# Exposure and Empathy on SLA of Intonation

Robert Esposito

Department of Spanish and Portuguese, Rutgers University

# Author Note

The authors have no conflicts of interest to disclose.

Correspondence concerning this article should be addressed to Robert Esposito, Department of Spanish and Portuguese, Rutgers University, 15 Seminary Pl., New Brunswick, NJ 8901, USA, Email: [rme70@rutgers.edu](mailto:rme70@rutgers.edu)

# Abstract

The abstract goes here.

*Keywords*: intonation, prosody, Spanish, individual differences

# Exposure and Empathy on SLA of Intonation

INTRODUCTION GOES HERE.

# 1. Background and Motivation

SUMMARY OF LIT REVIEW GOES HERE

## 1.1 L2 Intonation Acquisition

Investigation in first language (L1) intonational phonology has exploded since the groundbreaking work by Ladd ([2008](#ref-ladd2008intonational)) and colleagues ([Pierrehumbert, 1980](#ref-pierrehumbert1980phonology)), but research on second language acquisition (SLA) of intonation has lagged behind. Second language (L2) learners face challenges acquiring the target language’s intonation system largely because of its many-to-many relationship between sound and meaning. That is, intonation serves to signal various and distinct communicative functions across syntax, pragmatics, and semantics, such as constituent focusing, affective meaning, politeness, and sentence modality. Taking the last item as an example, sentence modality in Spanish may be indicated solely by intonation; the utterance ‘Mariano habla del tiempo’ *Mariano talks about the weather* may variably be a statement or question depending on the intonational contour.

Current SLA models, such as Speech Learning Model (SLM) revised ([Flege & Bohn, 2021](#ref-flege2021revised)), primarily focus on segmental phonology. SLA suprasegmental phonology has received much less attention, one of the only models developed being the L2 Intonation Learning Theory (LILt, [Mennen, 2015](#ref-mennen2015beyond)). Similar to the SLM revised, the LILt model proposes that intonation acquisition outcomes depend on cross-language similarities and differences on one or more of four dimensions (systemic, realizational, semantic, and frequency). L2 learners whose target language converges on one or more dimensions between their L1 and L2 are predicted to exhibit positive transfer, whereas differences in dimensions will produce negative transfer.

A ripe area for investigating L1-L2 transfer is L1 English L2 Spanish perception of Spanish neutral yes-no questions that end in a low boundary tone, such as those found in Caribbean ([Armstrong, 2010](#ref-armstrong2010puerto); [Bailey, 2014](#ref-bailey2014intonational); [Willis, 2010](#ref-willis2010dominican)), Galician Spanish ([Pérez Castillejo & Fuente Iglesias, 2024](#ref-perez2024basic)), as well as other varieties ([Astruc et al., 2010](#ref-astruc2010venezuelan)). There has been little research directly investigating the acquisition of a low boundary tone to signal yes-no questions in L2 Spanish learners, possibly due to variety biases in pedagogical material; that is, L2 Spanish learners are not typically exposed to the varieties that feature this element in the classroom.

George ([2024](#ref-george2024l2)) investigated the acquisition of L2 Spanish sentence modality, cued by intonation alone, across 7 Spanish varieties. Participants completed a two-alternative forced choice task in which they had to determine if an utterance was an interrogative or not. The task included 21 utterances, evenly divided between broad-focused declarative statements, information-seeking yes-no questions, and wh-questions, with each of the 7 dialects contributing one utterance of each sentence type. Although participants performed at ceiling, the majority of misidentifications were found in the Puerto Rican yes-no questions, which feature a low boundary tone.

In a similar vein, Casillas et al. ([2023](#ref-casillas2023using)) presented a conceptual replication of Brandl et al. ([2020](#ref-brandl2020development)). 225 L1 English L2 Spanish participants, with a range of proficiency levels from functionally monolingual to advanced, completed a two-alternative forced choice task, determining if a given utterance was an interrogative or not. 8 Spanish varieties were included, two of which (Puerto Rican and Cuban) signaled information-seeking yes-no questions with a low boundary tone. Participants performed least accurately with these two varieties compared to the other 6 varieties.

The acquisition of Venezuelan Spanish by L1 English L2 Spanish learners, another variety that signals information-seeking yes-no questions with a low boundary tone, has also been investigated by Trimble ([2013b](#ref-trimble2013perceiving)). He presented a conceptual replication of Face ([2007](#ref-face2007role)), including L2 Spanish learners of varying proficiency levels, as well as advanced L2 learners who had no study abroad experience or had studied in Spain or Venezuela. Participants completed a two-alternative forced task in which they had to accurately broad-focus declaratives and information-seeking yes-no questions. All participants performed at ceiling when classifying Spain utterances, as well as Venezuelan declaratives, but accurate identification of Venezuelan yes-no questions depended on both proficiency and study abroad experience, such that higher proficiency participants were more accurate and participants who had studied abroad in Venezuela were most accurate. These findings corroborate George ([2024](#ref-george2024l2)) and Casillas et al. ([2023](#ref-casillas2023using)) in that the low boundary tone signaling yes-no questions in Spanish poses issues for L1 English L2 Spanish learners.

On the other hand, Bedialauneta Txurruka ([2023](#ref-bedialauneta2023perception)) found contradicting results. She investigated the perception of Argentinian information-seeking yes-no questions, which are signaled by a circumflex rise-fall boundary tone, in L1 English L2 Spanish learners who have been learning for at least fives years (no proficiency measures were reported). Participants completed a two-alternative forced choice task to identify Peninsular Spanish and Argentinian Spanish broad-focus declaratives and information-seeking yes-no questions. Interestingly, L2 participants identified Argentinian utterances with accuracy that did not prove to be meaningfully different from L1 Argentinian Spanish speakers. She attributes the unexpected findings to secondary cues to intonation ([Face, 2007](#ref-face2007role)), or possibly that the relative height of the last accentual peak before the final fall is enough to condition interrogativity for L1 English speakers ([Gabriel et al., 2010](#ref-gabriel2010argentinian)).

The present study aims to directly target L2 Spanish acquisition of the low boundary tone for yes-no questions, specifically in Puerto Rican and Cuban Spanish. Following the LILt model and SLA research that includes these varieties, it is expected that L1 English L2 Spanish learners will face difficulties correctly identifying this sentence type for these dialects, but that accurate identification will be conditioned in part by the learner’s proficiency level.

## 1.2 L2 Intonation in the Language Classroom

Despite the various systematic functions of intonation within language, it still is sidelined in the L2 classroom and pedagogical materials ([Jenkins, 2004](#ref-jenkins20045); [Rao, 2019](#ref-rao2019key)). As with other linguistic areas, cross-linguistic transfer effects may impose meaningful issues for L2 learners, such as if they wish to express politeness. Herrero and Devı́s ([2020](#ref-herrero2020unintentional)), for example, investigated L1 Mandarin Chinese L2 Spanish learners production of polite requests in Spanish evaluated by L1 Spanish speakers. They found that even if the utterance by the L2 learners included lexical-grammatical aggression mitigators (such as ‘gracias’ *please*), L1 Spanish speakers relied primarily on intonation to evaluate the utterance’s politeness. In a similar vein, Estebas-Vilaplana ([2014](#ref-estebas2014evaluation)) investigated cross-linguistic differences between English and Spanish responses to wh- questions. A broad pitch range is polite in English, but over-excited in Spanish; a narrow pitch range is rude in English, but polite in Spanish. These studies work to highlight the importance of intonation in the L2 classroom.

