% accuracy) in the identification of Y/N questions and statements at gate 1. Figure 4.1 shows that native speakers were 74% accurate in question identification and 69% accurate in statement identification while L2 speakers were 80% accurate in question identification and 78% accurate in statement identification.

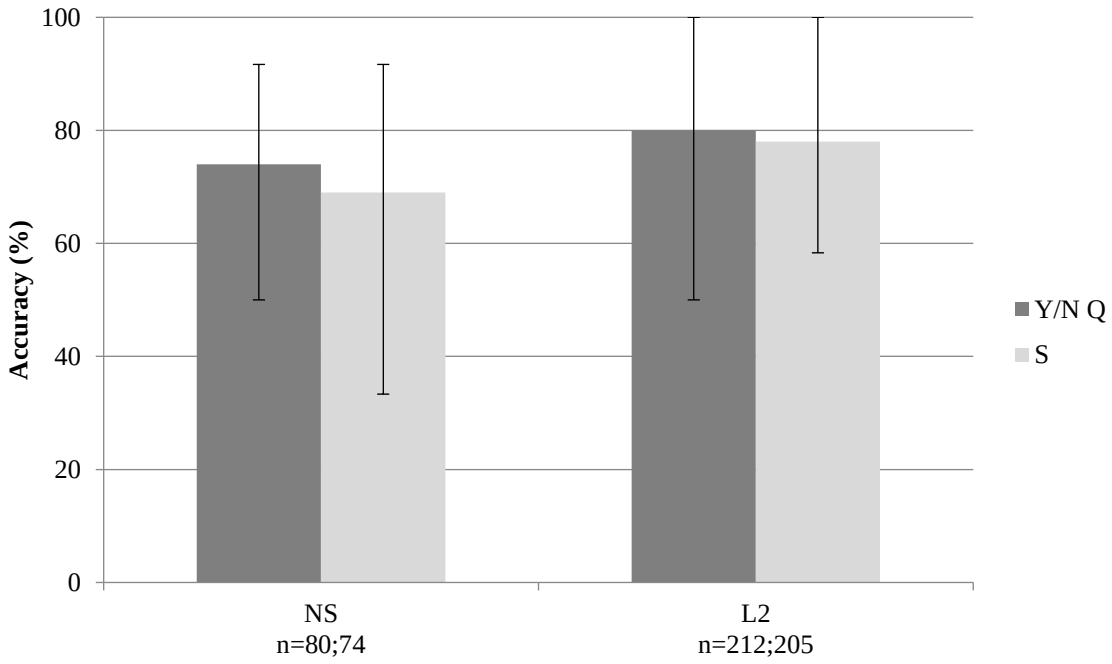

Figure 4.1. Mean percentage of accurate responses for the native speaker (NS) and L2 speaker groups at gate 1 by utterance type. For each group, Y/N questions (dark grey bars) are compared to statements (light grey bars). The number of accurate responses (maximum 108 for NS, 264 for L2ers) are also indicated. The bars indicate the minimum and maximum percentage of accuracy for each bar.

Because condition (question or statement) was expected to have a different effect on the two groups (NS and L2), a generalized linear mixed effects analysis of accuracy as a function of condition by group was performed. Utterance and participant were random effects. No significant effects were found (Table 4.1). This finding suggests that both groups were equally accurate in the identification of questions versus statements.

Table 4.1

Generalized mixed effects model for accuracy as a function of condition by group

Fixed Effects	Beta estimate	Standard error	z-value	p-value
Intercept	1.511	0.211	7.150	< .001
Condition (Statement)	-0.169	0.218	-0.775	.439
Group (NS)	-0.434	0.316	-1.371	.170
Condition*Group	-0.070	0.378	-0.186	.852

A range in accuracy responses was attested at gate 1 in both groups (Figures 4.2 and 4.3 for native speakers and L2 speakers respectively) (see Appendix F for the number of accurate responses at all gates per participant).

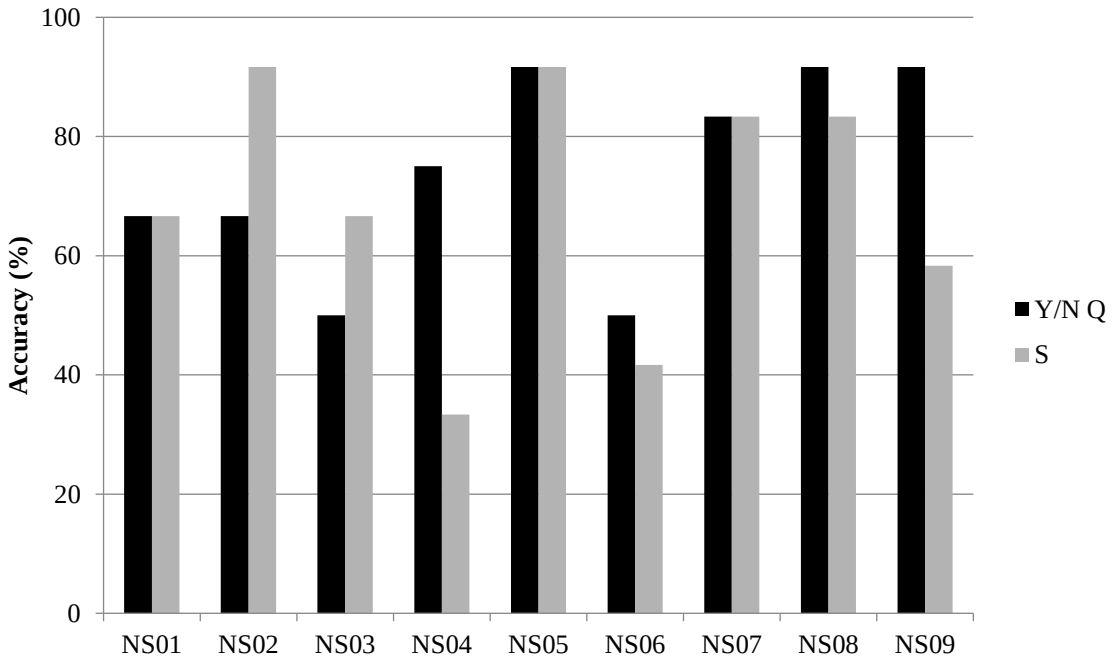

Figure 4.2. Individual NS accuracy (%) in identifying utterance type (Y/N Question versus Statement) at gate 1.

Among the native speakers (Figure 4.2), NS01, NS05, and NS07 were equally accurate in identifying questions and statements (67%, 92%, and 83% respectively) while NS02 (67% versus 92%), NS03 (50% versus 67%), and NS08 (92% versus 83%) differed by only a few answers (out of a maximum of 12 for questions and statements respectively).

Two native speakers had noticeably different accuracy rates between Y/N questions and statements. NS04 was 75% accurate (9 out of 12) in the identification of questions but only 34% (4 out of 12) accurate in the identification of statements while NS09 was 92% (11 out of 12) accurate in identifying Y/N questions but only 58% (7 out of 12) accurate in identifying statements. One native speaker's accuracy rates were at chance or just below; NS06 was 50% (6 out of 12) accurate in the identification of Y/N questions and 42% (5 out of 12) accurate in the identification of statements. These three speakers did not share any characteristics (age at the time of testing, gender, country of origin, or number of months spent learning English formally). Details are presented in Table 4.2.

Table 4.2

Native speakers who presented noticeable differences between Y/N question accuracy and statement accuracy

Speaker	Age	Gender	Country of Origin	Formal Instruction English
NS04	33	F	Venezuela	156 months
NS09	32	F	Chile	96 months
NS06	22	M	Mexico	156 months

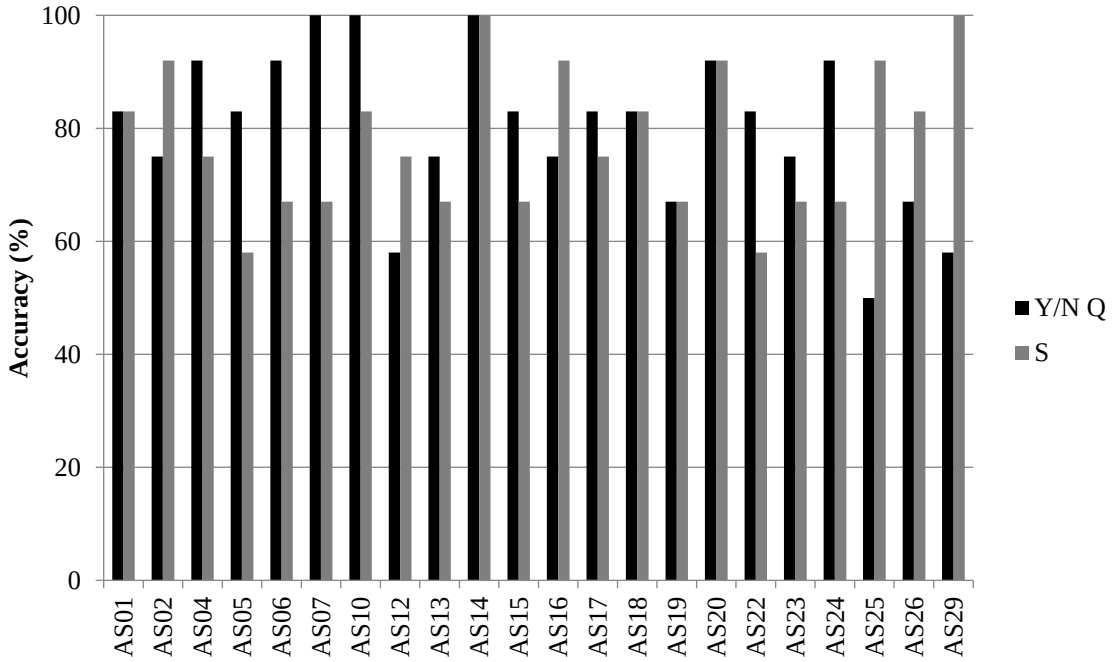

Figure 4.3. Individual L2 accuracy (%) in identifying utterance type (Y/N Question versus Statement) at gate 1.

Among the L2 speakers (Figure 4.3), AS01, AS14, AS18, AS19, and AS20 were equally accurate in identifying questions and statements (83%, 100%, 83%, 67%, and 92% respectively). The following speakers differed by only a few answers (out of a maximum of 12 for questions and statements respectively): AS02 (75% versus 92%), AS04 (92% versus 75%), AS05 (83% versus 58%), AS06 (92% versus 67%), AS07 (100% versus 67%), AS10 (100% versus 83%), AS12 (58% versus 75%), AS13 (75% versus 67%), AS15 (83% versus 67%), AS16 (75% versus 92%), AS17 (83% versus 75%), AS22 (83% versus 58%), AS23 (75% versus 67%), AS24 (92% versus 67%), and AS26 (67% versus 83%).

Two speakers had noticeably different accuracy rates between Y/N questions and statements. AS25 was only 50% accurate (6 out of 12) in the identification of questions but 92% (11 out of 12) accurate in the identification of statements while AS29 was only 58% (7 out of 12) accurate in identifying Y/N questions but 100% (12 out of 12) accurate in identifying statements. These L2

speakers were similar in age at the time of testing, age of first contact with Spanish, percentage of Spanish use, number of months of formal Spanish instruction, and native speaker judge rating. Their details are reported in Table 4.3.

Table 4.3

L2 speakers who presented noticeable differences between Y/N question accuracy and statement accuracy

Speaker	Age	Gender	Age of first contact with SPA	Percentage of Spanish use	Formal Instruction Spanish	NS judge rating
AS25	23	F	18	2.5%	24 months	1.8
AS29	20	M	18	5%	18 months	1

It must be noted, however, that these two speakers share some of these characteristics with other L2 speakers who did not show the same pattern in perception.

An analysis between the L2 speaker variables (accentedness, age at the time of testing, age of first contact with Spanish, self-rating, percentage of Spanish use, months of formal education in Spanish) and accuracy rates was also conducted. A Pearson correlation test revealed a positive correlation ($r\ (20) = .46$, $p = .02$) between question accuracy and accentedness scores (Figure 4.4) while a Spearman correlation test revealed a positive correlation ($r_s\ (20) = .45$, $p = .03$) between question accuracy and percentage of use of Spanish (Figure 4.5).

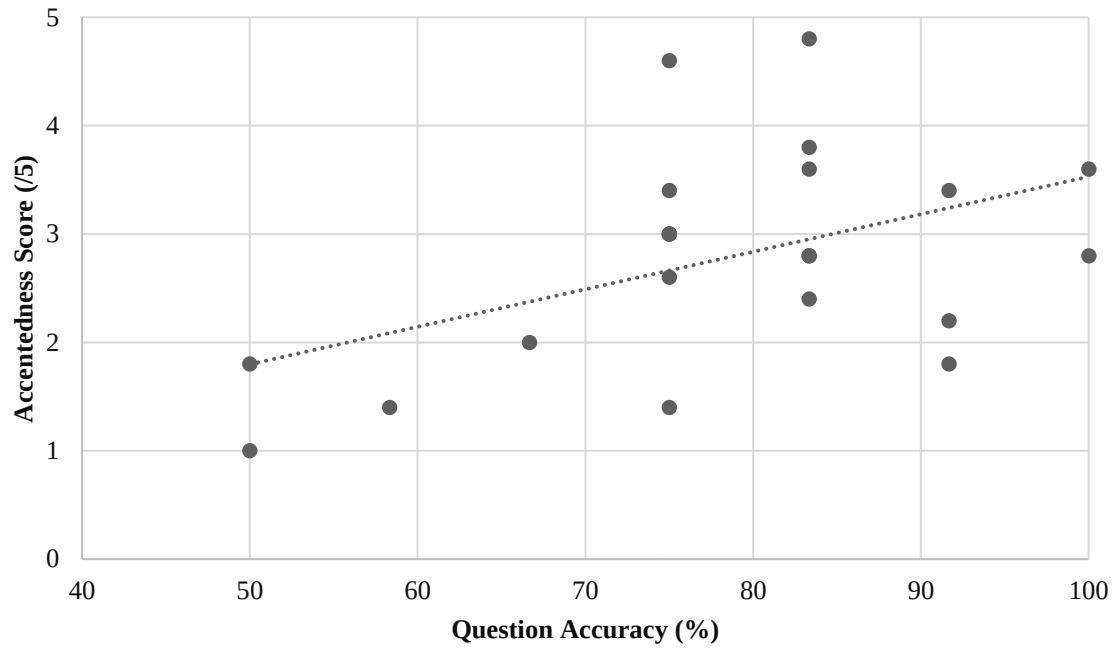

Figure 4.4. Correlation between question accuracy rate and accentedness score for L2 speakers.

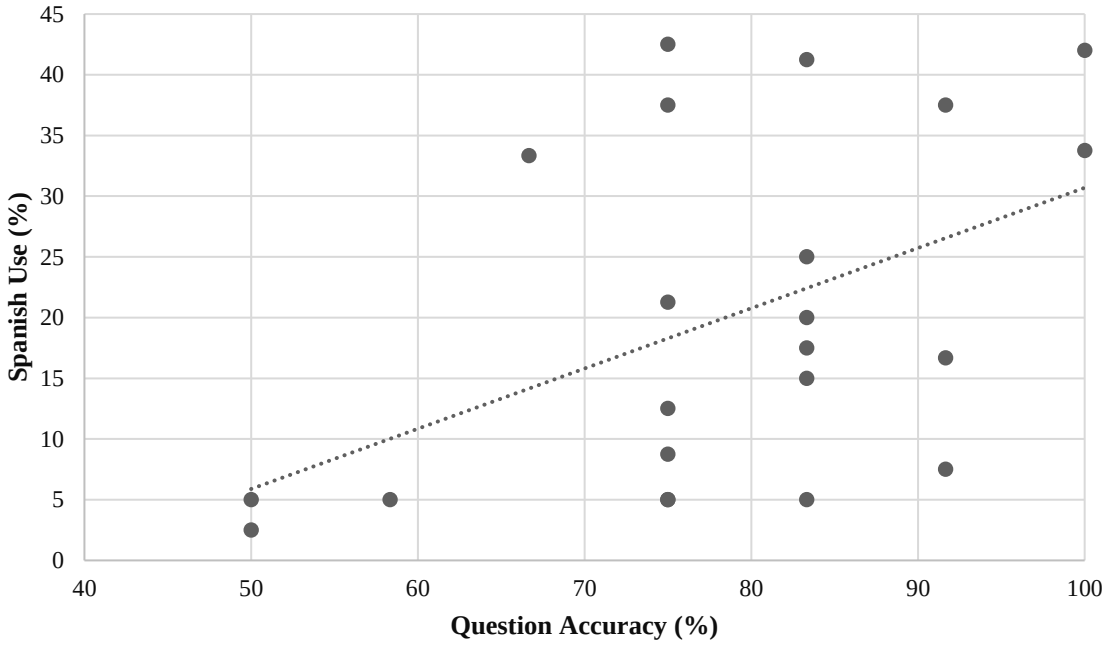

Figure 4.5. Correlation between question accuracy rate and percentage of use of Spanish for L2 speakers.

4.2.2 Perception Reaction Times

The reaction time (RT) data reported here are for accurate responses only (as is usually the case with RT data, Jiang, 2012). A total of 569 tokens (152 native speaker tokens and 417 L2 speaker tokens) were used. Three tokens that were over 6000 ms were eliminated because they were a result of distraction on the part of the participant. The rest of the data were not screened further for outliers because generalized linear mixed models were used for statistical analysis and these models do not require a strict 2SD cut-off from the mean (Baayen & Milin, 2010). Furthermore, the reaction time values presented here were not transformed because 1) data transformation has the potential to lessen the impact of outliers (Whelan, 2008) and 2) unless the researcher has theoretical reasons to transform the reaction time data (e.g., significant age difference in groups that could ultimately have an impact on motor skills), the data should remain in its original raw form (Lo & Andrews, 2015). Figure 4.6 shows the mean group RTs for Y/N questions and statements for the native speakers and L2 speakers at gate 1.