Research on and evaluating methodologies of teaching intonation in the classroom remain limited. Olea ([2019](#ref-olea2019effectiveness)) studied the production of a single English intonation contour for implication in advanced L1 Spanish L2 English speakers. Participants completed a pre-test, a four-session intervention, and a post-test. The intervention consisted of listening to and then repeating the targeted intonation pattern. Participants showed meaningful improvement from the pre- to post-test, indicating the effectiveness of this intervention for production.

Sonsaat-Hegelheimer and Levis ([2025](#ref-sonsaat2025intonation)) investigated the acquisition of final intonation contours (falls, rises, and fall-rises), both production and perception, in L1 Turkish L2 English learners in a similar pre-test, intervention, and post-test design. The intervention in this case consisted of a three-week training in which participants were trained on a variety of material. Participants first were explicitly taught the sound-to-meaning mapping, and then participated in a variety of communicative activities that required accurate perception and production of the contours, as well as shadowing exercises. Importantly, Turkish’s utterance-level prosody is more constrained than Germanic languages ([Kabak, 2016](#ref-kabak2016refin)); boundary tones of declaratives, on the whole, tend to be low, and it was predicted that participants would have most issue with the rises and fall-rises, in line with predictions from the LILt model ([Mennen, 2015](#ref-mennen2015beyond)). These predictions were borne out in the pre-test, where participants had greater issues with the rises and fall-rises. In the post-test, participants displayed meaningful improvements in both the perception and production of these problem final contours, indicating that with training, cross-linguistic influence may be inhibited.

Both Olea ([2019](#ref-olea2019effectiveness)) and Sonsaat-Hegelheimer and Levis ([2025](#ref-sonsaat2025intonation)) used training methodologies that relied on intentional learning paradigms. However, these studies represent a very small minority of what teaching intonation in the classroom looks like. As intonation is not typically explicitly or intentionally taught in the classroom, students most likely rely on L1-L2 transfer ([Ortega-Llebaria & Colantoni, 2014](#ref-ortega2014l2)) or incidental learning through input from the instructor or other auditory input. Incidental learning in the current investigation is based on a definition provided by Hulstijn ([2003](#ref-hulstijn2003incidental)): “the ‘picking up’ of words and structures, simply by engaging in a variety of communicative activities… during which the learner’s attention is focused on the meaning rather than on the form of language” (p. 1). The majority of research on intentional and incidental learning has focused primarily on vocabulary (and spelling), and secondarily on morphology and syntax. Phonetics and phonology, on the other hand, has not received attention where the research goal makes clear a distinction between intentional and incidental learning.

The current investigation aims to determine if L1 English L2 Spanish learners are able to rapidly acquire accurate perception of Spanish neutral yes-no questions that are marked by low boundary tones, which according to the LILt model should present difficulties, through training on naturalistic conversations between L1 speakers of varieties that feature this sound-to-meaning mapping. The training provided is incidental in nature, as the participants must answer a content-focused question based on each conversation, and not the targeted linguistic form.

## 1.3 Individual Differences in Empathy

As of now, individual pragmatic differences in learners in regards to adult L2 intonational phonology acquisition has received little attention. Although individual differences have been observed for intonation processing ([J. Bishop, 2016](#ref-bishop2016individual); [J. B. Bishop et al., 2015](#ref-bishop2015individual); [Diehl et al., 2008](#ref-diehl2008resolving); [Ward & Hirschberg, 1988](#ref-ward1988intonation)), researchers have recently proposed that individual differences in *empathy* may account for some of the processing variation observed. Empathy, one’s ability to infer the intentions of others ([Baron-Cohen & Wheelwright, 2004](#ref-baron2004empathy)), has recently become a focus of attention in L1 and L2 intonation research. Empathy has been associated with theory of mind, mindreading, mentalizing, and perspective-taking ([Baron-Cohen, 2011](#ref-baron2011zero); [Carruthers, 2009](#ref-carruthers2009mindreading); [Frith & Frith, 2003](#ref-frith2003development)). Typically, empathy is categorized by *affective empathy*, or emotional alignment with the interlocutor, and *cognitive empathy*, to recognize and understand what an interlocutor feels or thinks. The latter categorization, cognitive empathy, has clear ties to the linguistic domain: empathy is necessary to comprehend an interlocutor’s motives in verbal interactions, especially for the interpretation of non-literal meaning.

A growing body of research has provided evidence that the construct empathy explains some variability in intonation perception in the L1 ([Esteve-Gibert et al., 2016](#ref-esteve2016role), [2020](#ref-esteve2020empathy); [Orrico et al., 2025](#ref-orrico2025individual); [Orrico & D’imperio, 2020](#ref-orrico2020individual)), and preliminary evidence in the L2. For example, Casillas et al. ([2023](#ref-casillas2023using)) found an interaction between proficiency and empathy score in L1 English L2 Spanish learners, such that lower-proficiency learners with higher empathy scores, measured with the Empathy Quotient ([Baron-Cohen & Wheelwright, 2004](#ref-baron2004empathy)), performed more accurately when identifying an utterance as a question or not. Apart from this one example, there has been no modern SLA research relating empathy to intonation or phonology in general.

Although the link between empathy and intonation processing has been established in L1 and L2 research, the mechanism and motivation for the link has not been thoroughly and explicitly investigated. One possibility is that L2 learners with higher empathy store more information about the linguistic content, even if it is not the focus of a given communicative activity. For example, Melchers et al. ([2017](#ref-melchers2017oxtr)) found that higher-empathy participants more accurately responded to (non-linguistic) questions about brief videos more accurately than their lower-empathy participants. Although Melchers et al. ([2017](#ref-melchers2017oxtr)) described their videos-and-questions methodology as an “implicit learning paradigm”, and that differences in “implicit perception” may explain the results between higher- and lower-empathy participants, it must be clear that “implicit” has received a multitude of definitions within and without linguistics. For this reason, what Melchers et al. ([2017](#ref-melchers2017oxtr)) described as “implicit” will here be reinterpreted as “incidental” ([Hulstijn, 2003](#ref-hulstijn2003incidental)). That is, participants incidentally learned and retained information from the brief videos. In a similar way, participants in a linguistic study may incidentally acquire vocabulary through reading (e.g., [Horst et al., 1998](#ref-horst1998beyond)).

The current research aims to contribute to the SLA research by investigating a possible mechanism through which individual differences in empathy impact L2 intonation processing. Inspired by Melchers et al. ([2017](#ref-melchers2017oxtr)), it seems possible that higher-empathy individuals store more information during social interactions than lower-empathy individuals, which may help them acquire a novel boundary tone to meaning association more rapidly than their proficiency-matched lower-empathy peers.

# 2. Present Study

I investigated the relationship between proficiency, empathy, and the rapid acquisition of a novel intonation contour in L1 English L2 Spanish learners. This study was preregistered on the Open Science Framework (https://osf.io/p6vuz) and designed to address the following research questions:

1. Do higher-empathy L1 English L2 Spanish speakers make greater use of prosodic cues to determine if a Spanish utterance is an interrogative?
2. Do higher-empathy L1 English L2 Spanish speakers make greater use of prosodic input from dialects to which they have little or no exposure to extract prosodic features?

With respect to RQ1, it is expected that higher-empathy learners will more accurately identify utterances as interrogatives, in line with Casillas et al. ([2023](#ref-casillas2023using)). As for RQ2, it should be recalled that Casillas et al. ([2023](#ref-casillas2023using)) found that participants performed less accurately when identifying yes-no questions from Spanish dialects with a low boundary tone (as opposed to a high boundary tone). As such, it is predicted that when participants are exposed to these dialects from realistic conversations before performing the given task, higher-empathy individuals will store more information about the form and function of the low boundary tone as opposed to lower-empathy individuals, and thus perform more accurately ([Melchers et al., 2017](#ref-melchers2017oxtr)).

# 3. Methodology

INTRODUCE METHODOLOGY

## 3.1 Participants

Two experiments were launched on the online experimental platform Prolific.ac: one for the experimental group and one for the control group, the difference solely being that the experimental group was exposed to an additional task (see: [Section 3.2.1](#sec-exposure)).

Upon beginning either experiment, all participants responded to the question “Are you familiar with Caribbean (e.g., Caribbean, Puerto Rican) or Galician Spanish?” (a proxy for if the individual is familiar with a dialect that has a final fall contour for absolute interrogatives), to which they could choose from a drop-down menu either “yes” or “no”. Participants were recruited until both groups had 100 participants each that responded “no” to this question.

To address RQ1, all participants in the control group were included in the analysis (n = 140). For RQ2, only participants who responded that they were unfamiliar with the above-specified dialects were included.

Both groups were compensated at a rate of $12.00 per hour. Participants in the control group were paid $3.00, while participants in the experimental group, which was approximately five minutes longer, were paid $4.00 upon completing all tasks. The median time of completion for the control group was approximately # A tibble: 1 × 1, mean\_exp\_time, , 1 14.0 minutes, while the median time for the experimental group was approximately # A tibble: 1 × 1, mean\_exp\_time, , 1 16.2 minutes.

The pool of participants for both groups were filtered primarily by using criteria set in Prolific.ac. Participants self-reported as being adult L1 English speakers born, raised, and currently living in the Northeastern US. Furthermore, participants had to have reported growing up as a monolingual English speaker. They reported no hearing difficulties and were required to use headphones on a personal computer.

In sum, all participants were adult L1 English L2 Spanish learners with varying levels of Spanish proficiency, ranging from functionally monolingual to highly proficient.

## 3.2 Tasks

The experimental and control groups performed essentially the same tasks: a 2AFC task, a lexical decision vocabulary assessment, and a Likert-type questionnaire to assess empathy. The experimental group completed an additional “exposure” task first. The tasks were programed in Python using PsychoPy3 ([Peirce et al., 2019](#ref-peirce2019psychopy2)) and hosted on Pavlovia. All code and material used to generate the tasks are freely available on the Open Science Framework.

### 3.2.1 Exposure Stimuli

Participants in the experimental group were first exposed to three naturalistic phone conversations. Each conversation was between two speakers of either Dominican, Puerto Rican, or Galician Spanish talking about everyday topics, such as going to the grocery store or making plans for dinner. These dialects were chosen specifically because they feature a final fall when producing neutral absolute interrogatives. Each pair of speakers was given a script to read, and they were told that they could modify any vocabulary to make it more natural to them, but not overall structure. Each conversation lasted approximately one minute. After each conversation, participants answered one fact-based yes-or-no question about the dialogue. [Table 1](#tbl-stimuli) documents the total number of absolute interrogatives produced in each conversation, as well as the count of absolute interrogatives that featured a final fall. In total, participants were exposed to 21 absolute interrogatives, 16 (76.19%) of which were produced with a final fall.

Table 1

Stimuli for exposure trials in experimental group.

| dialect | time (seconds) | # of absolute interrogatives | # of absolute interrogatives with final fall |
| --- | --- | --- | --- |
| Dominican | 60 | 10 | 7 |
| Galician | 47 | 5 | 3 |
| Puerto Rican | 36 | 6 | 6 |
| TOTAL | 143 | 21 | 16 |

### 3.2.2 2AFC

Both the experimental group and the control group completed the same 2AFC task. Participants were presented an audio file that was either a statement (broad focus) or a question (absolute interrogative). Participants were told to determine as quickly and accurately as possible if the utterance they heard was a question or not. At the start of each audio, they were prompted on-screen with the question, “Is this a question?” and they responded with their keyboard by typing either “1” for “yes” or “0” for “no”.

The auditory stimuli consisted of 88 items total. The audio stimuli were taken from a subset of those used in Casillas et al. ([2023](#ref-casillas2023using)), specifically from Cuban, Puerto Rican, Mexican, and Castilian speakers. Each speaker produced the same 22 utterances, 11 of which were broad focus statements and 11 of which were absolute interrogatives. The sentences were made up of three function words following a subject-verb-object (SVO) word order, the default in Spanish. The object was a noun with penultimate stress. The four native speakers produced the same items in a quiet room using professional recording equipment. The items were presented to the speaker on a screen, and they were asked to silently read the item to familiarize themselves with the content, and then to read it aloud. All utterances were segmented using Praat ([Boersma & Weenink, 2018](#ref-BoersmaWeenink2018)). A detailed description of the auditory stimuli is provided in the OSF repository from Casillas et al. ([2023](#ref-casillas2023using)) at: https://osf.io/zxkdt.

For RQ1, all 88 items were used for analysis; for RQ2, only 22 critical items were used for analysis. Those 22 critical items were absolute interrogatives from the Puerto Rican and Cuban speakers, who produced all absolute interrogatives with a final fall.

### 3.2.3 LexTALE

Spanish proficiency was assessed by administering the Lexical Test for Advanced Learners of Spanish (LexTALE-ESP, henceforth LexTALE) ([Izura et al., 2014](#ref-izura2014lexical); [Lemhöfer & Broersma, 2012](#ref-lemhofer2012introducing)). The LexTALE is a lexical decision experiment that provides a standardized assessment of vocabulary size in Spanish. Participants are exposed to a series of real words and pseudowords on the computer screen, and they must decide if the word is real or fake. LexTALE scores can range from -20 to 60. Monolingual Spanish speakers generally score above 50, whereas individuals with little or no knowledge tend to score negative. Adult learners with low to medium proficiency range from 0 to 25, and advanced learners generally score above 25. Proficiency here is treated as a continuous variable.

[Table 2](#tbl-lextale) shows the descriptive statistics from the participant samples used to address RQ1 and RQ2.