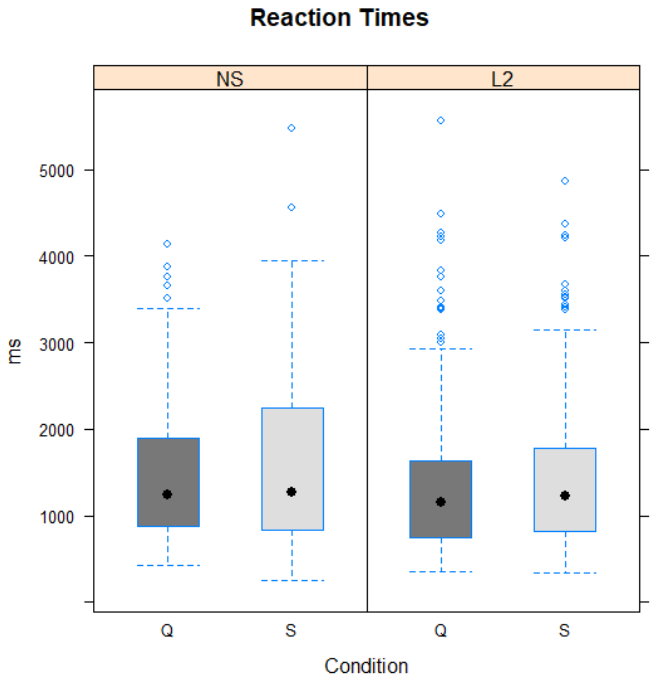

Figure 4.6. Mean group reaction times (ms) at gate 1 for native speakers (left panel) and L2 speakers (right panel). For each pair, question RTs (Q) are on the left (in dark grey) and statement (S) RTs are on the right (in light grey).

As shown in Figure 4.6, identifying questions and statements required approximately the same amount of time for both groups. Native speakers showed no real difference between question and statement identification while L2 speakers required slightly more time for statements than questions. The means (with their respective standard deviations) and the medians for each group and each condition are presented in Table 4.4.

Table 4.4

Mean (with standard deviations) and median RT values for Y/N questions and statements for both native and L2 speakers.

	Question		Statement	
	Mean (SD)	Median	Mean (SD)	Median
NS	1570 ms (941)	1240	1629 ms (1096)	1273
L2	1394 ms (905)	1153	1448 ms (871)	1231

Because condition (question or statement) was expected to have a different effect on the groups (NS and L2), a linear mixed effects analysis of reaction time as a function of condition by group was performed. Utterance and participant were random factors. No significant effects were found (Table 4.5).

Table 4.5

Mixed effects model for RT as a function of condition by group

Fixed Effects	Beta estimate	Standard error	t-value	p-value
Intercept	1375.60	106.84	12.875	< .001
Condition (Statement)	86.66	81.45	1.064	.287
Group (NS)	42.82	195.51	0.730	.465
Condition*Group	48.04	157.51	0.368	.712

Individual results are shown in Figures 4.7 (native speakers) and 4.8 (L2 learners). In these figures, boxplots show the distribution of RT values for questions and statements of each participant (see Appendix G for reaction time values at all 3 gates for all participants).

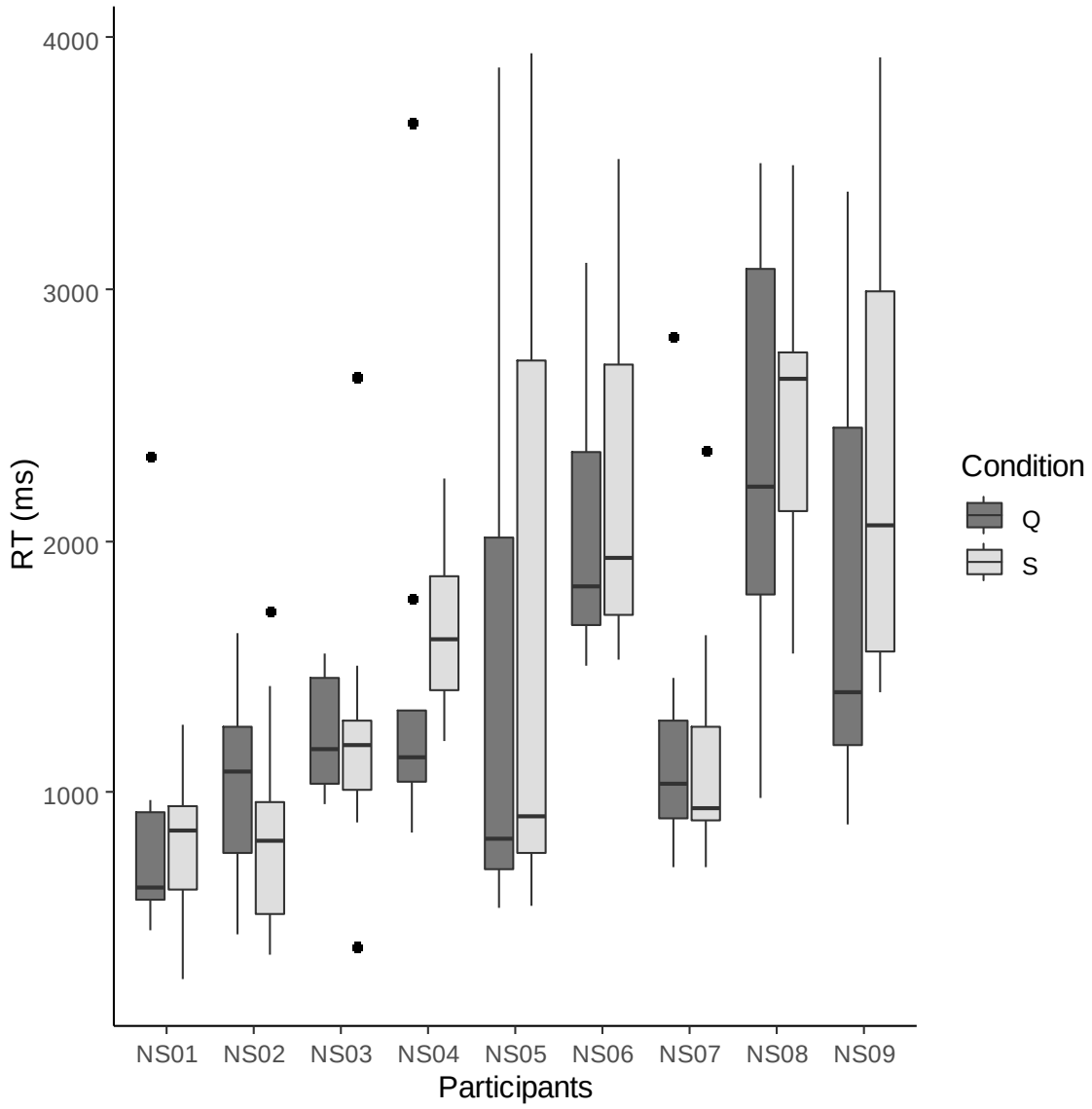

Figure 4.7. Native speaker RTs (in ms) for questions (dark grey) and statements (light grey).

The native speaker group (Figure 4.7) was expected to show similar values between question and statement RTs. While NS02 required more time to identify questions versus statements and speakers NS04, NS08, and NS09 required more time to identify statements versus questions, there were no statistically significant differences *between* question RTs and statement RTs of each individual native speaker (see Appendix H for individual values).

In the L2 speaker group (Figure 4.8), AS02 and AS06 required more time to identify questions than statements. Speakers AS15, AS18, AS19, AS20, and AS22 required more time to identify statements versus questions. A t-test or a Wilcoxon signed rank test (based on the normality of the distribution) on each pair of data revealed that only AS06, AS18, and AS20 had statistically significant differences (**AS06**: $Z = 32$, $p = .05$; **AS18**: $Z = 8$; $p = .04$; **AS20**: $t(10) = -3.15$, $p = .01$) (see Appendix H for individual values). These three L2 speakers did not share similar characteristics in terms of age at the time of testing, age of first contact with Spanish, percentage of use of Spanish, native speaker judge rating, or months of formal instruction of Spanish. Their characteristics are summarized in Table 4.6.

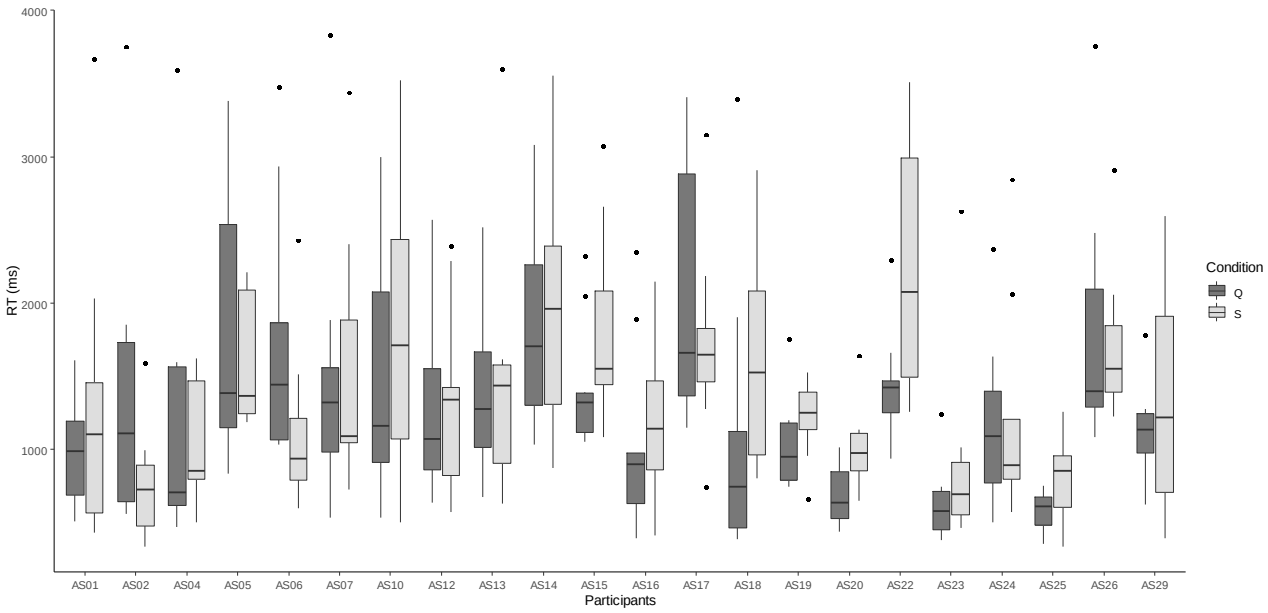

Figure 4.8. L2 learner RTs (in ms) for questions (dark grey) and statements (light grey).

Table 4.6

L2 speakers who had statistically significantly different RT values in the identification of Y/N questions versus statements

Speaker	Gender	Age	Age of first contact with Spanish	Percentage of Spanish use	Formal Instruction Spanish	NS judge rating
AS06	F	33	17	12.5%	60 months	3.4
AS18	M	25	19	5%	12 months	2.8
AS20	F	23	16	7.5%	12 months	1.8

An analysis between the L2 speaker variables (accentedness, age at the time of testing, age of first contact with Spanish, self-rating, percentage of Spanish use, months of formal education in Spanish) and RTs was also conducted. A Pearson correlation test revealed a positive correlation between accentedness ratings and RTs (**Question RT**: $r(20) = .59$, $p = .003$; **Statement RT**: $r(20) = .46$, $p = .02$). L2 speakers with more native-like scores also had longer reaction times in both question and statement identification (Figure 4.9).

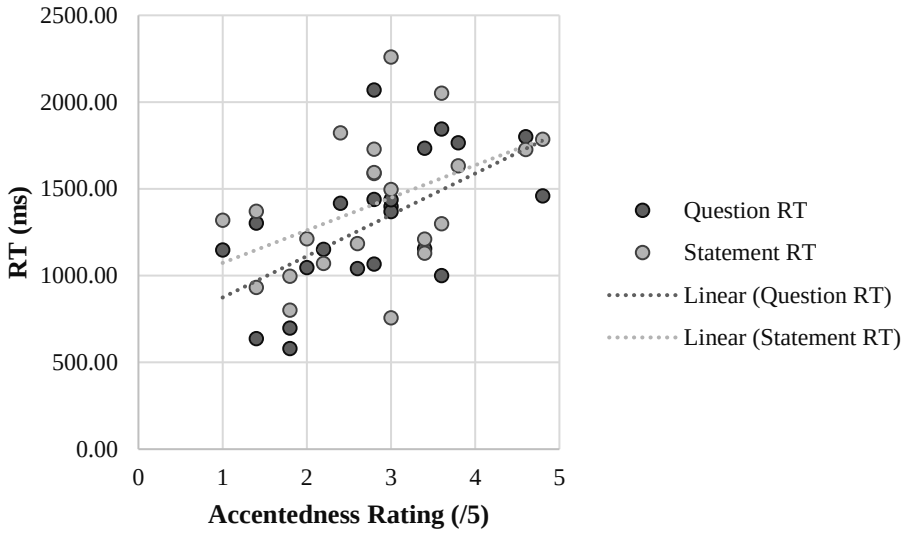

Figure 4.9. L2 speaker correlations between accentedness scores and question RT (dark grey circles) and accentedness scores and statement RT (light grey circles).

4.3 Production Results

The results reported here are based on the productions of Y/N questions and statements that closely resembled what was expected in each context (details in Appendix C). Seventy-one percent of native speaker productions (153 out of a possible 216 tokens) and 79% (417 out of a possible 528 tokens) of L2 speaker productions were used for the calculation (see Appendix I for the total number of utterances per participant). Wh-questions and utterances that were of poor sound quality were excluded. Utterances with pauses were also excluded as they would have created two or more intonation phrases per utterance making the comparison across speakers impossible. For each remaining utterance, the initial boundary tone (IBT) and pre-nuclear peak (PNP) (Figure 4.10) were labelled in Praat (Boersma & Weenink, 2016). The IBT was identified as the first regular pitch movement detected by the software and PNP was the highest point in the post-tonic syllable following the first stressed syllable of the utterance. The values were measured in Hertz (Hz) for each speaker. Mean question and statement values were then calculated for each participant and

those differences were transformed into semitones in order to compare across speakers (Gut, 2013; Nolan, 2003).

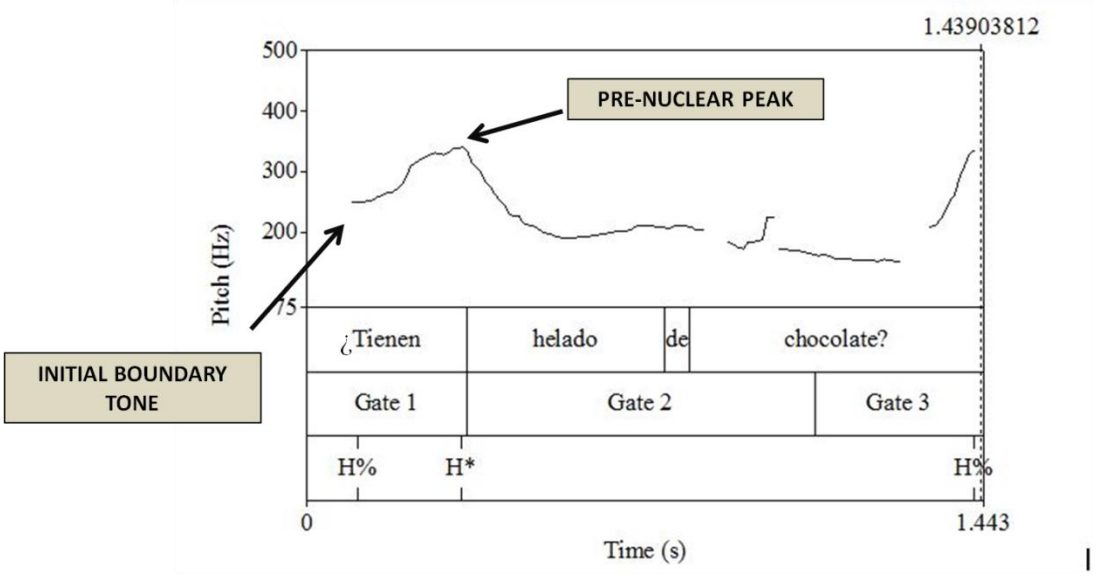

Figure 4.10. ¿Tienen helado de chocolate? 'Do they have chocolate ice-cream?' as realized by a female native (Mexican) Spanish speaker.

4.3.1 Initial Boundary Tone

Figure 4.11 shows mean initial boundary tone differences between Y/N questions and statements for both the native speakers (left) and L2 learners (right) in semitones. The native speaker group produced a greater difference than the L2 speaker group between Y/N questions and statements in the initial boundary tone (NS: 0.36 semitones; L2 speakers: 0.29 semitones). However, such a small difference is not considered reliably discriminable by the human ear (Nooteboom, 1997).

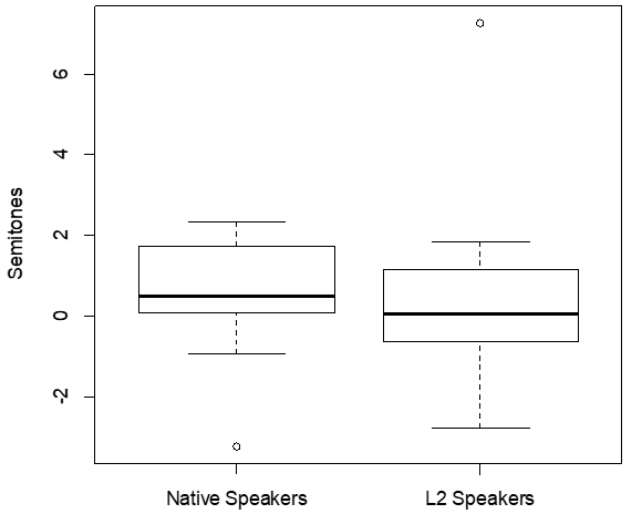

Figure 4.11. Mean difference in semitones between Y/N question and statement initial boundary tone by group.