A Bayesian regression model was fit to LexTALE score as a function of *group* (baseline condition = control group) to determine if the sampled scores were practically equivalent. The intercept estimate was (β = 0.33, HDI = [-1.37, 2.07], ROPE = 0.09, MPE = 0.65). The experimental group did not meaningfully differ from the intercept ((β = -0.05, HDI = [-2.50, 2.34], ROPE = 0.07, MPE = 0.52)).

Table 2

Descriptive statistics of LexTALE scores.

| Dataset | Mean | SD | Min | Max |
| --- | --- | --- | --- | --- |
| RQ1 | 0.35 | 8.906327 | -17 | 41 |
| RQ2 (control) | 0.36 | 8.071048 | -16 | 31 |
| RQ2 (experimental) | 0.28 | 10.141504 | -26 | 55 |

### 3.2.4 Empathy Questionnaire

The construct empathy was assessed using the Empathy Quotient [EQ; Baron-Cohen and Wheelwright ([2004](#ref-baron2004empathy))]. The EQ consists of 60 items, 40 critical items and 20 distractors, that must be rated on a four-point Likert scale, ranging from “strongly agree” to “strongly disagree”. To avoid response bias, choices indicating empathic responses are coded to elicit “agree” responses in half the target items and “disagree” responses in the other half. The target items are scored with 2 or 1 points based on if the participant responds “strongly” or “slightly” to the “empathic” response, and 0 if the participants responds with the “non-empathic” response. The EQ is scored by summing the total points to produce a single value indicating an individual’s level of empathy. The minimum score is 0 (low empathy) and the maximum score is 80 (high empathy).

[Table 3](#tbl-eq) shows the descriptive statistics of EQ scores from the participant samples used to address RQ1 and RQ2.

A Bayesian regression model was fit to EQ score as a function of *group* (baseline condition = control group) to determine if the sampled scores were practically equivalent. The intercept estimate was (β = 39.45, HDI = [37.08, 41.73], ROPE = 0, MPE = 1). The experimental group had on average a slightly higher average, but did not meaningfully differ from the control group ((β = 2.06, HDI = [-1.14, 5.34], ROPE = 0.02, MPE = 0.9)).

Table 3

Descriptive statistics of EQ scores.

| Dataset | Mean | SD | Min | Max |
| --- | --- | --- | --- | --- |
| RQ1 | 42.09286 | 11.93260 | 15 | 74 |
| RQ2 (control) | 41.74000 | 12.07604 | 15 | 74 |
| RQ2 (experimental) | 41.63000 | 12.43135 | 12 | 67 |

## 3.3 Procedure

Participants were recruited on Prolific.ac and complete all tasks in a single session. The experimental and control groups identically completed tasks in the same order, with the exception that the experimental group first completed the exposure task. Afterwards, they completed the 2AFC, the LexTALE, and finally the EQ questionnaire.

I planned to collect data from 200 total individuals, 100 for the experimental group and 100 for the control group. Participants were recruited until 100 participants were found for each group who answered “no” to the question “Are you familiar with Caribbean (e.g., Caribbean, Puerto Rican) or Galician Spanish?” For RQ2, only these 200 participants were used for analysis; for RQ1, the total number of participants collected in the control group (n = 140) was used for analysis.

## 3.4 Statistical Analysis

I report two primary statistical analyses. The first analysis addresses RQ1, which was pre-registered, but the specific model structure was not. The second analysis addresses RQ2, which was pre-registered for both the research question and the model structure. Both analyses model response accuracy using Bayesian multilevel logistic regression. Both models’ likelihoods were Bernoulli distributed with logit link functions. All continuous variables were standardized. The criterion, *response*, was coded as “1” for correct responses and “0” for incorrect responses. That is, the models represented the probability of accurately responding to the prompt, “Is this a question?”.

The first analysis addressing RQ1 considered response accuracy for the population effects *caribbean* (‘0’ for Castilian and Mexican dialects; ‘1’ for Cuban and Puerto Rican dialects), *utterance type* (broad focus declarative and absolute interrogative), *LexTALE score* (i.e., Spanish proficiency), *EQ* (empathy quotient), and the higher order interactions. I specified group-level effects for participants, speaker variety, and items. The slope for *utterance type* varied for the participant effect, as did the *LexTALE* by *EQ* interaction for the speaker variety effect. The intercept represented the log-odds response accuracy of a declarative utterance produced by a non-Caribbean speaker at the average proficiency and EQ levels. For this model, all participants from the control experiment (i.e., participants not exposed to the naturalistic conversation stimuli), including those who responded that they were familiar with the previously specified dialects, were included, as well as all stimuli.

The second analysis addressing RQ2 considered response accuracy for the population effects *LexTALE score*, *EQ*, and *group* (experimental, control), as well as the higher order interactions. I specified group-level effects for participants, speaker variety, and items. The slope for *LexTALE* by *EQ* interaction varied for the speaker variety effect. For this model, 200 participants total were included, 100 from each the control and experimental group. All participants responded “no” to the question “Are you familiar with Caribbean (e.g., Caribbean, Puerto Rican) or Galician Spanish?”. Furthermore, only the critical items, absolute interrogatives produced by the Cuban and Puerto Rican speakers, were included in the analysis. The intercept represented the control group’s log-odds response accuracy of an absolute interrogative produced by a Caribbean Spanish speaker at the average proficiency and EQ levels.

For all models, I included regularizing, weakly informative priors ([Gelman et al., 2017](#ref-gelman2017prior)). I sample from the posterior distribution of a given model for statistical inferences. To assess my preregistered hypotheses, I established a region of practical equivalence (ROPE) around a point null value of 0 (see [Kruschke, 2018](#ref-kruschke2018rejecting)) using the following formula:

For all models, median posterior point estimates are reported for each parameter of interest, along with the 95% highest density interval (HDI), the percent of the region of the HDI contained within the ROPE, and the maximum probability of effect (MPE). For statistical inferences, I focus on estimation rather than decision-making rules, though, generally, a posterior distribution for a parameter in which 95% of the HDI falls outside the ROPE and a high MPE (i.e., values close to 1) are taken as compelling evidence for a given effect. I conducted all analyses using R and fit all models using the probabilistic programming language stan via the R package brms ([Bürkner, 2017b](#ref-burkner2017brms), [2017a](#ref-burkner2017advanced)).

# 4. Results

## 4.1 RQ1

The first analysis addressed RQ1, which aimed to determine if the findings from Casillas et al. ([2023](#ref-casillas2023using)) could be replicated. In their results, response accuracy to broad focus declaratives varied as a function of LexTALE score and EQ, such that low-proficiency participants with higher empathy performed more accurately than their proficiency-matched lower-empathy peers. Importantly, there was no effect found for the interaction between proficiency and EQ for absolute interrogatives.