A Mann-Whitney test indicated that the difference in semitones between Y/N questions and statements in native speakers was not statistically significantly different from the difference in semitones between Y/N questions and statements in L2 speakers, $U = 83$, $p. = .50$ (see Appendix J for means and standard deviations of initial boundary tone values in Hertz per participant as well as visual representation of data for each participant). Figures 4.12 and 4.13 show semitone differences for individual native speakers and L2 speakers respectively.

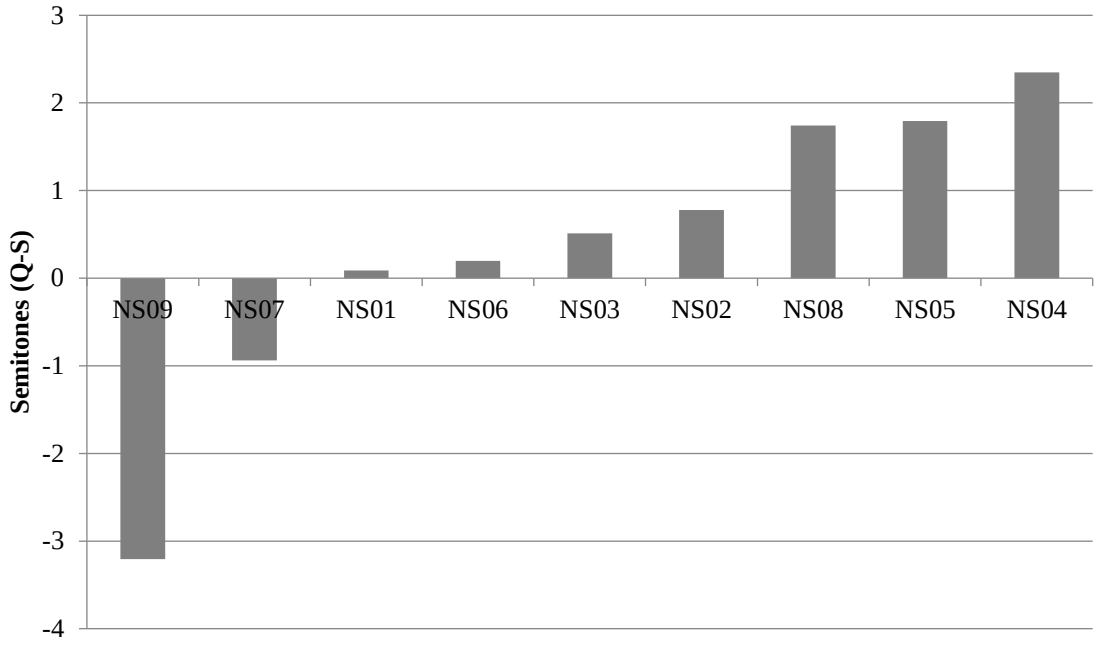

Figure 4.12. Initial boundary tone difference in semitones between Y/N questions and statements for native speakers.

A closer look at the individual native speakers (Figure 4.12) revealed that 5 of 9 participants had differences that were smaller than 1.5 semitones which according to Nooteboom (1997) is the threshold of discriminability for the human ear. Of the remaining 4 native speakers, three (NS08, NS05, and NS04) realized the height difference in the expected (positive) direction while one (NS09) realized the difference in the unexpected (negative) direction. This last finding is not attested in any of the literature on Spanish intonation reported in Chapter 2. Some information about the three speakers with the expected (positive) trend are presented in Table 4.7.

Table 4.7

Native speakers who realized the initial boundary tone in Y/N questions at the expected higher value than comparable statements

Speaker	Age	Gender	Country of Origin	Formal Instruction English
NS08	35	M	Argentina	84 months
NS05	35	M	Venezuela	156 months
NS04	33	F	Venezuela	156 months

These speakers (along with NS09, who is discussed below) were close in age and all came from the southern cone. The remaining native speakers were all in their 20s. There seems to be no other factor that sets the speakers in Table 4.7 apart from the other native speakers. The initial boundary tone difference for speaker NS09 surpassed the 1.5 semitone threshold as well however, the direction was unexpected; the production of the initial boundary tone in statements was much higher than Y/N questions. This particular speaker also stood out in the perception accuracy task as she was 92% accurate in the identification of Y/N questions but only 58% accurate in the identification of statements.

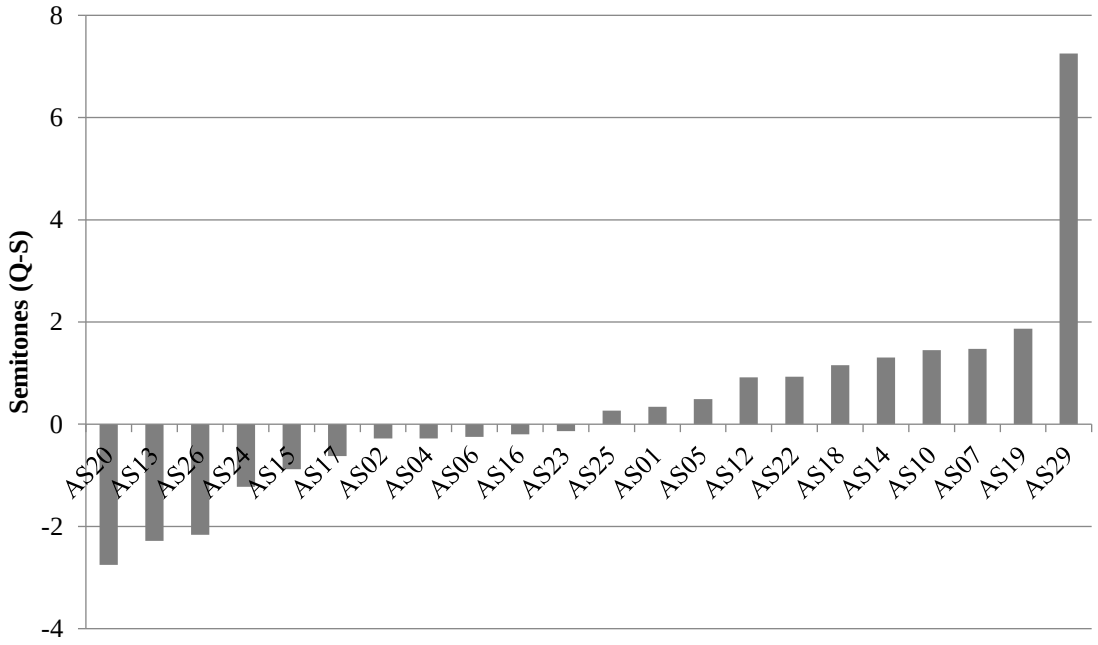

Figure 4.13. Initial boundary tone difference in semitones between Y/N questions and statements for L2 speakers.

The majority of L2 speakers (Figure 4.13) produced a difference between Y/N questions and statements of less than 1.5 semitones. Two speakers produced a difference greater than 1.5 semitones in the expected (positive) direction: AS19 produced a 1.8 semitone difference while speaker AS29 produced a 7.2 semitone difference.[19] Three speakers realized a difference of 1.5 semitones in the (unexpected) negative direction: AS20, AS13, AS26 produced differences of -2.7, -2.2, -2.1 respectively. Once again, these negative values are not attested in any of the literature on Spanish intonation reported in Chapter 2. Tables 4.8 and 4.9 report the demographic data for the L2 speakers with positive and negative values greater than 1.5 semitones respectively.

[19] This particular speaker (AS29) was a 20-year-old male whose mean question value was 187.54 Hz and mean statement value was 123.36 Hz. This seemingly unusual semitone difference may be a result of the low frequency of his voice. The difference of 64.16Hz at such a low frequency has a greater magnitude than it would at a higher frequency (Hart et al., 1990). In fact, the exact same difference at a higher frequency, for example, 400 Hz - 347.82 Hz, translates to 2.9 semitones. While the physical distance is the same in both cases, the semitones reflect the magnitude of the difference for the human ear.

Table 4.8

L2 speakers who realized the initial boundary tone in Y/N questions 1.5 semitones higher than comparable statements

Speaker	Gender	Age	Age of first contact with Spanish	Percentage of Spanish use	Formal Instruction Spanish	NS judge rating
AS19	F	34	20	33%	120months	2
AS29	M	20	18	5 %	18 months	1

Table 4.9

L2 speakers who realized the initial boundary tone in statements 1.5 semitones higher than comparable Y/N questions (unexpected)

Speaker	Gender	Age	Age of first contact with Spanish	Percentage of Spanish use	Formal Instruction Spanish	NS judge rating
AS20	F	23	16	7.5%	12 months	1.8
AS13	M	22	16	5%	48 months	3
AS26	M	23	17	42.5%	72 months	4.6

The information reported in both tables (4.8 and 4.9) does not suggest that these speakers had characteristics that could group them together while distinguishing them from the rest of the L2 speaker group. Furthermore, an analysis between L2 speaker variables (accentedness, age at the

time of testing, age of first contact with Spanish, self-rating, percentage of Spanish use, months of formal education in Spanish) and semitone differences in the production of the initial boundary tone revealed no significant correlations. Overall, in the production of the initial boundary tone, L2 speakers patterned the same way that native speakers did: the majority did not realize a 1.5 semitone difference, some showed the expected trend of Y/N questions realized at a 1.5 semitone or greater value compared to statements, and a few unexpectedly reached the 1.5 semitone threshold in the opposite direction.

4.3.2 Pre-Nuclear Peak

Figure 4.14 shows mean pre-nuclear peak differences between Y/N questions and statements for both the native speakers (left) and L2 learners (right) in semitones; the native speaker group produced a greater difference than L2 speakers. Specifically, the native speaker group produced a difference that was above the discriminability threshold at 1.8 semitones while the L2 speakers' difference fell just below the threshold at 1.3 semitones.

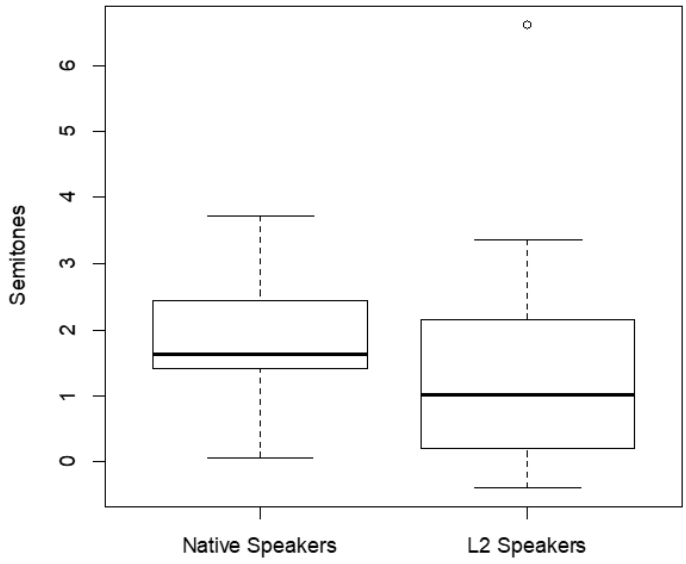

Figure 4.14. Mean difference in semitones between Y/N question and statement pre-nuclear peak by group.

A Mann-Whitney test indicated that the difference in semitones between Y/N questions and statements in native speakers was not statistically significantly different from the difference in semitones between Y/N questions and statements in L2 speakers, *U = 67, p. = .17*. (see Appendix K for means and standard deviations of pre-nuclear peak values in Hertz per participant as well as visual representation of data for each participant).

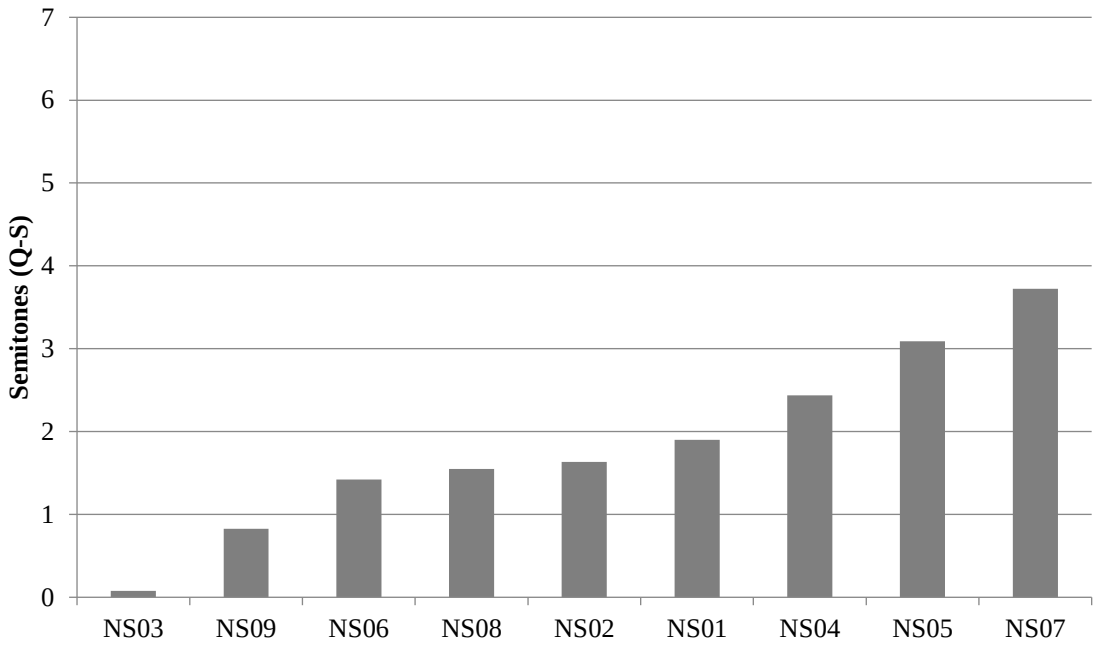

Figure 4.15. Pre-nuclear peak semitone values for native speakers.

The majority of native speakers (Figure 4.15) produced a difference of 1.5 semitones or greater indicating that their pre-nuclear peak in questions was higher than that of their statements. Only three speakers did not reach the 1.5 semitone threshold with values of 0 semitones, 0.8 semitones, and 1.4 semitones for NS03, NS09, and NS06 respectively. These three speakers did not appear to share any demographic characteristics. The details are reported in Table 4.10 below.

Table 4.10

Native speakers who did not realize pre-nuclear peak questions at 1.5 semitones or higher with respect to comparable statements

Speaker	Age	Gender	Country of Origin	Formal Instruction English
NS03	21	M	Argentina	240 months
NS09	32	F	Chile	96 months
NS06	22	M	Mexico	156 months

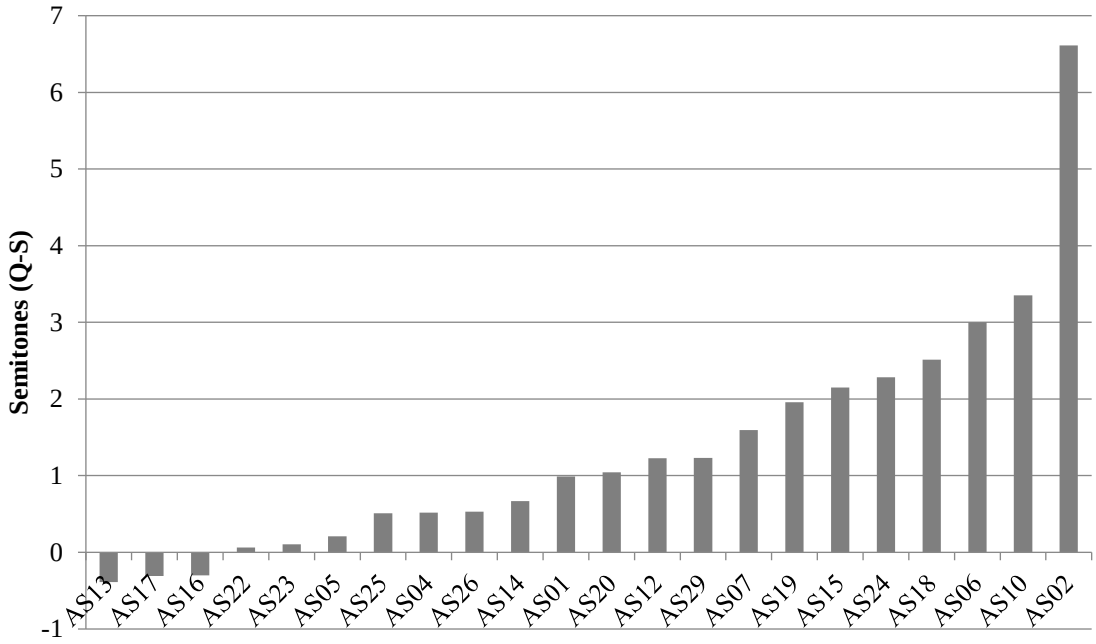

Figure 4.16. Pre-nuclear peak semitone values for L2 speakers.

The majority of L2 speakers (Figure 4.16) did not produce a difference that was greater than 1.5 semitones between their question and statement production. Their values ranged from -0.3 to 1.2 semitones. Eight L2 speakers (AS07, AS19, AS15, AS24, AS18, AS06, AS10, AS02) produced

values greater than 1.5 semitones. Their values ranged from 1.5 to 6.6 semitones.[20] None of these speakers shared any characteristics that distinguished them from the rest of the L2 speaker group (Table 4.11).