The log-odds of a correct response for a non-Caribbean declarative utterance at the average proficiency and EQ levels was 0.18, or approximately 54.49%, (β = 0.18, HDI = [-0.43, 0.81], ROPE = 0.22, MPE = 0.72). Comparatively, participants performed less accurately for non-Caribbean absolute interrogatives ((β = -0.76, HDI = [-1.17, -0.34], ROPE = 0, MPE = 1)). In the case of Caribbean utterances, participants patterned similarly: declarative utterances ((β = 0.27, HDI = [−0.60, 1.12], ROPE = 0.16, MPE = 0.72)) were accurately identified more often than absolute interrogatives ((β = −2.34, HDI = [−3.39, −1.27], ROPE = 0.00, MPE = 1.00)). Although the median estimate for Caribbean declaratives is slightly above the median estimate of non-Caribbean declaratives, it should be noted that their HDIs significantly overlap, suggesting that there is no meaningful difference between the two estimates. Caribbean absolute interrogatives, on the other hand, are on the whole more negative than non-Caribbean absolute interrogatives, suggesting that individuals had overall more difficulty identifying Caribbean absolute interrogatives.

Figure ([**fig?**](#ref-fig)) plots response accuracy as a function of utterance type and proficiency (left panel) and EQ (right panel). Surprisingly, proficiency alone did not have a strong relationship with non-Caribbean declarative utterances ((β = −0.02, HDI = [−0.28, 0.24], ROPE = 0.59, MPE = 0.56)), and in fact had a negative relationship with non-Caribbean absolute interrogative utterances ((β = −0.15, HDI = [−0.38, 0.08], ROPE = 0.32, MPE = 0.90)) and Caribbean declarative utterances ((β = −0.16, HDI = [−0.44, 0.13], ROPE = 0.33, MPE = 0.86)). Caribbean absolute interrogative utterances, on the other hand, had a positive relationship with proficiency ((β = 0.20, HDI = [−0.01, 0.43], ROPE = 0.16, MPE = 0.97)).

Empathy, as opposed to proficiency, was a stronger predictor of response accuracy in general. Response accuracy to non-Caribbean declarative utterances ((β = 0.24, HDI = [−0.03, 0.49], ROPE = 0.13, MPE = 0.96)) and absolute interrogative utterances ((β = 0.21, HDI = [−0.02, 0.43], ROPE = 0.15, MPE = 0.96)) had largely positive posterior distributions. To a lesser extent, Caribbean declarative utterances had a positive relationship with empathy ((β = 0.10, HDI = [−0.18, 0.38], ROPE = 0.43, MPE = 0.77)), whereas there was little evidence to suggest a relationship between response accuracy to Caribbean absolute interrogative utterances and empathy ((β = −0.01, HDI = [−0.23, 0.21], ROPE = 0.68, MPE = 0.54)).

Lastly, the model estimates the proficiency by EQ interaction for each utterance type. There was weak evidence to suggest that the proficiency effect was modulated by EQ for non-Caribbean absolute interrogative utterances ((β = 0.21, HDI = [−0.02, 0.45], ROPE = 0.16, MPE = 0.96)). Figure ([**fig?**](#ref-fig)) plots the relationship. When holding standardized EQ values of -1, 0, and +1 constant for non-Caribbean absolute interrogative utterances across proficiency, the slope is more negative for lower EQ. In other words, learners with higher empathy at any given proficiency level tended to respond more accurately.

## 4.2 RQ2

The second analysis addressed RQ2, which questions whether L2 Spanish learners with higher empathy make greater use of prosodic cues when exposed to an unfamiliar variety of Spanish when compared to lower empathy individuals. This analysis included only Caribbean absolute interorgatives.

Figure ([**fig?**](#ref-fig))- summarizes the posterior distribution of the response accuracy model, illustrating point estimates ±66% and 95% HDIs in graphical form. The log-odds of a correct response at the average proficiency and EQ levels in the control group was -0.27, or approximately 43.29%, (β = -0.27, HDI = [-0.98, 0.40], ROPE = 0.19, MPE = 0.79). The experimental group had a slightly higher median estimate, but it should be noted that the HDI distribution is very large ((β = −0.08, HDI = [−0.87, 0.70], ROPE = 0.20, MPE = 0.57)). This is to say, group by itself is not an adequate predictor of response accuracy.

The effect of proficiency in both groups was similarly positive (control: (β = 0.24, HDI = [−0.06, 0.51], ROPE = 0.15, MPE = 0.95); experimental: (β = 0.23, HDI = [−0.09, 0.50], ROPE = 0.17, MPE = 0.94)), and the effect was practically equivalent between the two groups ((β = -0.01, HDI = [-0.34, 0.31], ROPE = 0.48, MPE = 0.53)). There was no evidence that empathy alone had an effect on response accuracy in either group (control: (β = 0.03, HDI = [−0.24, 0.28], ROPE = 0.57, MPE = 0.58); experimental: (β = 0.00, HDI = [−0.28, 0.29], ROPE = 0.54, MPE = 0.51)).

Although no effect was found for empathy alone, an interaction effect between empathy and proficiency was found in only the experimental group. The relationship is plotted in ([**fig?**](#ref-fig)). In the left panel, response accuracy remains constant at standardized EQ values of -1, 0, and +1 across different proficiency levels in the control group ((β = 0.00, HDI = [−0.32, 0.33], ROPE = 0.48, MPE = 0.51)). In the experimental group, on the other hand, the slope of the proficiency effect increases for higher EQ values ((β = 0.22, HDI = [−0.09, 0.52], ROPE = 0.20, MPE = 0.93)). It should be noted that the effect here is weak, as indicated by the HDI including 0 and the given ROPE value. However, the bulk of the posterior is positive, and the MPE value close to 1 indicates higher certainty in the effect’s direction. In other words, learners with high empathy (black lines) with above average proficiency tended to respond more accurately to Caribbean absolute interrogatives than their proficiency-matched peers only in the experimental group.

To illustrate the interaction of empathy and proficiency, I compared predicted accuracy for learners at low (-1 SD) versus high (+1 SD) empathy across two levels of proficiency (+1 SD and +2 SD). At 1 SD above mean proficiency, learners with higher empathy were approximately 10.41 percentage points (HDI = [-11.05, 31.23]) more accurate than those with low empathy, which corresponds to a 21.79% relative increase in accuracy (HDI = [-26.66, 85.03]). At 2 SD above mean proficiency, learners with higher empathy were approximately 19.66 percentage points (HDI = [-11.30, 51.05]) more accurate than those with low empathy, which corresponds to a 40.48% relative increase in accuracy (HDI = [-33.23, 165.91]).

# 5. Discussion

INTRODUCTION TO DISCUSSION.

## 5.1 RQ1

The first research question, *Do higher-empathy L1 English L2 Spanish speakers make greater use of prosodic cues to determine if a Spanish utterance is an interrogative?*, was a partial replication of Casillas et al. ([2023](#ref-casillas2023using)), and informed by previous research on the relationship between empathy and intonation ([Esteve-Gibert et al., 2016](#ref-esteve2016role), [2020](#ref-esteve2020empathy); [Orrico et al., 2025](#ref-orrico2025individual); [Orrico & D’imperio, 2020](#ref-orrico2020individual)). Unlike previous research on empathy’s impact on L2 phonological acquisition, which primarily focused on pronunciation accuracy (ADD CITATIONS), my study continued the line investigated from Casillas et al. ([2023](#ref-casillas2023using)) by investigating its relation to speech perception. I found that empathy, as measured by the EQ (Baron-Cohen and Wheelwright ([2004](#ref-baron2004empathy))), did modulate response accuracy, specifically for non-Caribbean utterances. Caribbean utterances, on the other hand, did not show an effect for empathy. These findings do not directly contradict Casillas et al. ([2023](#ref-casillas2023using)), but they are slightly different. Where Casillas et al. ([2023](#ref-casillas2023using)) found a relationship between accuracy and empathy for declaratives, they did not find such for absolute interrogatives. This relationship was expected by them based on previous research on empathy and absolute interrogatives in L1 Salerno Italian speakers ([Orrico & D’imperio, 2020](#ref-orrico2020individual)). Furthermore, although the same stimuli were used in the present study and Casillas et al. ([2023](#ref-casillas2023using)), a different distribution of dialects and sentence types were used: where Casillas et al. ([2023](#ref-casillas2023using)) features six non-Caribbean dialects and two Caribbean dialects, and four sentence types (wh- question, absolute interrogatives, broad focus declaratives, and narrow focus declaratives), the current study featured two non-Caribbean and two Caribbean dialects, and only two sentence types (absolute interrogatives and broad focus declaratives). Unlike Casillas et al. ([2023](#ref-casillas2023using)), the current study specifically modeled Caribbean versus non-Caribbean dialects. As such, it is possible that the same findings could have been found in the study this is based on if such an analysis were realized.

Proficiency, as measured by the LexTALE-ESP ([Izura et al., 2014](#ref-izura2014lexical); [Lemhöfer & Broersma, 2012](#ref-lemhofer2012introducing)), was taken into consideration in this cross-sectional design. Surprisingly, proficiency was not found to have an effect for non-Caribbean declaratives, and in fact showed negative trends for non-Caribbean absolute interrogatives and Caribbean declaratives. Caribbean absolute interrogatives, on the other hand, showed a positive relationship with proficiency. The general findings on proficiency directly contradict previous research on the relationship between L2 intonation development and target language proficiency ([Brandl et al., 2020](#ref-brandl2020development); [Casillas et al., 2023](#ref-casillas2023using); [Nibert, 2006](#ref-nibert2006acquisition); [Trimble, 2013a](#ref-trimble2013acquiring)). The validity of the LexTALE as a proxy measure of global proficiency has been questioned ([Puig-Mayenco et al., 2023](#ref-puig2023lextale)). However, these findings remain surprising, as Casillas et al. ([2023](#ref-casillas2023using)) utilized the same measure of proficiency.

A relationship between empathy and proficiency was also considered. It was found that only non-Caribbean absolute interrogative utterances supported such a relationship. In this relationship, individuals with higher empathy were “protected” from the negative-going proficiency trend, such that high-empathy high-proficiency individuals responded more accurately than their lower-empathy proficiency-matched peers. Although an interaction effect was also found by Casillas et al. ([2023](#ref-casillas2023using)), the trend was particularly different: they found that empathy had a “compounding effect” on the relationship between response accuracy and proficiency, such that lower-proficiency individuals with higher empathy showed more accuracy than their proficiency-matched peers only in wh- questions.

As previously noted, the current study directly models Caribbean versus non-Caribbean dialects, unlike Casillas et al. ([2023](#ref-casillas2023using)). The decision to do so was two-fold: (1) Caribbean dialects featured the lowest response accuracy by participants in Casillas et al. ([2023](#ref-casillas2023using)); (2) Caribbean absolute interrogative stimuli differed from non-Caribbean absolute interrogative stimuli in that the former has a final fall, while the latter has a final fall. These differences may have resulted in some of the differences in the results, such as no empathy effect being found for absolute interrogatives in Casillas et al. ([2023](#ref-casillas2023using)), but such an effect being found in the present study. It may be the case that only the Caribbean varieties did not feature such an effect, and that confounded the overall effect.

A possibility as to why no effect for empathy was found for Caribbean utterances is the lack of exposure to those dialects. For example, Orrico and D’imperio ([2020](#ref-orrico2020individual)) investigated L1 Salerno Italian speakers processing Italian sentences with varying nuclear configurations to predict degree of perceived positive bias (i.e., expected positive answer). They found that exposure to non-native dialects affected the way pitch span was mapped onto perceived bias, mediated by empathy. That is, high-empathy listeners were more affected by degree of non-native dialect exposure. As such, it may be that it is not simply enough to have higher empathy to have greater response accuracy when determining if an utterance is an interrogative or not in Spanish based on intonation alone, but exposure to the specific categorical tune-to-meaning association is also necessary.

WRAP IT UP.

# 6. RQ2

Although Casillas et al. ([2023](#ref-casillas2023using)) found a relationship between empathy and intonation processing in L1 English L2 Spanish learners, it remained unclear why having a higher EQ should drive greater accuracy when determining if a Spanish utterance is an interrogative or not. The second research question, *Do higher-empathy L1 English L2 Spanish speakers make greater use of prosodic input from dialects to which they have little or no exposure to extract prosodic features?*, aimed to provide greater understanding of the cognitive relationship between empathy and intonation processing. For the second research question, a between-subjects design compared L1 English L2 Spanish learners who self-reported as unfamiliar with Caribbean or Galician dialects (i.e., dialects with a final fall for absolute interrogatives). Participants in the experimental group were exposed to approximately three total minutes of conversation between three different groups of people. In total, they were exposed to 21 absolute interrogatives, 16 of which had a final fall (approximately 76.1904762%).

Previous research in psychology demonstrated that higher-empathy individuals exposed to short video clips responded more accurately to questions about the videos than their lower-empathy peers, suggesting that higher-empathy individuals retain more environment details while watching social interactions ([Melchers et al., 2017](#ref-melchers2017oxtr)). Furthermore, Orrico and D’imperio ([2020](#ref-orrico2020individual)) found that both high empathy and exposure to a non-native dialect was necessary to cause phonological category interference. As such, the current study aimed to determine if higher-empathy L1 English L2 Spanish learners benefited greater from exposure to an unfamiliar intonation-meaning mapping, specifically Spanish absolute interrogatives with a final fall, when later determining if an utterance is an interrogative or not.

No effect was found for empathy alone, but there was evidence for a weak interaction between empathy and proficiency only for individuals exposed to the short conversations between speakers who produce absolute interrogatives with final falls. That is, individuals at the average proficiency level did not benefit from higher empathy, but individuals with higher empathy and above-average proficiency outperformed on average their proficiency-matched peers. No effect of empathy, alone or interacting with proficiency, was found in the control group, the subjects of which were not exposed to the short conversations and also reported having no prior familiarity with the aforementioned dialects. These results corroborate the idea that higher empathy individuals retain more detail from social interactions ([Melchers et al., 2017](#ref-melchers2017oxtr)), but in second-language processing, empathy specifically works to modulate a proficiency effect when processing novel intonation-meaning mappings. Participants must have a certain proficiency in a language, as well as higher empathy, to benefit from rapid exposure to an unfamiliar intonation-to-meaning mapping.