Table 4.11

L2 speakers who realized the pre-nuclear peak of their Y/N questions at least 1.5 semitones higher with respect to comparable statements

Speaker	Gender	Age	Age of first contact with Spanish	Percentage of Spanish use	Formal Instruction Spanish	NS judge rating
AS07	F	22	11	42%	60 months	2.8
AS19	F	34	20	34%	120months	2
AS15	F	20	10	17.5%	120months	2.4
AS24	M	36	18	17%	32 months	3.4
AS18	M	25	19	5%	12 months	2.8
AS06	F	33	17	12.5%	60 months	3.4
AS10	M	23	16	15%	12 months	4.8
AS02	M	29	15	37.5%	83 months	3

[20] Speaker AS02 was a 29-year-old male whose mean question value was 203.75 Hz and mean statement value was 139.07 Hz. This seemingly unusual semitone difference may be a result of the low frequency of his voice. See previous note for details.

An analysis between L2 speaker variables (accentedness, age at the time of testing, age of first contact with Spanish, self-rating, percentage of Spanish use, months of formal education in Spanish) and pre-nuclear peak production did not result in any significant correlations.

4.4 Relationship between Perception and Production

This section examines the relationship between the findings in the perception and production tasks presented in sections 4.2 and 4.3 respectively. A correlation test was performed between accuracy in perception and pitch height differences in production. For each group, a separate correlation was performed for the initial boundary tone production and the pre-nuclear peak production.

Figures 4.17 (native speakers) and 4.18 (L2 speakers) plot the relationship between the percentage of identification accuracy of Y/N questions and statements and the semitone difference between questions and statements in the production of the **initial boundary tone**.

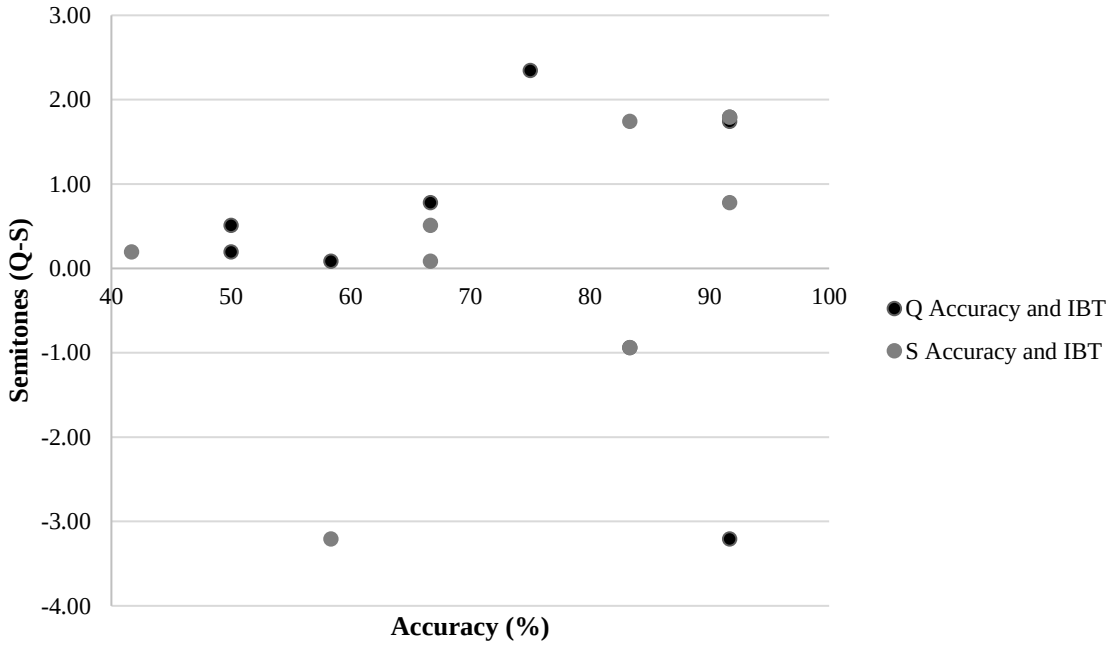

Figure 4.17. Relationship between percentage of identification accuracy (questions: black circles, statements: grey circles) and semitone difference in production of the initial boundary tone for native speakers.

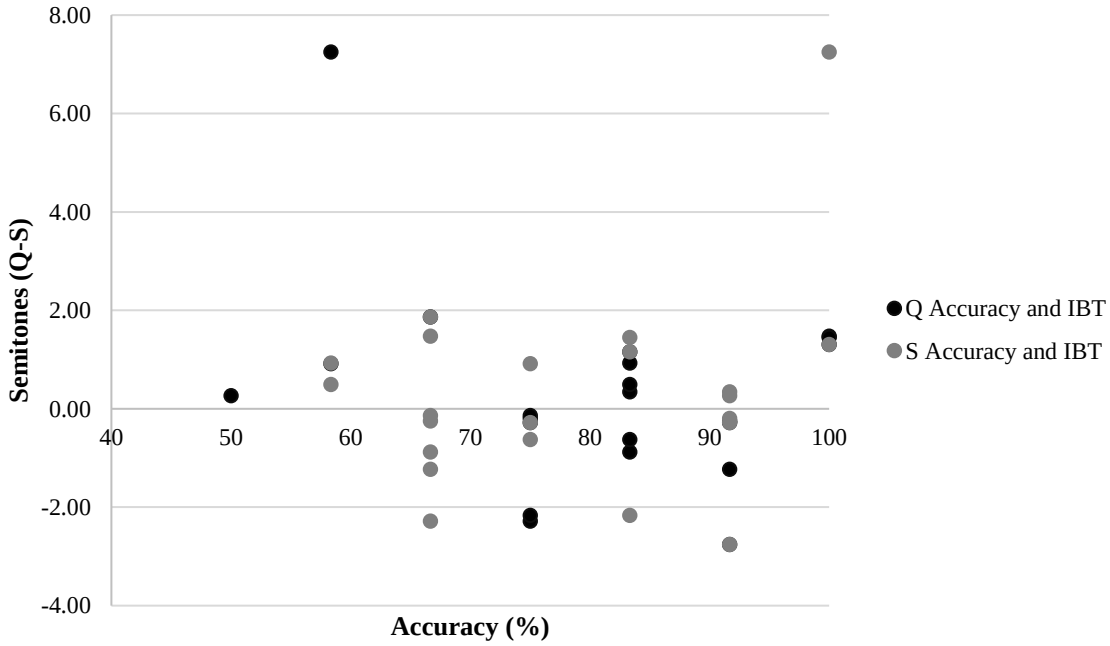

Figure 4.18. Relationship between percentage of identification accuracy (questions: black circles, statements: grey circles) and semitone difference in production of the initial boundary tone for L2 speakers.

With an increase in the percentage of accurate identification, an increase in semitone difference was expected to occur. A higher percentage of identification accuracy would suggest greater familiarity with the question-statement difference in perception and a higher semitone difference between question and statement production would be considered more target-like. The results of a Spearman correlation test revealed that there was no significant correlation between question or statement accuracy and production of initial boundary tone for either speaker group: **NS questions:** r_s (7) = .11, p = .77; **NS statements:** r_s (7) = .14, p = .71; **L2 questions:** r_s (20) = -.02, p = .89; **L2 statements:** r_s (20) = .08, p = .71.

The same analysis was conducted on the relationship between the percentage of identification accuracy of Y/N questions and statements and the semitone difference between questions and statements in the production of the **pre-nuclear peak**. Individual relationships are plotted in Figures 4.19 and 4.20 for native speakers and L2 speakers respectively.

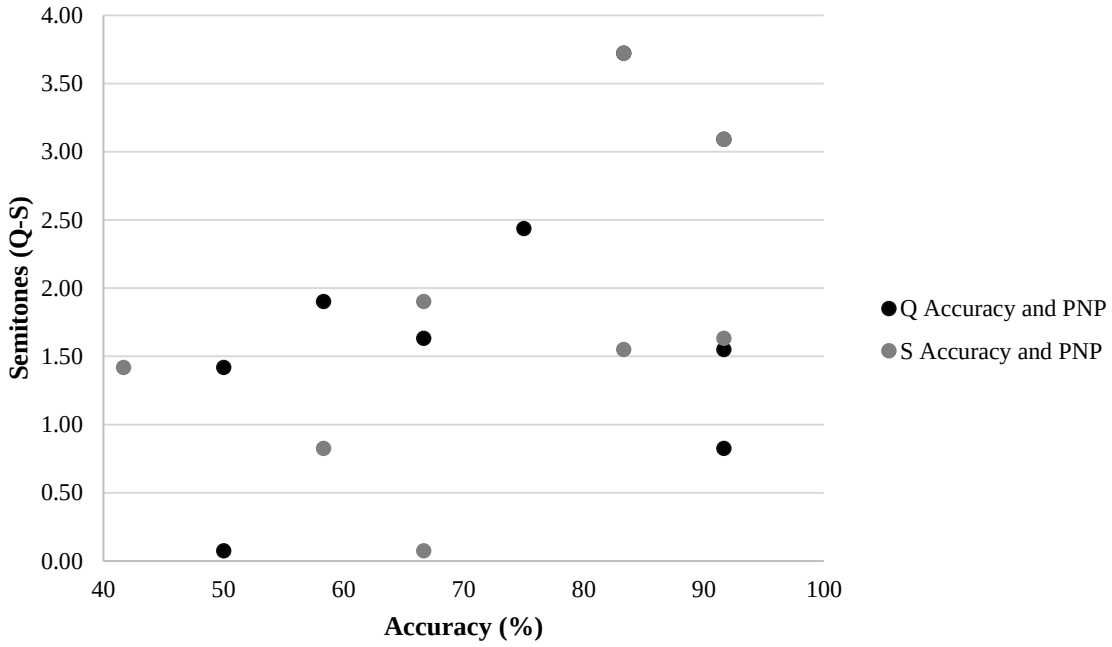

Figure 4.19. Relationship between percentage of identification accuracy (questions: black circles, statements: grey circles) and semitone difference in production of the pre-nuclear peak for native speakers.

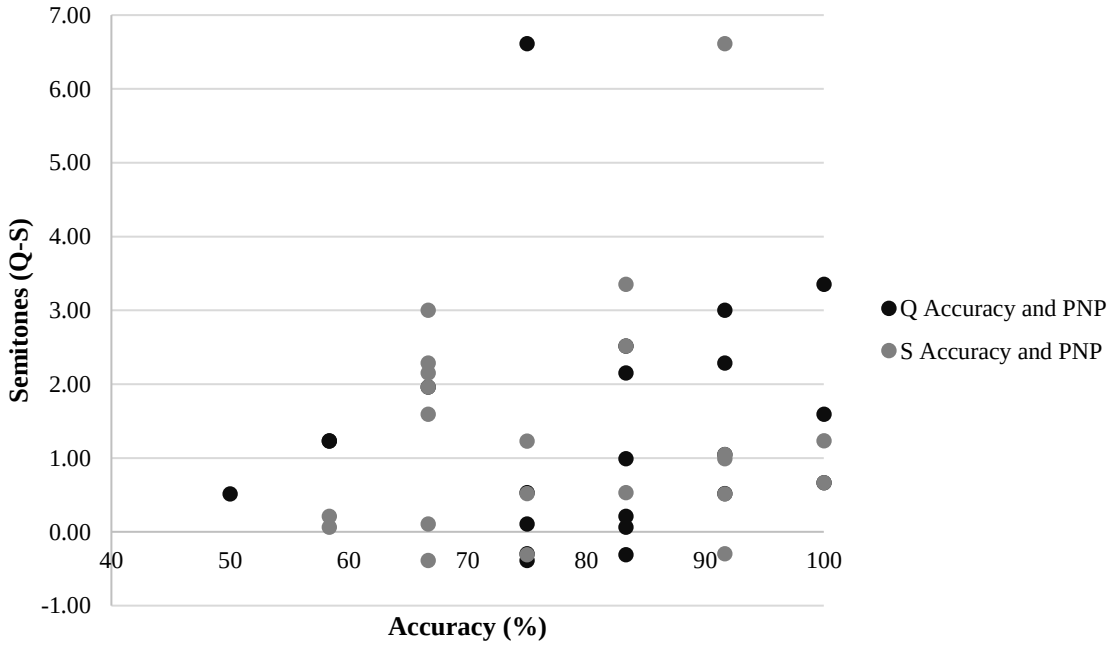

Figure 4.20. Relationship between percentage of identification accuracy (questions: black circles, statements: grey circles) and semitone difference in production of the pre-nuclear peak for L2 speakers.

With an increase in the percentage of accurate identification, an increase in semitone difference was expected to occur. A higher percentage of identification accuracy would suggest greater familiarity with the question-statement difference in perception and a higher semitone difference between question and statement production would be considered more target-like. The results of a Spearman correlation test revealed no significant correlation between question or statement accuracy and production of pre-nuclear peak for either speaker group: **NS questions:** r_s (7) = .34, p = .36; **NS statements:** r_s (7) = .34, p = .36; **L2 questions:** r_s (20) = .26, p = .23; **L2 statements:** r_s (20) = .11, p = .61.

4.5 Summary

This chapter has presented the results following the research questions and predictions made in Chapter 3. In perception, there were no statistically significant differences between native speakers and L2 speakers. Both groups were successful in the identification of Y/N questions and statements and both groups had comparable reaction times. Individual results showed variation in both the native speakers and the L2 speakers. Some speakers in both the native speaker and the L2 group stood out as having a very big difference between the question accuracy rate and the statement accuracy rate but these speakers did not share any characteristics that could separate them from the rest of their groups. Among the L2 speakers, question accuracy correlated positively with accentedness scores and percentage of Spanish use such that those L2 speakers who reported a greater percentage of use of Spanish and received more native-like accentedness scores were also more accurate in question identification.

The individual results of reaction time data revealed that none of the native speakers had statistically significant differences between their own Y/N question RT and statement RT. Only 3 L2 learners had statistically significant differences between their own Y/N question RT and

statement RT but these speakers did not appear to share any characteristics. A positive correlation was found between accentedness ratings and RT such that those L2 speakers who received a more native-like scores were also slower in identifying both Y/N questions and statements.

In production, native and L2 speakers showed no group differences between Y/N questions and statements in the initial boundary tone. Both groups realized a difference between Y/N questions and statements that was below 1.5 semitones. An individual analysis in both the native speaker and the L2 speaker groups revealed that some speakers realized the initial boundary difference at a value that was at least 1.5 semitones or greater, however, these speakers did not share any obvious characteristics. Furthermore, no relationship was found between L2 speaker variables and the production of the initial boundary tone. In the production of the pre-nuclear peak, the native speaker group realized the difference at a value that surpassed the 1.5 semitone threshold while the L2 speaker was just below this value. The group differences, however, were not statistically significant. The individual analysis revealed that not all native speakers or L2 speakers realized the pre-nuclear peak height at a 1.5 semitone difference or greater. However, these speakers did not share characteristics. Furthermore, no relationship was found between L2 speaker variables and the production of the pre-nuclear peak.

The relationship between perception and production was analyzed by looking at perception accuracy and production. No statistically significant correlations were found for either group.

We now turn to the following chapter to discuss these findings by reviewing the research questions and predictions. We will then consider the contributions we intended to make as stated in Chapter 1. Finally, Chapter 5 will also consider the shortcomings of the study and possible future directions.

Chapter 5
Discussion and Conclusion

5.1 Introduction

This chapter begins with a discussion of the results presented in Chapter 4 in light of the research questions and predictions presented in Chapter 3. After a detailed analysis of each research question, we will turn our focus to contributions that this study sought to make as stated in Chapter 1. The chapter concludes with possible future avenues of research.

5.2 Evaluation of Research Questions

5.2.1 Question1

The first question was: Are L2 learners of Spanish able to accurately perceive the higher initial boundary tone and pre-nuclear peak of Spanish Y/N questions with respect to comparable statements? Participants were expected to be highly accurate in the identification of both utterance types and, in fact, both native and L2 speakers were at least 69% accurate or higher with no significant difference between the two groups. Individual accuracy scores revealed that two native speakers had very different accuracy rates between Y/N questions and statements: NS04 was 75% accurate (9 out of 12) in the identification of questions but only 34% (4 out of 12) accurate in the identification of statements while NS09 was 92% (11 out of 12) accurate in identifying Y/N questions but only 58% (7 out of 12) accurate in identifying statements. Furthermore, one native speaker correctly identified Y/N questions at chance (6 out of 12) and was only 42% (5 out of 12) accurate in the identification of statements. These speakers did not share any common characteristics.

Among the L2 speakers, two stood out as having noticeably different accuracy scores for questions when compared to statements. NS25 was only 50% accurate (6 out of 12) in the identification of questions but 92% (11 out of 12) accurate in the identification of statements while AS29 was only 58% (7 out of 12) accurate in identifying Y/N questions but 100% (12 out of 12) accurate in identifying statements. These two speakers shared some common characteristics. They were close in age (early 20s), shared the same age of first contact with Spanish (18 years old), had a low percentage of Spanish use (2.5% and 5% respectively), received 24 and 18 months of formal Spanish instruction respectively, and received low native speaker judge ratings (1.8 and 1 respectively). However, the other L2 speakers in this study shared some of these characteristics as well but did not show the same pattern in perception. Upon further analysis of the L2 speaker variables (accentedness, age at the time of testing, age of first contact with Spanish, self-rating, percentage of Spanish use, months of formal education in Spanish) and accuracy rates, a positive correlation was found between question accuracy and accentedness scores and between question accuracy and percentage of use of Spanish. With respect to the L2 speakers, the speaker variables that correlated with higher accuracy also correlated with each other. In section 3.2 we saw that a higher accentedness score positively correlated with a higher percentage of Spanish use. It is not difficult to imagine that a greater use of Spanish would also include more opportunities for listening and perhaps the greater amount of input has allowed these L2ers to be more accurate in the perception of Y/N questions in Spanish.