Within the language classroom, these findings suggest that only a small set of students (those with both higher empathy and higher proficiency) will benefit from brief incidental tasks involving intonation. It is clear that exposure alone is not sufficient for all individuals to acquire target intonation in production ([Herrero et al., 2020](#ref-herrero2020perception)) or perception ([Herrero & Devı́s, 2020](#ref-herrero2020unintentional)), and it is also clear that intentional learning ([Olea, 2019](#ref-olea2019effectiveness); [Sonsaat-Hegelheimer & Levis, 2025](#ref-sonsaat2025intonation)) reaps wider benefits than those found in the present study. As such, the results of this study point towards a greater need of more target-focused training for intonation, especially for dialects that have traditionally less exposure in the language classroom (e.g., Caribbean dialects). Indeed, the skills necessary to successfully complete the task in this study require “sociolinguistic competence” ([Canale, 1987](#ref-canale1987measurement)), an essential component of successful language use.

In sum, the present work contributes to our understanding of how empathy relates to speech processing, specifically for intonation. Current models of L2 phonology, such as the SLM-r ([Flege & Bohn, 2021](#ref-flege2021revised)), fail to address suprasegmental acquisition. Although the LILt model ([Mennen, 2015](#ref-mennen2015beyond)) has presented an opportunity to investigate L2 acquisition of intonation, it has primarily focused on transfer effects. The results of this study, as well as Casillas et al. ([2023](#ref-casillas2023using)), has demonstrated the need for such models to account for individual differences in learner outcomes.

## 6.1 Limitations and Future Directions

While the current study provides evidence that L2 learners with higher empathy and higher proficiency benefit more from incidental learning of a new intonation-meaning mapping than their lower-empathy proficiency-matched peers, a longitudinal study would be required to know if these gains are maintained. Furthermore, it is not clear if the positive effect would continue to be found after repeated exposure. That is, do the lower-empathy individuals eventually “catch up” to their higher-empathy peers with increasing incidental learning moments, or do these individuals require intentional learning?

Another query, one repeated from Casillas et al. ([2023](#ref-casillas2023using)), is how much of an effect does familiarity with a dialect have on response accuracy? Casillas et al. ([2023](#ref-casillas2023using)) found that participants responded more accurately to wh- and absolute interrogatives from their self-reported familiar dialect, but not broad focus or narrow focus declaratives. Furthermore, they did not take proficiency or empathy into account. Based on the current findings, it does seem that empathy interacts with exposure. As previously discussed, Orrico and D’imperio ([2020](#ref-orrico2020individual)) found that empathy interacted with exposure to non-native dialects in L1 category perception. It is important to note that Orrico and D’imperio ([2020](#ref-orrico2020individual)) investigated individuals with exposure to *any* dialect other than their non-native one. Extending this to an L2 context, one of two hypotheses can be proposed: (1) high-empathy L2 learners, as opposed to lower-empathy L2 learners, exposed to various dialects in their L2 have greater pick-up of novel sound-meaning mappings in general; or (2) high-empathy L2 learners, as opposed to lower-empathy L2 learners, exposed to a specific dialect have greater sound-meaning mapping accuracy for that specific dialect. The current study pokes at the second hypothesis, but the small amount of exposure to the target dialect, and the focus on a single nuclear configuration, does not allow for a generalizable answer.

Lastly, the lack of an effect (for non-Caribbean broad declaratives) and negative effect (for non-Caribbean absolute interrogatives and Caribbean declaratives) of proficiency was particularly surprising. It appears unlikely that the LexTALE-ESP ([Izura et al., 2014](#ref-izura2014lexical); [Lemhöfer & Broersma, 2012](#ref-lemhofer2012introducing)), the measure of proficiency used in this study, is inadequately correlated with intonational development, at least when it comes to identifying these sentence types as an interrogative or not, as Casillas et al. ([2023](#ref-casillas2023using)) found a positive effect for proficiency, also measured with the LexTALE-ESP, across the board.

## 6.2 Conclusion

The current study investigated the impact that incidental learning and empathy has on the acquisition of a novel intonation-meaning mapping. I found that individuals with above-average empathy and above-average proficiency were more sensitive to the novel mapping after a brief period of listening to naturalistic conversation compared to their lower-empathy proficiency-matched peers, and also compared to those who did not have the “exposure” period. This study contributes to our understanding of how individual differences impact L2 learner intonation outcomes, as well as providing evidence that incidental learning of intonation may only be beneficial for a subset of L2 learners. These results encourage more research in pragmatic skill as it relates to L2 learner outcomes, and how these findings, and those of Casillas et al. ([2023](#ref-casillas2023using)), relate to findings from L1 research on empathy and language processing.

# 7. References

Armstrong, M. E. (2010). Puerto rican spanish intonation. *Transcription of Intonation of the Spanish Language*, 155–189.

Astruc, L., Mora, E., & Rew, S. (2010). Venezuelan andean spanish intonation. *Transcription of Intonation of the Spanish Language*, 191–226.

Bailey, A. A. (2014). Intonational phonology of cuban spanish: A preliminary model. *Proc. SpeechProsody 2014*, 809–813.

Baron-Cohen, S. (2011). *Zero degrees of empathy: A new theory of human cruelty* (Vol. 30). Penguin uk.

Baron-Cohen, S., & Wheelwright, S. (2004). The empathy quotient: An investigation of adults with asperger syndrome or high functioning autism, and normal sex differences. *Journal of Autism and Developmental Disorders*, *34*, 163–175.

Bedialauneta Txurruka, I. (2023). Perception of spanish declarative questions and statements by L2 spanish speakers. In R. Skarnitzl & J. Volín (Eds.), *Proceedings of the 20th international congress of phonetic sciences* (pp. 2591–2596). Guarant International. https://doi.org/<https://doi.org/10.1017/S0025100323000348>

Bishop, J. (2016). Individual differences in top-down and bottom-up prominence perception. *Proceedings of Speech Prosody*, *2016*, 668–672.

Bishop, J. B., Chong, A. J., & Jun, S.-A. (2015). Individual differences in prosodic strategies to sentence parsing. *ICPhS*.

Boersma, P., & Weenink, D. (2018). *Praat: Doing phonetics by computer [computer program]* (Version 6.4.39). <https://praat.org>

Brandl, A., González, C., & Bustin, A. (2020). The development of intonation in L2 spanish: A perceptual study. In *Hispanic linguistics* (pp. 11–32). John Benjamins Publishing Company.

Bürkner, P.-C. (2017a). Advanced bayesian multilevel modeling with the r package brms. *arXiv Preprint arXiv:1705.11123*.

Bürkner, P.-C. (2017b). Brms: An r package for bayesian multilevel models using stan. *Journal of Statistical Software*, *80*, 1–28.

Canale, M. (1987). The measurement of communicative competence. *Annual Review of Applied Linguistics*, *8*, 67–84.

Carruthers, P. (2009). Mindreading underlies metacognition. *Behavioral and Brain Sciences*, *32*(2), 164–182.

Casillas, J. V., Garrido-Pozú, J. J., Parrish, K., Arroyo, L. F., Rodrı́guez, N., Esposito, R., Chang, I., Gómez, K., Constantin-Dureci, G., Shao, J., et al. (2023). Using intonation to disambiguate meaning: The role of empathy and proficiency in L2 perceptual development. *Applied Psycholinguistics*, *44*(5), 913–940. https://doi.org/<https://doi.org/10.1017/S0142716423000310>

Diehl, J. J., Bennetto, L., Watson, D., Gunlogson, C., & McDonough, J. (2008). Resolving ambiguity: A psycholinguistic approach to understanding prosody processing in high-functioning autism. *Brain and Language*, *106*(2), 144–152.

Estebas-Vilaplana, E. (2014). The evaluation of intonation: Pitch range differences in english and in spanish. In *Evaluation in context* (pp. 179–194). John Benjamins Publishing Company.

Esteve-Gibert, N., Portes, C., Schafer, A., Hemforth, B., & D’Imperio, M. (2016). The role of individual empathic skills on the online processing of intonational meaning. *Architectures and Mechanisms for Language Processing (AMLaP)*.

Esteve-Gibert, N., Schafer, A. J., Hemforth, B., Portes, C., Pozniak, C., & D’Imperio, M. (2020). Empathy influences how listeners interpret intonation and meaning when words are ambiguous. *Memory & Cognition*, *48*, 566–580.

Face, T. L. (2007). The role of intonational cues in the perception of declaratives and absolute interrogatives in castilian spanish. *Journal of Experimental Phonetics*, *16*, 185–225.

Flege, J. E., & Bohn, O.-S. (2021). The revised speech learning model (SLM-r). In R. Wayland (Ed.), *Second language speech learning: Theoretical and empirical progress* (pp. 3–83). Cambridge University Press. <https://doi.org/10.1017/9781108886901.002>

Frith, U., & Frith, C. D. (2003). Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *358*(1431), 459–473.

Gabriel, C., Feldhausen, I., Pešková, A., Colantoni, L., Lee, S.-A., Arana, V., & Labastı́a, L. (2010). Argentinian spanish intonation. *Transcription of Intonation of the Spanish Language*, 285–317.

Gelman, A., Simpson, D., & Betancourt, M. (2017). The prior can often only be understood in the context of the likelihood. *Entropy*, *19*(10), 555.

George, A. (2024). L2 intonation perception in learners of spanish. *Perspectives on Input, Evidence, and Exposure in Language Acquisition: Studies in Honour of Susanne E. Carroll*, *69*, 144. https://doi.org/<https://doi.org/10.1075/lald.69.06geo>

Herrero, C., & Devı́s, E. (2020). Unintentional impolite intonation in L2 spanish requests produced by chinese workers living in madrid. *Proceedings of the 10th International Conference on Speech Prosody*, 848–852.

Herrero, C., Planelles, M., & Alhmoud, Z. (2020). Perception of L2 spanish polite requests and impolite commands by chinese migrant workers living in spain. *Lengua y Migración*, *12*(2), 65–85.

Horst, M., Cobb, T., & Meara, P. (1998). *Beyond a clockwork orange: Acquiring second language vocabulary through reading*.

Hulstijn, J. H. (2003). Incidental and intentional learning. *The Handbook of Second Language Acquisition*, 349–381.

Izura, C., Cuetos, F., & Brysbaert, M. (2014). Lexical test for advanced learners of spanish. *Psicológica*.

Jenkins, J. (2004). 5. Research in teaching pronunciation and intonation. *Annual Review of Applied Linguistics*, *24*, 109–125.

Kabak, B. (2016). Refin (d) ing turkish stress as a multifaceted phenomenon. *Paper Delivered at Second Conference on Central Asian Languages and Linguistics (ConCALL-2)*.

Kruschke, J. K. (2018). Rejecting or accepting parameter values in bayesian estimation. *Advances in Methods and Practices in Psychological Science*, *1*(2), 270–280.

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

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of english. *Behavior Research Methods*, *44*(2), 325–343.

Melchers, M., Montag, C., Markett, S., Niazy, N., Groß-Bölting, J., Zimmermann, J., & Reuter, M. (2017). The OXTR gene, implicit learning and social processing: Does empathy evolve from perceptual skills for details? *Behavioural Brain Research*, *329*, 35–40.

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. https://doi.org/<https://doi.org/10.1007/978-3-662-45168-7>

Nibert, H. J. (2006). The acquisition of the phrase accent by beginning adult learners of spanish as a second language. *Selected Proceedings of the 2nd Conference on Laboratory Approaches to Spanish Phonetics and Phonology*, 131–148.

Olea, M. S. A. U. (2019). The effectiveness of the instruction of the intonation of english implications to L1 spanish trainee teachers. *Logos: Revista de Lingüı́stica, Filosofı́a y Literatura*, *29*(2), 399–414.

Orrico, R., & D’imperio, M. (2020). Individual empathy levels affect gradual intonation-meaning mapping: The case of biased questions in salerno italian. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *11*(1), 1–39.

Orrico, R., Gryllia, S., Kim, J., & Arvaniti, A. (2025). Individual variability and the h\*~ l+ h\* contrast in english. *Language and Cognition*, *17*, e9.

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*(2), 331–353.

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*, 195–203.

Pérez Castillejo, S., & Fuente Iglesias, M. de la. (2024). Basic intonation patterns of galician spanish. *Languages*, *9*(2), 57.

Pierrehumbert, J. B. (1980). *The phonology and phonetics of english intonation* [PhD thesis]. Massachusetts Institute of Technology.

Puig-Mayenco, E., Chaouch-Orozco, A., Liu, H., & Martı́n-Villena, F. (2023). The LexTALE as a measure of L2 global proficiency: A cautionary tale based on a partial replication of lemhöfer and broersma (2012). *Linguistic Approaches to Bilingualism*, *13*(3), 299–314.

Rao, R. (2019). *Key issues in the teaching of spanish pronunciation*. Routledge London, United Kingdom.

Sonsaat-Hegelheimer, S., & Levis, J. (2025). Is intonation learnable in the classroom? Evidence from turkish learners of english. *Language Teaching Research*, 13621688241312502.

Trimble, J. C. (2013a). *Acquiring variable L2 spanish intonation in a study abroad context* [PhD thesis, University of Minnesota]. https://doi.org/<https://hdl.handle.net/11299/175354>

Trimble, J. C. (2013b). Perceiving intonational cues in a foreign language: Perception of sentence type in two dialects of spanish. *Selected Proceedings of the 15th Hispanic Linguistics Symposium*, 78–92.

Ward, G., & Hirschberg, J. (1988). Intonation and propositional attitude: The pragmatics of l\*+ HLH%. *Proceedings of the Fifth Eastern States Conference on Linguistics*, 512–522.

Willis, E. (2010). Dominican spanish intonation. *Transcription of Intonation of the Spanish Language*, 123–153.

# 8. Appendix A

# 9. This Section Is an Appendix

# 10. Appendix B

# 11. Another Appendix