Reaction times were expected to be longer overall for L2 speakers compared to the native speaker group; however, no statistically significant difference was found between the two groups. Furthermore, there were no differences between questions and statements within each group. An individual analysis was also conducted on question and statement RTs for each participant. For each of the native speakers, none of the differences between Y/N question and statement RTs were

statistically significant. For L2 speakers, only three had statistically significant differences between Y/N question and statement RTs. Of these 3 speakers, only 1 had a difference in the expected direction (i.e., question RT longer than statement RT: participant AS06). The other two participants (AS18 and AS20) had statistically significant differences in the unexpected direction (i.e., statement RTs longer than question RTs). These speakers did not share any similar characteristics in terms of age, gender, first contact with Spanish, percentage of use of Spanish, native speaker judge rating, or months of formal instruction of Spanish. An analysis between L2 speaker variables and RTs revealed a positive correlation between accentedness scores and RTs. Speakers with higher (more native-like) accented scores were also slower in both question and statement RTs. This finding is in contradiction with the expectation that RTs should indicate difficulty. We saw that accentedness scores positively correlated with percentage of Spanish use such that those L2 speakers who were rated as more native-like also reported a greater use of Spanish in their daily lives. We might expect that L2 speakers who have greater contact with Spanish and are more native-like would be faster in their identification of questions and statements, but the correlation was in the opposite direction. It seems that further investigation into what RTs convey in terms of intonation is required.

Considering the results of these L2 speakers, the SLM (Flege, 1995) would suggest that those speakers who are able to accurately perceive Y/N questions and statements have created a new category for this distinction in their L2 grammar and the PAM-L2 (Best & Tyler, 2007) would suggest that the question-statement height difference in the L2 has not been assimilated to the L1. Furthermore, those L2 speakers who do not accurately perceive Y/N questions and statements in the L2, are having difficulty in the semantic and frequency dimensions as Mennen's (2015) model suggested. The pattern we observe in the native speakers of this study, however, make the application of the speech models somewhat challenging. In fact, the native speakers themselves

do not all pattern the way the literature on Spanish intonation suggested they would. In Flege's terms, the native Spanish speakers do not show the existence of these categories. A possibility we must consider then is that the Y/N question-statement distinction in Spanish can be signaled in more than one way. The findings suggest that not all native Spanish speakers will signal a Y/N question with a high pitch compared to statements despite what the literature on Spanish intonation has reported thus far. Before exploring this option any further, we must also evaluate the production results of the current study.

5.2.2 Question 2

The second question was: Are learners of Spanish able to accurately produce a higher initial boundary tone and pre-nuclear peak of Spanish Y/N questions with respect to comparable statements? L2 speakers were expected to produce a higher initial boundary tone and pre-nuclear peak in *their own* Y/N question production with respect to *their own* statement production. However, these differences in L2 production were not expected to reach native speaker differences. Some between-speaker variation was expected.

5.2.2.1 Initial Boundary Tone

There was no statistically significant difference between the native speaker group and the L2 group in the production of the initial boundary tone. Group averages placed both groups well below the 1.5 semitone threshold discussed in the results (NS: 0.36 semitones; L2 speakers: 0.29 semitones). Individually, the native speakers showed two trends. Five of the 9 speakers had production values that fell below the 1.5 semitone difference (positive and negative values are included here). The remaining 4 native speakers realized initial boundary tone height differences that were greater than 1.5 semitones with three speakers (NS08, NS05, and NS04) in the expected (positive) direction

and one speaker (NS09) in the unexpected (negative) direction. These speakers were all in their 30s and from the southern cone but there seems to be no other factor that sets these speakers aside from the other native speakers.

The majority of L2 speakers did not realize the initial boundary tone with a 1.5 semitone difference. Only 5 speakers were able to produce a semitone difference greater than 1.5 semitones: AS19 and AS29 in the expected (positive) direction and AS20, AS13, and AS26 in the unexpected (negative) direction. These speakers did not seem to have any common characteristics and an analysis of the L2 speaker variables (accentedness, age at the time of testing, age of first contact with Spanish, self-rating, percentage of Spanish use, months of formal education in Spanish) and the production results did not reveal any statistically significant correlations.

5.2.2.2 Pre-Nuclear Peak

In the production of the pre-nuclear peak, the native speaker group produced a difference in semitones (>1.5) between Y/N questions and statements that was reliably discriminable for the human ear while the L2 speaker group was just below the 1.5 semitone threshold. This difference, however, was not statistically significant. Three native speakers (NS03, NS09, and NS06) did not realize the pre-nuclear peak of their Y/N questions higher than 1.5 semitones compared to their statements. These speakers did not appear to share any characteristics.

In the L2 speaker group, 8 of the 22 speakers tested showed a 1.5 or greater semitone difference in the expected direction, however, these 8 speakers did not share any characteristics that distinguished them from the rest of the L2 speaker group. Furthermore, there was no significant correlation between L2 speaker variables (accentedness, age at the time of testing, age of first contact with Spanish, self-rating, percentage of Spanish use, months of formal education in Spanish) and pre-nuclear peak production.

If we consider the L2 production of the initial boundary tone and/or the pre-nuclear peak, Flege's (1995) SLM would suggest that those speakers who produced a difference of 1.5 semitones, where the question value was higher than the statement value, have created a new category in their L2 based on the accurate perception of the L2. However, given that native speakers showed a pattern that was unexpected with respect to the literature on initial boundary tones and pre-nuclear peaks in Y/N questions and statements in Spanish, we may have to reconsider what the target-like production in Spanish really is. Given the findings presented in Chapter 4 and summarized above, it is possible that there exists more than one way in which Spanish signals the Y/N question-statement distinction. If this variety is present among native Spanish speakers, it is likely that these (and other) L2 speakers have received different types of input.

At this point, it is difficult to determine what these different possibilities may look like and if there are any other characteristics (e.g., age, variety of Spanish, level of education) that characterize different groups who make use of different cues. The native speaker data here had three speakers (NS08, NS05, and NS04) who produced a higher initial boundary tone **and** pre-nuclear peak in Y/N questions compared to statements where the height difference was greater than 1.5 semitones. These three speakers are all in their 30s and from the southern cone. However, there is another native speaker from the southern cone also in her 30s (NS09) who does not match this pattern at all. Furthermore, there are a few native speakers who only realized a higher pre-nuclear peak in questions but showed no height difference in the realization of the initial boundary tone (NS02, NS01, NS07).

5.2.3 Question 3

The third question was: What is the relationship between perception and production of the initial boundary tone and pre-nuclear peak in L2 Spanish? We expected the data to show a positive

correlation between accuracy in perception and semitone difference in production. Correlations were conducted on native speaker and L2 speaker data separately. The initial boundary tone production was separated from the pre-nuclear peak production. There was no statistically significant correlation for any of the relationships tested.

A possible explanation for the non-significance of these results comes from Cangemi, Krüger, and Grice (2015).[21] Their study focused on L1 German intonation, however, their ideas about the relationship between perception and production of intonation merit consideration. Cangemi et al. (2015) suggested that there could be 3 possible relationships between perception and production. The first type of relationship is *independent* where some speakers are overall more intelligible, and some listeners are overall more reliable, but they are not the same people. The second type of relationship is a *link* where some individuals are both accurate in their productions and reliable in their perceptual judgements. Finally, the third type of relationship is an *interaction* in which some speakers might be overall more intelligible and some listeners might be overall more reliable but there is no such thing as a "universally intelligible speaker" or a "universally intelligible listener" and thus no individual who is both (making this relationship different from the link option). Cangemi et al. (2015) tested the acoustic correlates of focus, which, in German, is signaled by peak alignment and peak height but can also be signaled by the duration of the target word, the duration of the first word, and the number of pre-nuclear pitch accents. These authors found that the same speaker could produce contrasts which were well decoded by a certain listener, but poorly decoded by a different listener. At the same time, the same listener could reliably decode contrasts as produced by a given speaker, while being less reliable with productions from a different speaker. The findings in Cangemi et al. (2015) support the hypothesis of an interaction (not a link) between

[21] I was unaware of this study when I began the current study.

perception and production of intonation in this case. If we apply their findings to the study in this dissertation, we may find that a participant who has reliable perceptual abilities may not have equally reliable accuracy in production or vice versa. It would be necessary to test the intelligibility of the production of our speakers with other listeners in order to determine if Cangemi et al.'s proposal can be applied here.

Related to this idea is the possibility that Spanish signals the question-statement distinction with more than one cue. The assumption in this study, based on the literature presented in Chapter 2, was that Spanish speakers produced a higher initial boundary tone *and* pre-nuclear peak to signal Y/N questions compared to statements and that this was the difference that listeners would attend to when making this distinction in perception. However, based on the performance of the native speakers in this study, we may suggest that Spanish has more than one way to signal the Y/N question-statement distinction. Other potential cues may include peak alignment and/or speech rate. A few possibilities are explored below. Prieto (2004), for example, found that her participants relied on a combination of alignment and scaling (height) to distinguish sentence type in Spanish. Specifically, Y/N questions and statements showed phonetic differences in their alignment. While both utterance types presented peak alignment in the post-tonic syllable, interrogatives had a fixed place while declaratives varied according to the stress pattern of the word and the upcoming L tone. Alternatively, Gussenhoven (2004) has suggested that a peak delay (alignment) may replace greater peak height altogether. He explains that a higher pitch peak will take longer to reach than a lower one if rate of change is the same. Therefore, higher peaks will tend to be later than lower peaks. Peak delay can therefore be used as a substitute for greater peak height. Finally, a study by van Heuven and van Zanten (2005) found that Y/N questions had consistently faster speech rate than comparable statements in three typologically different languages (Malay, English, and Dutch). These authors suggested that speech rate could be a secondary cue to the question-

statement distinction comparable to the pitch height difference. In sum, studies have found that pitch height may be accompanied by differences in alignment, or it may be entirely replaced by another phonetic feature like alignment or speech rate. The literature discussed in Chapter 2 assumes that the Y/N question-statement distinction in Spanish is based exclusively on the pitch height difference but if other phonetic features can signal the question-statement distinction, this would explain why there is no significant correlation in the current study.

Having considered the results of the current study with respect to the research questions and hypotheses, we now turn to the general contributions of the current study to the field of L2 intonation and the study of Spanish intonation.

5.3 Contributions

This study has investigated the perception and production of the initial boundary tone and pre-nuclear peak of Spanish Y/N questions and statements by L2 speakers. One of the contributions of this study is the explicit focus on the relationship between L2 perception and L2 production of intonation. As discussed in section 2.9, only a handful of studies have investigated this relationship in L2 intonation, but the findings have varied. The results of the current study, as seen in section 4.4, found no significant correlations between L2 perception and production of intonation. Interestingly, no correlation between native Spanish perception and production of intonation was found either. These findings suggest two possibilities. The first one is that the relationship between (L2) perception and production is not as straightforward as past research may have assumed it to be. In section 5.2.3 we explored Cangemi et al. (2015)'s proposal with respect to the relationship between perception and production. They proposed that the relationship is an *interaction* in which some speakers might be overall more intelligible and some listeners might be overall more reliable but there is no such thing as a "universally intelligible speaker" or a "universally intelligible

listener" and thus no individual who is both. This scenario would explain the lack of significant correlations. The other possibility is that there was no significant correlation in the current study because the Y/N question-statement distinction in Spanish may be signaled in more than one way despite what the literature has reported thus far. The individual variation in the native Spanish speaker data in the current study suggests the existence of more than one way to signal the Y/N question-statement distinction. It follows that L2 speakers are likely to have received input from native speakers that vary as well. Furthermore, if speakers signal the Y/N question-statement distinction in more than one way in production, we must also consider the possibility that listeners are not all relying on the same cue (i.e., pitch height difference) in perception. This possibility would explain the lack of significant correlations here.

The L2 speech models discussed in Chapter 2 would have some difficulty predicting this result. The SLM and the PAM-L2 consider the perception-production relationship to be occurring within the same individual. There is no discussion about the possibility of accurate perceivers not being accurate producers and vice versa. Mennen's (2015) LILt, while not directly concerned with the perception-production relationship, supports the proposal presented in Flege's (1995) SLM. Mennen's model, however, is innovative in that it considers what is specific to L2 intonation and where L2ers may have difficulty in acquiring the intonation of an L2 by giving researchers a framework for cross-linguistic comparison. The L2 models, especially Mennen's, will benefit from the findings of the current study because the current findings have brought to our attention an aspect that needs to be considered in L2 speech models: multiple (intonation) cues that convey the same meaning. In Mennen's model, both the semantic and the frequency dimension would require additional detail as a result of the current findings. The semantic dimension is quite self-explanatory as it concerns the meaning that intonation conveys. However, the frequency dimension

would become important in determining exactly what the distribution of the multiple cues would be.

In light of the possible multiple ways in which the Y/N question-statement distinction may be signaled by native speakers, this dissertation also contributes to what Prieto and Roseano (2019) identify as one of the current issues in Spanish intonation research: a lack of a *full* description of Spanish intonation. While the last 20 years have seen several overviews of some of the most common features of Spanish intonation (e.g., Hualde & Prieto, 2015; Prieto & Roseano, 2010; Sosa, 1999), a fine-grained analysis is still missing. One of the missing pieces of information that has emerged from this study is that the height difference that has been described as signaling the Y/N question-statement distinction is not quantified in any way. In this study we chose to transform individual pitch height differences into semitones in order to compare across speakers (as recommended by Gut, 2013 and Nolan, 2003). Furthermore, we choose a minimum threshold of 1.5 semitones because, according to Nooteboom (1997), it is the threshold of discriminability for the human ear. It is unclear however if this cut-off is linguistically meaningful or not. We must consider the possibility that a different (higher) value, in the case of those who use this height difference to signal the Y/N question-statement distinction, is the minimum threshold. Furthermore, while we can assume the existence of a minimum value that distinguishes Y/N questions from statements, we did not discuss the possibility of a maximum value. In theory, the difference could be as great as the speaker's anatomy allows. This possibility may explain the L2 findings (speaker AS29 in Figure 4.15 and Speaker AS02 in Figure 4.18) that seem to stand out from the rest of the speakers. The production of these speakers may be a result of their anatomy[22]

[22] As described in notes 19 and 20 in Chapter 4, the pitch values at such low frequencies (typical of many male speakers) will result in a higher semitone value.

coupled with the potential to make the difference as big as possible. While it is clear that this height difference requires further investigation, the analysis here contributes to a more complete description of Spanish intonation.

A full intonational description of Spanish also benefits from the methodological choices made here. Some studies have found noticeable differences between intonation patterns of scripted experimental settings and those of spontaneous speech (e.g., Face, 2003; Hedberg & Sosa, 2002). As a result of these findings, the methodological decision is not a trivial one, particularly if we aim to generalize our findings to everyday speech. While earlier studies in intonation relied heavily on descriptions from non-contextualized sentences with doubtful semantics (Kohler, 2006), there has been a clear shift towards application of a variety of empirical methods that integrate intonation into a context (Prieto and Roseano, 2019). The perception and production tasks in the current study were designed with the goal of avoiding highly structured, non-contextualized tasks that were very common in earlier intonation studies. The perception task used here was based on Grosjean's (1980) gating paradigm where participants listened to increments (gates) of an utterance and were asked to determine whether they heard a question or a statement at each increment (gate). This task allowed us to focus on the part of the utterance (the initial part) that was of interest here without additional manipulation of the sound files. Furthermore, this task began with a brief narrative of the speaker used allowing the listener to establish a range for the speaker. The production task in this study was a narrative designed to elicit a question or a statement by removing one utterance and asking the participant to provide it. In this task, the participant produced an utterance that was in context allowing us to move away from highly structured sentence reading tasks.

This section has presented the contributions of this study to the field of intonation research. The relationship between L2 perception and production has not been thoroughly investigated. The

findings here, while not significant, open the field up to new possibilities both in terms of the relationship overall as well as what this means for Spanish intonation particularly in the Y/N question-statement distinction. Furthermore, the results of this study have suggested that there are possibly more ways to signal the Y/N question-statement distinction in Spanish.

This study has also shed some light on what is still in need of investigation in Spanish intonation research. The height difference that is reported by many sources is not quantified in any way. In the current study, we have attempted to quantify this height difference. The results suggest that more research is needed.

Finally, the methodological choices in this study have allowed us to put both the perception and production task in context, a need that was identified long ago by Kohler (2006) and is now becoming common practice (Prieto and Roseano, 2019).

We now turn to the following section to consider some future avenues of research.

5.4 Future Work

After considering the contributions of the current study, we wish to consider what can be done in future research.

First, a more find-grained analysis of the possible multiple ways of signaling the Y/N question-statement distinction in Spanish should be undertaken, as the results of the current study suggest that there is a possibility that Spanish speakers signal this difference in more than one way. A study dedicated exclusively to the understanding of how many ways this difference can be signaled would give us a more complete picture of what the intonational inventory of Spanish looks like. In turn, this information would allow us to understand the variety of input that L2 speakers are exposed to.

With respect to the L2 speakers, the current study operated under the assumption that this L2 group was relatively homogenous. While it is true that these participants were able to carry out the tasks with relative ease suggesting that they all had similar abilities, it is possible that a more fine-grained analysis would reveal the existence of sub-groups. It is difficult to draw conclusions about this group if there are multiple proficiency groups, as we would expect them to be at different stages of acquisition and we would expect this to be reflected in the results. In the current study, accentedness ratings ranged from 1 to 4.8. While this range suggests some kind of difference between the L2 speakers, it is unclear what influenced the judges' scores. Future studies should consider the addition of a grammatical or a lexical assessment tool such as a cloze test.

Finally, intonation research would benefit from exploring different methodologies to obtain generalizable results. The perception task in the current study was intended to 1) use unmodified stimuli and b) isolate and focus on a specific part of the contour. While participants were able to carry out the task, it is important to remember that this task is somewhat unnatural when compared to everyday conversation context. Tasks that involve some (minimal) manipulation but that mimic the more natural course of conversation are worth considering. One option for a perception of intonation task may include a context where the listener is provided with 2 possible utterances at a specific point (e.g., one question and one statement). The participant would be asked to choose which of the two, based on a given context, is the best fit. This type of task would require ensuring that the two options only differ in one specific feature (e.g., peak height OR peak alignment) in order to test that particular feature and not have other confounding variables. It would require some manipulation of the other parameters within the intonation contour. In the production task, participants were also overall able to comfortably carry out the task, but a few native speakers observed that it seemed like a strange or unusual task. This could be a result of the fact that this task was designed primarily with L2 speakers in mind and mimics exercises that are typical of an

L2 classroom. An alternative could be an object moving task with pictorial instructions (Ito & Speer, 2006). Participant pairs would have a map of where certain objects are located and they would have the goal of making the other person recreate the layout by asking if they have a particular object and then by telling them where to put that object.

5.5 Conclusion

This dissertation has analyzed the perception and production of Y/N questions and statements by L2 Spanish speakers. Specifically, the initial boundary tone and pre-nuclear peaks of Y/N questions and statements were discussed. Findings revealed that this advanced group of L2 speakers was successful in perception. In production, some L2 speakers reached the expected differences in the realization of the initial boundary tone and pre-nuclear peak. The correlation between perception and production did not reveal any statistically significant results. An interesting finding, however, emerged among the native speakers. Not all the native Spanish speakers performed as expected (as presented in section 2.3.1.2) suggesting the possibility of multiple ways to signal the Y/N Q-S distinction in Spanish. This finding could explain the lack of significant correlations and open avenues for new research.

References

Baayen, R. H., & Milin, P. (2010). Analyzing Reaction Times. *International Journal of Psychological Research*, *3*(2). http://www.redalyc.org/html/2990/299023506003/

Baltazani, M., Kainada, E., Lengeris, A., & Nikolaidis, K. (2015). The prenuclear field matters: Questions and statements in Standard Modern Greek. *Proceedings of the 18th International Congress of Phonetic Sciences*, 10–14.

Bartels, C. (1999). *The intonation of English statements and questions: A compositional interpretation*. Garland Pub.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Beckman, M. E., Hirschberg, J., & Shattuck-Hufnagel, S. (2005). The original ToBI system and the evolution of the ToBI framework. In S.-A. Jun (Ed.), *Prosodic Typology* (pp. 8–54). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199249633.001.0001

Beckman, M., Hirschberg, J., & Shattuck-Hufnagel, S. (2005). The original ToBI system and the evolution of the ToBI framework. In S.-A. Jun (Ed.), *Prosodic Typology* (pp. 8–54). Oxford University Press.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn & M. J. Munro (Eds.), *Second-language Speech Learning: The Role of Language Experience in Speech Perception and Production. A Festschrift in Honour of James E. Flege* (pp. 13–34). John Benjamins Publishing Company. https://doi.org/10.1075/lllt.17.07bes

Boersma, P., & Weenink, D. (2016). *Praat: Doing Phonetics by Computer*. http://www.praat.org/

Bohn, O.-S., & Flege, J. E. (1990). Perception and production of a new vowel category by adult second language learners. In J. Leather & A. James (Eds.), *New Sounds 90: Proceedings of the 1990 Amsterdam symposium on the acquisition of second language speech* (pp. 37–56). University of Amsterdam.

Bolinger, D. (1978). Intonation Across Languages. In *Universals of Human Language* (Joseph H. Greenberg Ed., Vol. 2). Stanford University Press.

Bongaerts, T., Mennen, S., & Slik, F. van der. (2000). Authenticity of pronunciation in naturalistic second language acquisition: The case of very advanced late learners of Dutch as a second language. *Studia Linguistica*, *54*(2), 298–308. https://doi.org/10.1111/1467-9582.00069

Canellada, M. Josefa., & Madsen, J. Kuhlmann. (1987). *Pronunciación del español: Lengua hablada y literaria*. Castalia.

Cangemi, F., Krüger, M., & Grice, M. (2015). Listener-Specific Perception of Speaker-Specific productions in Intonation. In S. Fuchs, D. Pape, C. Petrone, & P. Perrier (Eds.), *Individual Differences in Speech Production and Perception* (pp. 123–145). Peter Lang.

Chen, A. (2003). Reaction time as an indicator of discrete intonational contrasts in English. *EUROSPEECH*, 97–100.

Ching, M. K. L. (1982). The Question Intonation in Assertions. *American Speech*, *57*(2), 95–107.

Cournane, A. (2014). In Search of L1 Evidence for Diachronic Reanalysis: Mapping Modal Verbs. *Language Acquisition*, *21*(1), 103–117.

Face, T. L. (2003). Intonation in Spanish declaratives: Differences between lab speech and spontaneous speech. *Catalan Journal of Linguistics*, *2*, 115–131.

Face, T. L. (2005). F0 Peak Height and the Perception of Sentence Type in Castilian Spanish. *Revista Internacional de Lingüística Iberoamericana*, *3*(2 (6)), 49–65.

Face, T. L. (2007). The role of intonational cues in the perception of declaratives and absolute interrogatives in Castilian Spanish. *Estudios de Fonética Experimental*, *16*, 186–225.

Falé, I., & Faria, I. H. (2006). Time course of Intonation Processing in European Portuguese: A Gating paradigm approach. *IV Jornadas En Tecnologia Del Habla*, 265–270.

Flege, J. E. (1995). Second language speech learning: Theory, findings and problems. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-language Research* (pp. 233–277). York Press.

Flege, J. E., MacKay, I. R., & Meador, D. (1999). Native Italian speakers' perception and production of English vowels. *The Journal of the Acoustical Society of America*, *106*(5), 2973–2987.

Flege, J. E., Munro, M. J., & Mackay, I. R. A. (1995). Effects of age of second-language learning on the production of English consonants. *Speech Communication*, *16*(1), 1–26.

Gambi, C., & Pickering, M. J. (2018). Models linking production and comprehension. In E. M. Fernández & H. Smith Cairns (Eds.), *The Handbook of Psycholinguistics* (First, pp. 157–181). John Wiley & Sons, Inc.

Geluykens, R. (1987). Intonation and speech act type: An experimental approach to rising intonation in queclaratives. *Journal of Pragmatics*, *11*(4), 483–494.

Geluykens, R. (1988). On the Myth of Rising Intonation in Polar Questions. *Journal of Pragmatics*, *12*, 467–485.

Gili Fivela, B. (2012). *Testing the perception of L2 intonation* (A. Stella & M. G. Busà, Eds.; pp. 17–30). CLEUP.

Gordon, M. K. (1999). The intonational structure of Chickasaw. *Proceedings of the 14th International Congress of Phonetic Sciences*, *3*, 1993–1996.

Grabe, E., Rosner, B. S., García-Albea, J. E., & Zhou, X. (2003). Perception of English intonation by English, Spanish, and Chinese listeners. *Language and Speech*, *46*(4), 375–401.

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception and Psychophysics*, *28*(4), 267–283.

Gunlogson, C. (2003). *True to form: Rising and falling declaratives as questions in English*. Routledge.

Gunlogson, C. (2002). Declarative questions. In B. Jackson (Ed.), *Proceedings of Semantic and Linguistic Theory XII* (pp. 124–143). Cornell University.

Gussenhoven, C. (2004). *The Phonology of Tone and Intonation*. Cambridge University Press.

Gut, U. (2013). Analysing phonetic and phonological variation on the suprasegmental level. In M. Krug & J. Schluter (Eds.), *Research Methods in Language Variation and Change* (pp. 244–259). Cambridge University Press. https://doi.org/10.1017/CBO9780511792519.017

Guy, G., Horvath, B., Vonwiller, J., Daisley, E., & Rogers, I. (1986). An intonational change in progress in Australian English. *Language in Society, 15*(01), 23–51.

Hart, J. T., Collier, R., & Cohen, A. (1990). *A perceptual study of intonation: An experimental-phonetic approach to speech melody*. Cambridge University Press.

Hedberg, N., & Sosa, J. M. (2002). The Prosody of Questions in Natural Discourse. *Proceedings of Speech Prosody 2002*, 375–378.

Hedberg, N., Sosa, J. M., & Fadden, L. (2004). Meanings and Configurations of Questions in English. *Proceedings of Speech Prosody 2004*. Speech Prosody, Nara, Japan.

Henriksen, N. C., Geeslin, K. L., & Willis, E. W. (2010). The development of L2 Spanish intonation during a study abroad immersion program in León, Spain: Global contours and final boundary movements. *Studies in Hispanic and Lusophone Linguistics*, *3*(1), 113–162.

Hualde, J. I., & Prieto, P. (2015). Intonational variation in Spanish. In S. Frota & P. Prieto (Eds.), *Intonation in Romance* (pp. 350–391). Oxford University Press.

Ito, K., & Speer, S. R. (2006). Using Interactive Tasks to Elicit Natural Dialogue. In S. Sudhoff, D. Lenertová, R. Meyer, S. Pappert, P. Augurzky, I. Mleinek, N. Richter, & J. Schlieβer (Eds.), *Methods in Empirical Prosody Research* (pp. 229–257). Walter de Gruyter.

Jiang, N. (2012). *Conducting Reaction Time Research in Second Language Studies*. Routledge.

Kohler, K. J. (2006). Paradigms in Experimental Prosodic Analysis: From Measurement to Function. In S. Sudhoff, D. Lenertová, R. Meyer, S. Pappert, P. Augurzky, I. Mleinek, N. Richter, & J. Schlieβer (Eds.), *Methods in Empirical Prosody Research* (pp. 123–152). Walter de Gruyter.

Kügler, F. (2017). Tone and intonation in Akan. In L. J. Downing & A. Rialland (Eds.), *Intonation in African Tone Languages* (pp. 89–130). De Gruyter.

Ladd, D. R. (2008). *Intonational Phonology*. Cambridge University Press.

Ladd, D. R., & Morton, R. (1997). The perception of intonational emphasis: Continuous or categorical? *Journal of Phonetics*, *25*(3), 313–342.

Levy, E. S. (2009). On the assimilation-discrimination relationship in American English adults' French vowel learninga). *J. Acoust. Soc. Am.*, *126*(5), 2670–2682.

Liu, C., & Rodriguez, A. (2012). Categorical perception of intonation contrasts: Effects of listeners' language background. *Journal of the Acoustical Society of America*, *131*(6), EL427–EL433.

Llisterri, J. (1995). Relationships between speech production and speech perception in a second language. *Proceedings of the 13th International Congress of Phonetic Sciences*, *4*, 92–99.

Lo, S., & Andrews, S. (2015). To transform or not to transform: Using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology*, *6*, 1–16.

Marasco, O. (2014, May). *L2 Spanish Speakers' Perception and Production of Secondary Cues in Y/N Question and Statement Intonation* [Poster]. Canadian Linguist Association (CLA), St. Catherines, ON, Canada.

Mennen, I. (2004). Bi-directional interference in the intonation of Dutch speakers of Greek. *Journal of Phonetics*, *32*(4), 543–563. https://doi.org/10.1016/j.wocn.2004.02.002

Mennen, I. (2015). Beyond Segments: Towards a L2 Intonation Learning Theory. In *Prosody and Language in Contact: L2 Acquisition, Attrition and Languages in Multilingual Situations* (pp. 171–188). Springer.

Navarro Tomás, T. (1944). *Manual de entonación española*. Hispanic Institute.

Nolan, F. (2003). Intonational equivalence: An experimental evaluation of pitch scales. *Proceedings of the 15th International Congress of Phonetic Sciences*, 771–774.

Nooteboom, S. (1997). The prosody of speech: Melody and rhythm. In W. J. Hardcastle & J. Laver (Eds.), *The Handbook of Phonetic Sciences* (pp. 640–673). Blackwell.

Ortega-Llebaria, M., & Colantoni, L. (2014). L2 English Intonation: Relations between Form-Meaning Associations, Access to Meaning, and L1 Transfer. *Studies in Second Language Acquisition*, *36*(02), 331–353. https://doi.org/10.1017/S0272263114000011

Patience, M., Marasco, O., Colantoni, L., Klassen, G., Radu, M., & Tararova, O. (2018). Initial Pitch Cues in English Sentence Types. *9th International Conference on Speech Prosody 2018*, 463–467. https://doi.org/10.21437/SpeechProsody.2018-94

Petrone, C., & Niebuhr, O. (2014). On the intonation of German intonation questions: The role of the prenuclear region. *Language and Speech*, *57*(1), 108–146.

Pierrehumbert, J. (1980). *The Phonology and Phonetics of English Intonation*. MIT.

Prieto, P. (2004). The search for phonological targets in the tonal space: H1 scaling and alignment in five sentence- types in Peninsular Spanish. In T. L. Face (Ed.), *Laboratory Approaches to Spanish Phonology* (pp. 29–59). Mouton de Gruyter.

Prieto, P., & Roseano, P. (2010). *Transcription of intonation of the Spanish language* (P. Prieto & P. Roseano, Eds.). LINCOM.

Prieto, P., & Roseano, P. (2019). Current issues and challenges in Spanish intonational research. In S. Colina & F. Martínez-Gil (Eds.), *The Routledge Handbook of Spanish Phonology* (pp. 222–236). Routledge.

Puga, K., Fuchs, R., Hudson, T., Setter, J., & Mok, P. (2018). The Perception-Production Link in Intonation: Evidence from German Learners of English. *Speech Prosody 2018*, 1–5.

R Core Team. (2017). *R: A language and environment for statistical computing.* https://www.R-project.org/

Radu, M. (2020). *The L2 Production of Spanish Taps and Trills by L1 Romanian and L1 French Speakers* [Doctoral Dissertation]. University of Toronto.

Radu, M., Colantoni, L., Klassen, G., Patience, M., Pérez-Leroux, A. T., & Tararova, O. (2018). The perception and interpretation of sentence types by L1 Spanish–L2 English speakers. *Linguistic Approaches to Bilingualism*. https://doi.org/10.1075/lab.17004.rad

Rialland, A. (2009). The African lax question prosody: Its realisation and geographical distribution. *Lingua*, *119*(6), 928–949. https://doi.org/10.1016/j.lingua.2007.09.014

Rochet, B. L. (1995). Perception and production of second-language speech sounds by adults. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-language Research* (pp. 379–410). York Press.

Rose, M. (2010). Differences in Discriminating L2 Consonants: A Comparison of Spanish Taps and Trills. In M. T. Prior, Y. Watanabe, & S.-K. Lee (Eds.), *Selected Proceedings of the 2008 Second Language Research Forum* (pp. 181–196). Cascadilla Proceedings Project.

Saindon, M. R., Trehub, S. E., Schellenberg, E. G., & van Lieshout, P. H. H. M. (2017). When is a Question a Question for Children and Adults? *Language Learning and Development*, *13*(3), 274–285. https://doi.org/10.1080/15475441.2016.1252681

Santiago Vargas, F., & Delais-Roussarie, E. (2012). Acquiring Phrasing and Intonation in French as Second Language: The Case of Yes-No Questions Produced by Mexican Spanish Learners. In Q. Ma, H. Ding, & D. Hirst (Eds.), *Proceedings of Speech Prosody 2012* (pp. 338–341).

Schneider, K., Dogil, G., & Möbius, B. (2011). Reaction Time and Decision Difficulty in the Perception of Intonation. *INTERSPEECH*, 2221–2224.

Sheldon, A., & Strange, W. (1982). The acquisition of/r/and/l/by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics*, *3*(03), 243–261.

Shokeir, V. (2008). Evidence for the stable use of uptalk in South Ontario English. *University of Pennsylvania Working Papers in Linguistics*, *14*(2), 16–21.

Sosa, J. Manuel. (1999). *La entonación del español: Su estructura fónica, variabilidad y dialectología*. Cátedra.

SuperLab (4.5). (2015). [Computer software]. Cedrus.

The CEFR Levels. (2020, January). Common European Framework of Reference for Languages (CEFR). https://www.coe.int/en/web/common-european-framework-reference-languages/level-descriptions

Trimble, J. C. (2013). Perceiving Intonational Cues in a Foreign Language: Perception of Sentence Type in Two Dialects of Spanish. *Selected Proceedings of the 15th Hispanic Linguistic Symposium*, 78–92.

Trofimovich, P., & Baker, W. (2006). Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition*, *28*(01), 1–30.

van Heuven, V. J., & van Zanten, E. (2005). Speech rate as a secondary prosodic characteristic of polarity questions in three languages. *Speech Communication*, *47*(1–2), 87–99.

Vion, M., & Colas, A. (2006). Pitch Cues for the Recognition of Yes-No Questions in French. *Journal of Psycholinguistic Research*, *35*(5), 427–445.

Warren, P. (2005). Patterns of late rising in New Zealand English: Intonational variation or intonational change? *Language Variation and Change*, *17*(02), 209–230.

Warren, Paul. (2016). *Uptalk: The phenomenon of rising intonation* (http://go.utlib.ca/cat/10352837). Cambridge University Press.

Whelan, R. (2008). Effective Analysis of Reaction Time Data. *The Psychological Record*, *58*(3), 475–482.

Zárate-Sández, G. (2015). *Perception and Production of Intonation among English-Spanish Bilingual Speakers at Different Proficiency Levels* [Doctoral Dissertation, Georgetown University]. https://repository.library.georgetown.edu/handle/10822/761510

Appendix A

Participant Characteristics

Table A1

Native Speaker Characteristics

Participant	Age at time of testing	Gender	Native Speaker Judge Rating (/5)	Variety of Spanish	Number of months of formal English instruction
NS01	21	F	5	Argentina	156
NS02	21	F	5	Spain	72
NS03	21	M	5	Argentina	240
NS04	33	F	4.8	Venezuela	156
NS05	35	M	4.5	Venezuela	156
NS06	22	M	4.8	Mexico	156
NS07	26	F	5	Uruguay	132
NS08	35	M	5	Argentina	84
NS09	32	F	4.8	Chile	96

Table A2

L2 Speaker Characteristics

Participant	Age at time of testing	Gender	Age of first contact with Spanish	Other L1	Self-rating (1-4)	Percentage of Use of Spanish	Native Speaker Judge Rating (/5)	Number of months of formal Spanish instruction
AS01	28	F	21		Advanced (3)	20	3.6	24
AS02	29	M	15		Near-native (4)	38	3	83
AS04	26	F	17		Near-native (4)	38	2.2	96

AS05	26	F	14		Near-native (4)	25	3.8	48
AS06	33	F	17		Advanced (3)	13	3.4	60
AS07	22	F	11	Cantonese	Intermediate (2)	42	2.8	60
AS10	23	M	16		Near-native (4)	20	4.8	12
AS12	21	F	15	Mandarin	Advanced (3)	15	1.4	72
AS13	22	M	16	Tamil	Advanced (3)	5	3	48
AS14	44	F	31		Advanced (3)	5	3.6	108
AS15	20	F	10		Intermediate (2)	18	2.4	120
AS16	23	F	14		Near-native (4)	21	2.6	96
AS17	28	F	15		Near-native (4)	41	2.8	84
AS18	25	M	19		Intermediate (2)	5	2.8	12
AS19	34	M	20		Advanced (3)	33	2	120
AS20	23	F	16		Advanced (3)	8	1.8	12
AS22	21	M	18		Advanced (3)	9	3	6
AS23	62	F	14		Advanced (3)	5	1.4	48
AS24	36	M	18	Somali	Advanced (3)	17	3.4	32
AS25	23	F	18		Advanced (3)	3	1.8	24
AS26	23	M	17		Near-native (4)	43	4.6	72
AS29	20	M	18		Intermediate (2)	5	1	18

Appendix B

Perception Task Stimuli

Table B1

*Perception task stimuli. For each pair of utterances, the gates are indicated with a slash in the question and the stressed syllables are indicated in **bold** in the statement.*

E 1_1	¿Tienen/ helado de choco/late?
E 1_2	**Tien**en he**la**do de choco**la**te.
E 2_1	¿Puedo/ ir al /parque?
E 2_2	**Pue**do ir al **par**que.
E 3_1	¿Puedo/ na/dar?
E 3_2	**Pue**do na**dar**.
E 4_1	¿Se puede/ pagar con tar/jeta?
E 4_2	Se **pue**de pa**gar** con tar**je**ta.
E 5_1	¿Le duele/ la gar/ganta?
E 5_2	Le **due**le la gar**gan**ta.
E 6_1	¿Tienen/ ca/fé?
E 6_2	**Tien**en ca**fé**.
E 7_1	Es/ tu/yo.
E 7_2	¿**Es tu**yo?
E 8_1	¿No se puede/ fu/mar?
E 8_2	No se **pue**de fu**mar**.
E 9_1	¿Esta es/ la nueva profesora de portu/gués?
E 9_2	**Es**ta es la **nue**va profe**so**ra de portu**gués**.
E 10_1	¿Está llo/vien/do?
E 10_2	Es**tá** llo**vien**do.
E 11_1	¿Están ce/rra/dos?
E 11_2	Es**tán** ce**rra**dos.
E 12_1	¿Hay cla/se /hoy?
E 12_2	**Hay cla**se **hoy**

Appendix C

List of Production Task Scenarios (Expected answer in **bold**)

Escena 1

Parte 1

Hace calor. Un papá llega con su hijo a la heladería. Al niño sólo le gusta el helado de chocolate así que el papá le pregunta al empleado de la heladería:

¿Tienen helado de chocolate?

Desafortunadamente no tienen y entonces el niño dice que no quiere nada.

¿Qué le dice el papá al empleado?

Parte 2

Después de algunos días padre e hijo pasan de nuevo por la misma heladería. Cuando el papá le pregunta si quiere helado el niño dice: *no quiero ir, no tienen lo que me gusta*. El padre mira y ve al empleado con un helado de chocolate y le dice al niño:

Tienen helado de chocolate.

El niño se queda contento.

¿Qué le dice el papá a su hijo?

Escena 2

Parte 1

Un niño quiere saber si puede ir al parque. Le pregunta a su mamá:

¿**Puedo ir al parque?**

La mamá le dice: 'sí, hijo puedes ir al parque pero tienes que volver a casa a las 8 para cenar'.

¿Qué le pregunta el niño a su mamá?

Parte 2

A las 9 el niño no ha llegado todavía y la familia está preocupada. El nuevo vecino quiere ayudar con la búsqueda del niño. La mamá le dice al vecino: *no sé si nos puede ayudar.* Y el vecino dice: sí claro,

Puedo ir al parque.

Y la madre: ¿seguro que puedes? ¿Sabes cómo llegar?

¿Qué dice el vecino?

Escena 3:

Parte 1:

Una niña y su mamá salen a caminar y pasan por una piscina. A la niña le encanta nadar y le pregunta a su mamá:

¿Puedo nadar?

La madre le dice que puede después de recoger su traje de baño.

¿Qué le pregunta la niña a la madre?

Parte 2:

En la clase de Luis la profesora muestra una imagen de un río y les pregunta cómo pueden cruzarlo. Luis dice:

Puedo nadar.

¿Qué dice Luis?

Escena 4:

Parte 1

Pedro quiere un celular nuevo. Cuando llega el momento de pagar se da cuenta de que no tiene efectivo. Saca su tarjeta del bolsillo y le pregunta al empleado:

¿Se puede pagar con tarjeta?

El empleado dice que hoy no funciona la máquina y entonces Pedro no puede pagar.

¿Qué le pregunta Pedro al cajero?

Parte 2

Pedro vuelve a su casa y quiere una pizza pero como no pasó por el banco todavía no tiene efectivo. Su amigo le dice que no hay ningún problema:

se puede pagar con tarjeta.

Y Pedro dice: perfecto! Aquí está mi Visa.

¿Qué le dice el amigo?

Escena 5:

Parte 1:
Juan está enfermísimo y perdió la voz. Su amiga María lo acompaña al médico. Cuando entran a su oficina y el médico ve que Juan se toca la garganta, el médico le pregunta:

¿Le duele la garganta?

Juan dice que sí con la cabeza.

¿Qué le pregunta el médico a Juan?

Parte 2:

El médico le recomienda que tome unas pastillas y que descanse un par de días. Al salir del médico, un amigo de Juan lo ve y le pregunta: oye Juan, ¿qué tienes? ¿Por qué fuiste al médico? y María contesta:

Le duele la garganta.

¿Qué le dice María al amigo de Juan?

Escena 6:

Parte 1:
Raúl entra en una panadería donde se puede desayunar. Tienen agua, jugos y té pero no los quiere. Le pregunta al empleado:

¿Tienen café?

El empleado dice que sí tienen y le prepara un café.

¿Qué le pregunta Raúl al empleado?

Parte 2:

En la clase los alumnos tienen que decir qué producto tiene cada país. Cuando llega el momento de Javier, la maestra le pregunta: ¿qué tienen en Colombia? Y Javier:

Tienen café.

¿Qué dice Javier?

Escena 7:

Parte 1:

María y Ana entran a un café cuando de repente suena el celular de Ana y tiene que contestar. Mientras tanto, María pide un par de cafés y se sienta en una mesa. Cuando Ana entra, ve el café y le pregunta a su amiga: ¿y ese café? ¿De quién es?

Y María: **es tuyo**. Lo pedí mientras estabas hablando por teléfono.

¿Qué le dijo María a Ana?

Parte 2:

Juan recoge a Ana en el aeropuerto. Ella no reconoce el coche y cuando entra dice:

¿y este coche? **¿Es tuyo?**

Y Juan: sí. Es mío. Me lo regaló papá.

¿Qué le preguntó María a Juan?

Escena 8:

Parte 1:

José y Luis dan un paseo por el parque. José es un fumador pero cuando saca un cigarrillo el amigo Mario le hace *no* con las manos. José dice:

¿Y por qué no? **¿No se puede fumar?**

Mario le dice que no y José guarda el cigarrillo.

¿Qué le preguntó José a Mario?

Parte 2:

José y Mario se van a una cafetería y José enciende su cigarrillo. Un empleado se acerca y les dice:

lo siento pero aquí **no se puede fumar**.

José apaga el cigarrillo.

¿Qué le dijo el empleado a José?

Escena 9:

Parte 1:

Llega una nueva profesora de portugués al departamento. María tiene que entregarle un ensayo pero no la conoce todavía. Mientras está en el departamento ve a una mujer que no conoce y le pregunta a la secretaria:

¿Esta es la nueva profesora de portugués?

La secretaria le dice que sí y María se presenta.

¿Qué le preguntó María a la secretaria?

Parte 2:

En una reunión de profesores, el jefe del departamento quiere presentar la nueva profesora de portugués al resto del claustro. A la nueva profesora le dieron el numero de salón equivocado así que cuando empieza la reunión ella no está. Finalmente logra encontrar el salón correcto y cuando entra el jefe del departamento la presenta diciendo:

Esta es la nueva profesora de portugués.

¿Qué dice el jefe del departamento?

Escena 10:

Parte 1:

Juan tiene que salir para el trabajo y no sabe si llevar el paraguas. Le pregunta a su hermana desde el sótano:

¿Está lloviendo?

Y la hermana: sí, está lloviendo. ¡Llévate un paraguas!

¿Qué le pregunta Juan a su hermana?

Parte 2:

Javier llama a su amigo Juan y le dice que quiere salir a jugar futbol en el parque. Juan ya está en el parque y le dice:

No se puede, **está lloviendo.**

Entonces, deciden ir a ver una película.

¿Qué le dice Juan a Javier?

Escena 11:

Parte 1:

María llega a la tienda para comprar zapatos y la puerta está cerrada. Ve a un empleado que sale de la tienda y le pregunta:

¿Están cerrados?

El empleado le explica a María que desafortunadamente se fue la luz y tuvieron que cerrar.

¿Qué le pregunta María al empleado?

Parte 2:

María vuelve a su casa y el hermano le pregunta: ¿Compraste los zapatos? Y María le explica:

No. **Están cerrados.**

¿Qué dijo María?

Escena 12:

Parte 1:

Hoy es el último día de clase y Juan llega tarde al salón. Cuando llega ve que no hay nadie y se pregunta:

¿Hay clase hoy?

Mira su correo y ve que el prof mandó un correo a todos avisándoles que no había clase.

¿Qué se pregunta Juan?

Parte 2:
Javier necesita ver algo en el auditórium pero cuando llega ve que está lleno de alumnos. Le pregunta a alguien qué está pasando y él le contesta:

Hay clase hoy.

Javier no sabía que el auditórium se usaba como clase también pero decide volver más tarde.

¿Qué le dice el chico a Javier?

Appendix D

The North Wind and the Sun: Spanish Version

El viento del norte y el sol discutían acerca de cuál de los dos sería el más fuerte, cuando, de repente, pasó un viajero envuelto en una amplia capa. Al verlo, convinieron en que el primero que consiguiera quitarle la capa sería el más fuerte. El viento del norte comenzó a soplar con mucha furia, pero, cuanto más soplaba, más se aferraba el viajante a su capa, hasta que el viento norte desistió. El sol brilló entonces con todo su esplendor, e inmediatamente, el viajante arrojó su capa. Así, el viento norte tuvo que reconocer la superioridad del sol.

Appendix E

Language Questionnaire

Department of Spanish & Portuguese

University of Toronto

Personal Information

- Sex: □ Male □ Female
- Year of Birth: _____
- Place of Birth: City _____ Country _____
- Occupation: _____
- Highest Level of Schooling: □ Secondary □ CEGEP/College/Professional □ University
- If you were not born in Canada, at what age did you move here? _____
- Are you left-handed or right-handed? _____

First Language

What is your first language? _____

What is the first language of: your mother? _____ your father? _____

Did you learn your first language from birth? □ Yes □ No

- If you answered 'No' to the question above, please explain:

  _____

Which language(s) did you speak at home as a child? _____

Is your first language the language with which you are the most comfortable? □ Yes □ No

- If you answered 'No' to the question above, please explain:

  _____

Education & Language Use

Which language(s) were you formally educated in? Where (i.e. country)?

Primary/Elementary School _____

High School _____

CEGEP/College _____

University _____

Which language(s) do you use (Indicate approximate percentage, e.g. 0, 50, 100%):

At school _____

At home _____

At work _____

In social situations _____

D. Second Languages

	Second Languages	
	A.	**B.**
At what age did you begin to learn your 2nd language?		
Where did you learn your 2nd language? Give place and years.		
Were your teachers native speakers of this language?		
Did you learn this language as a subject or was it the principal medium of instruction?	☐ Subject ☐ Medium of Instruction	☐ Subject ☐ Medium of Instruction
Have you ever spent time in an area where this language was the native language?	Where? How long?	Where? How long?
Approximately how many hours a week do you use this language? Specify for each of speaking, listening and reading.	Speaking: _____ hrs Listening: _____ hrs Reading: _____ hrs	Speaking: _____ hrs Listening: _____ hrs Reading: _____ hrs

- *Please rate your linguistic ability in each of your second languages in the following areas by checking the appropriate answer.*

	Beginner	**Intermediate**	**Advanced**	**Near-Native**
READING				
Language A				
Language B				
WRITING				
Language A				
Language B				
SPEAKING				
Language A				

Language B				
LISTENING				
Language A				
Language B				
OVERALL COMPETENCE				
Language A				
Language B				

Do you know any other second languages? Please specify:

_____

Appendix F

Number of accurate responses (maximum 12) for each participant at each gate.

	Question			Statement		
	Gate 1	Gate 2	Gate 3	Gate 1	Gate 2	Gate 3
NS01	7	8	11	8	9	12
NS02	8	9	12	11	11	12
NS03	6	10	12	8	8	10
NS04	9	9	11	4	3	7
NS05	11	11	11	11	12	12
NS06	6	7	10	5	6	12
NS07	10	9	12	10	10	12
NS08	11	10	12	10	10	11
NS09	11	10	11	7	8	11
AS01	10	11	12	10	11	12
AS02	9	11	12	11	12	12
AS04	11	12	12	9	10	12
AS05	10	11	12	7	7	12
AS06	11	10	11	8	11	11
AS07	12	11	12	8	8	11
AS10	12	11	12	10	11	12
AS12	7	8	11	9	8	12
AS13	9	9	12	8	10	12
AS14	12	12	12	12	12	12
AS15	10	9	12	8	10	11
AS16	9	10	11	11	12	12
AS17	10	9	12	9	10	12
AS18	10	10	12	10	11	12
AS19	8	10	12	8	8	12
AS20	11	11	12	11	11	12
AS22	10	10	11	7	7	11
AS23	9	9	10	8	9	11
AS24	11	10	11	8	8	11
AS25	6	10	12	11	11	12
AS26	8	10	12	10	10	12
AS29	7	9	10	12	12	12

Appendix G

Mean RTs (ms) for each participant at all three gates

	Question			Statement		
	1	2	3	1	2	3
NS01	894.99	350.26	288.65	796.42	335.68	212.24
NS02	1047.19	580.896	275.45	847.68	626.92	246.57
NS03	1232.01	918.15	260.45	1256.92	482.90	359.51
NS04	1435.36	755.45	86.184	1664.92	619.92	138.94
NS05	1511.69	539.35	-189.55	1791.70	526.99	75.1565
NS06	2066.50	553.80	98.16	2277.84	557.09	357.99
NS07	1221.09	657.50	313.95	1144.96	698.41	360.24
NS08	2341.42	1076.75	1344.91	2488.22	998.14	1370.67
NS09	1778.40	846.92	697.81	2354.72	570.62	696.75
AS01	999.96	437.01	-26.06	1282.90	261.93	-27.51
AS02	1399.04	596.56	765.79	757.056	561.57	343.41
AS04	1150.73	528.08	-13.88	1069.70	373.39	27.82
AS05	1766.15	707.71	-28.38	1633.36	985.46	296.41
AS06	1734.98	1223.94	990.75	1129.37	1087.44	720.49
AS07	1438.62	989.06	476.36	1588.69	1134.81	171.77
AS10	1459.31	708.60	373.55	1785.83	673.83	304.82
AS12	1302.52	997.27	291.18	1371.55	468.34	199.30
AS13	1369.86	880.61	542.95	1496.35	949.24	458.91
AS14	1844.87	814.92	414.04	2052.16	559.49	381.49
AS15	1416.44	633.14	439.86	1822.90	861.58	547.52
AS16	1040.76	387.47	-267.73	1185.42	744.99	-22.76
AS17	2070.19	1125.50	300.11	1729.13	990.43	398.99
AS18	1066.29	626.89	174.86	1593.41	893.84	666.64
AS19	1045.78	810.45	640.45	1211.93	715.33	793.99
AS20	697.05	664.26	211.04	996.64	467.25	367.49
AS22	1437.58	588.25	352.27	2259.67	488.99	558.28
AS23	637.78	306.28	-155.34	932.54	689.85	-11.24
AS24	1155.87	417.05	124.72	1209.58	170.48	-52.77
AS25	580.30	552.89	16.70	801.48	369.53	57.32
AS26	1715.33	1060.14	469.79	1725.86	686.21	1011.75
AS29	1148.14	480.92	411.49	1318.59	704.74	706.57

Appendix H

T-test or Wicoxon Signed Rank Test Results for Statement and Question RT data for each participant

	Test Type	Test Statistic[23]	df	p-value
NS01	Wilcoxon	12		0.81
NS02	t-test	0.75	7	0.47
NS03	t-test	-0.13	5	0.90
NS04	Wilcoxon	2		0.37
NS05	Wilcoxon	27		0.63
NS06	t-test	-1.60	4	0.18
NS07	Wilcoxon	29		0.92
NS08	t-test	0.46	6	0.65
NS09	t-test	-0.98	6	0.36
AS01	Wilcoxon	20		0.49
AS02	Wilcoxon	37		0.09
AS04	Wilcoxon	21		0.91
AS05	Wilcoxon	17		0.68
AS06	Wilcoxon	32		0.054*
AS07	Wilcoxon	11		0.68
AS10	Wilcoxon	16		0.27
AS12	t-test	-0.21	6	0.83
AS13	Wilcoxon	18		1
AS14	t-test	-0.57	11	0.5
AS15	Wilcoxon	9		0.25
AS16	Wilcoxon	18		0.65
AS17	Wilcoxon	27		0.65
AS18	Wilcoxon	8		0.04*
AS19	Wilcoxon	11		0.38
AS20	t-test	-3.15	10	0.01*
AS22	t-test	-2.30	6	0.06
AS23	Wilcoxon	13		0.5
AS24	Wilcoxon	18		1

[23] t for t-test and Z for the Wilcoxon signed rank test

AS25	t-test	-1.79	5	0.13
AS26	Wilcoxon	21		0.74
AS29	t-test	-0.91	5	0.40

Appendix I

Number of Y/N questions and statements used (out of 12) for initial boundary tone (IBT) and pre-nuclear peak (PNP) for all participants

	IBT		PNP	
	Question	**Statement**	**Question**	**Statement**
NS01	9	9	9	9
NS02	10	9	10	9
NS03	8	5	8	5
NS04	6	8	6	8
NS05	9	10	9	10
NS06	11	8	11	8
NS07	10	10	10	10
NS08	8	6	8	6
NS09	9	8	9	8
AS01	11	12	11	12
AS02	10	12	11	12
AS04	8	11	8	11
AS05	9	9	9	9
AS06	10	11	10	11
AS07	10	8	10	8
AS10	10	12	10	12
AS12	8	9	8	9
AS13	6	10	6	10
AS14	8	11	8	11
AS15	5	9	5	9
AS16	10	10	10	10
AS17	10	8	10	8
AS18	9	8	9	8
AS19	9	9	9	9
AS20	10	12	10	12
AS22	9	5	9	5
AS23	5	6	5	6
AS24	9	9	9	9
AS25	9	6	9	6
AS26	8	12	8	12
AS29	6	9	6	9

Appendix J

Initial boundary tone values (Table J1) and graphic representation for native speakers (Figure J1) and L2 speakers (Figure J2)

Table J1

Mean and standard deviation of the initial boundary tone for each participant in Hertz.

	Question		Statement	
	Mean	SD	Mean	SD
NS01	212.26	46.22	211.19	13.68
NS02	238.85	12.63	228.35	11.35
NS03	139.18	14.15	135.13	12.86
NS04	231.18	14.97	201.85	54.75
NS05	132.27	10.30	119.25	7.82
NS06	165.33	30.33	163.45	5.38
NS07	244.11	49.38	257.71	20.58
NS08	157.60	16.52	142.51	9.09
NS09	183.28	76.94	220.58	66.85
AS01	210.18	23.89	206.08	9.34
AS02	165.93	94.30	168.63	95.24
AS04	193.30	55.60	196.43	46.67
AS05	230.94	21.67	224.46	20.64
AS06	213.95	15.26	217.04	19.28
AS07	230.49	21.66	211.68	43.64
AS10	122.33	12.99	112.50	28.39
AS12	236.56	5.25	224.33	49.60
AS13	103.29	28.78	117.84	15.51
AS14	192.61	9.75	178.64	11.00
AS15	214.25	7.08	225.44	9.12
AS16	280.48	15.84	283.67	26.75
AS17	250.16	10.56	259.37	13.05
AS18	125.04	9.00	116.95	27.17
AS19	147.44	6.68	132.35	15.03
AS20	182.66	58.44	214.16	38.65
AS22	115.97	6.87	109.91	6.34
AS23	157.96	18.41	159.19	18.40
AS24	120.04	9.85	128.86	9.30
AS25	253.83	11.00	249.93	12.04

AS26	118.38	32.92	134.15	10.71
AS29	187.54	136.90	123.36	14.43

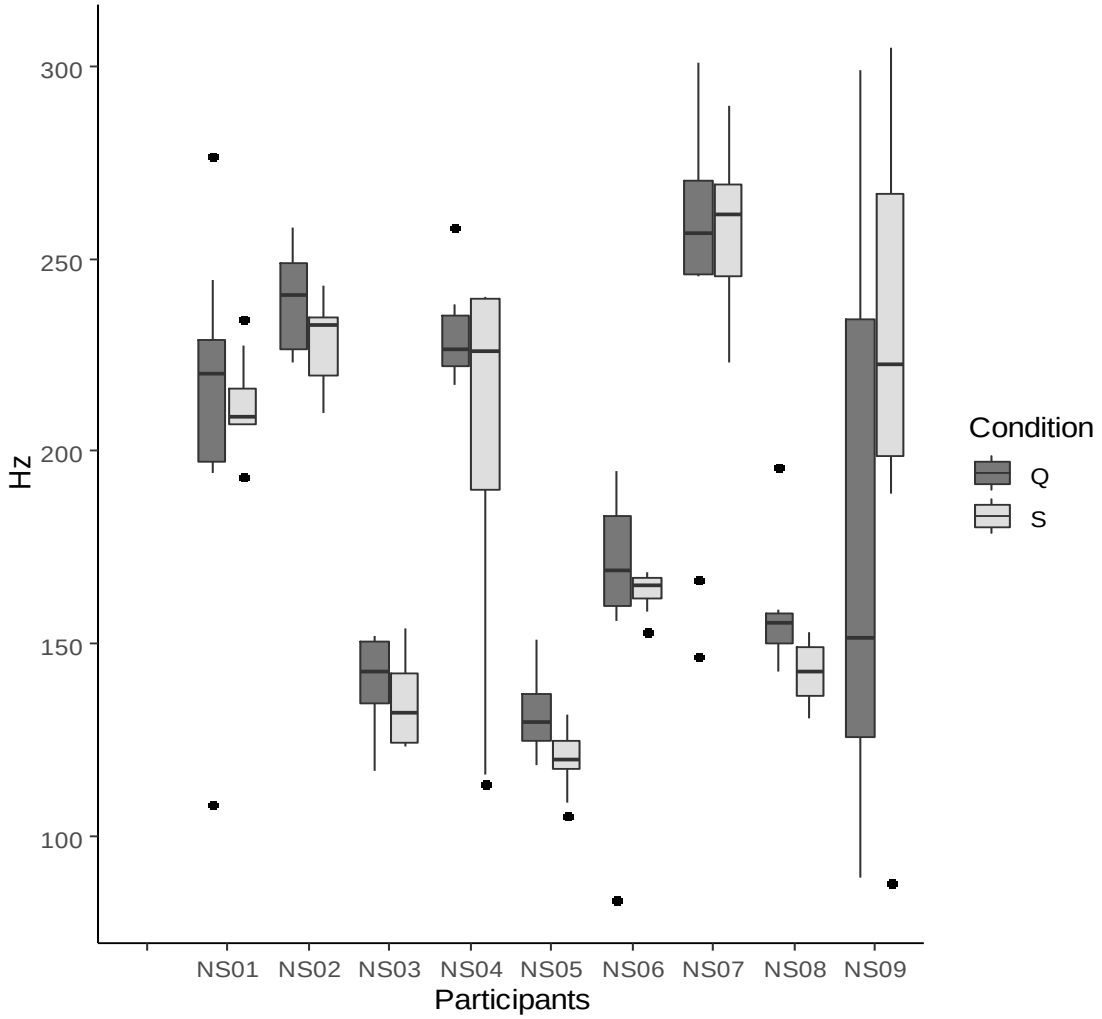

Figure J1. Initial boundary tone values for each native speaker in Hz.

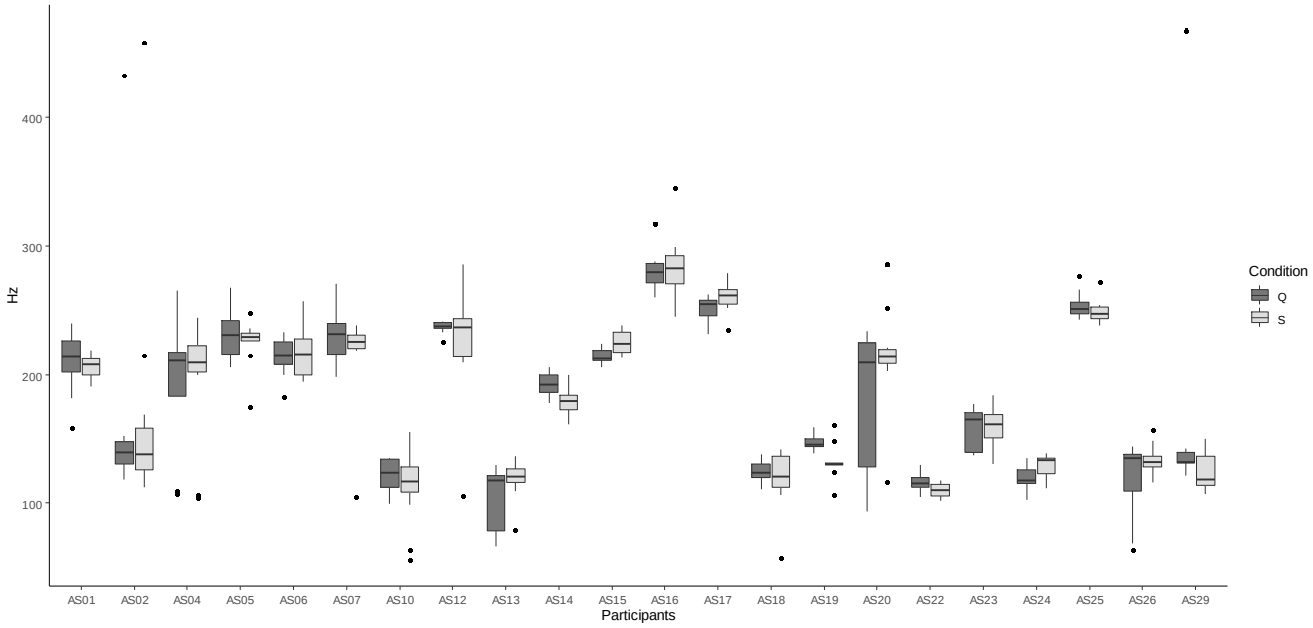

Figure J2. Initial boundary tone values for each L2 speaker Hz.

Appendix K

Pre-nuclear peak values (Table K1) and graphic representation for native speakers (Figure K1) and L2 speakers (Figure K2).

Table K1

Mean and standard deviation of the pre-nuclear peak for each participant in Hz.

	Question		Statement	
	Mean (Hz)	SD	Mean (Hz)	SD
NS01	230.25	17.51	206.30	21.19
NS02	246.48	20.90	224.29	8.88
NS03	148.05	9.53	147.40	6.31
NS04	289.48	32.25	251.49	61.46
NS05	166.97	14.59	139.67	15.34
NS06	193.06	17.76	177.86	10.27
NS07	387.89	46.70	312.83	56.70
NS08	180.95	15.75	165.45	20.46
NS09	276.52	66.88	263.64	73.31
AS01	227.98	25.13	215.32	18.60
AS02	203.75	103.43	139.08	28.27
AS04	223.11	11.87	216.52	15.79
AS05	241.31	36.90	238.38	25.28
AS06	250.94	26.16	211.02	25.78
AS07	251.70	42.63	229.57	16.68
AS10	149.08	29.98	122.83	15.48
AS12	230.96	19.24	215.14	12.00
AS13	135.80	13.23	138.88	12.46
AS14	191.74	9.64	184.50	13.48
AS15	241.83	41.46	213.58	13.24
AS16	255.99	8.36	260.46	13.15
AS17	268.64	22.98	273.49	24.57
AS18	131.67	10.57	113.87	7.43
AS19	164.50	11.19	146.89	13.28
AS20	228.95	47.35	215.53	36.74
AS22	112.44	5.99	112.04	5.01
AS23	174.85	33.40	173.80	20.04
AS24	146.38	38.15	128.28	12.38

AS25	236.88	18.05	229.99	11.49
AS26	142.12	8.25	137.83	7.78
AS29	138.62	5.94	129.09	16.55

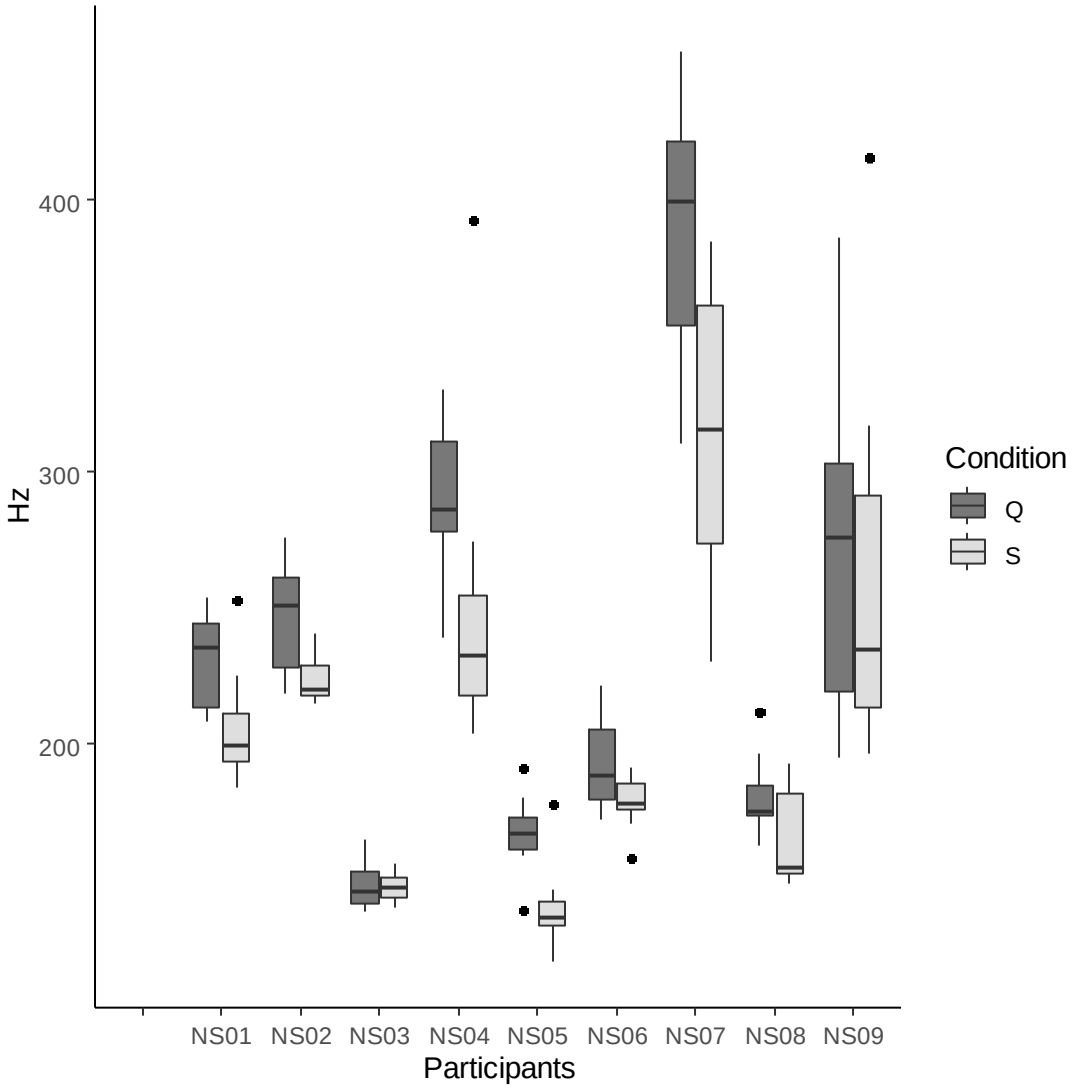

Figure K1. Pre-nuclear peak values for each native speaker in Hz.

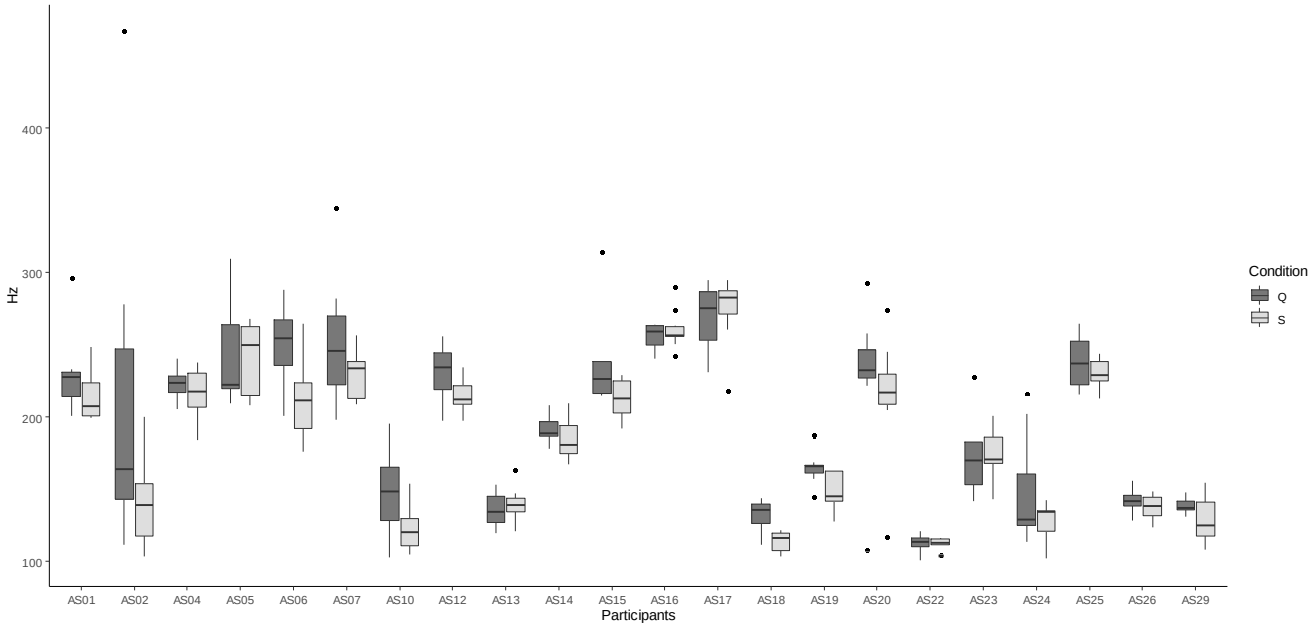

Figure K2. Pre-nuclear peak values for each