Production of final boundary tones in declarative utterances by English-speaking learners of Spanish

Germán Zárate-Sández
Department of Spanish, Western Michigan University, USA
german.zarate-sandez@wmich.edu

Abstract
Given differences in the intonational contours between Spanish and English in unmarked declarative utterances, the present study sought to determine how English-speaking learners of Spanish produce final boundary tones and how proficiency interacts with their performance. Participants were American college students learning Spanish at three proficiency levels: intermediate (n=17), high (n=20), and very high (n=18). In addition, groups of monolingual speakers of Spanish (n=17), English (n=17), and balanced English-Spanish bilinguals (n=16) were included for comparison. A storytelling task was used to elicit quasispontaneous speech. Syntactic, auditory, and acoustic criteria were employed to select eight analyzable utterances for each participant. End pitch values were automatically extracted, converted to ERB units, and averaged to produce one aggregate score per participant. Results revealed monolingual speakers of English had the highest boundary tone values, while speakers of Spanish had the lowest, which is consistent with the existence of high rising terminal contours (or uptalk) in American English. Scores for Spanish learners and bilingual speakers fell in the middle of the pitch scaling range. Intermediate learners clustered with English monolinguals, while scores for the high and very high proficiency learners were not statistically different among themselves or from other groups. Findings showed that, despite differences between English and Spanish, learners can reach native-like performance at a high proficiency level.

Index Terms: boundary tone, high rising terminal, intonation, second language, Spanish, uptalk

1. Introduction
The field of second language (L2) phonology has looked largely at segmental phenomena while issues related to the use and acquisition of suprasegmental features remain comparatively underexplored. As far as the L2 under investigation is concerned, most studies have dealt with prosody of English (e.g., [1], [2]) and Asian languages (e.g., [3], [4]), while the body of knowledge about the acquisition of Spanish—the most commonly studied L2 in the U.S. and popular in Europe as well—is still small. As a means to address these gaps in the current literature, the present study investigated Spanish learners’ production of utterance-final intonation in declarative utterances.

English speakers may face challenges when learning Spanish intonation because of differences between the two languages in unmarked declarative utterances. In terms of prenuclear peak alignment, for example, Spanish peaks tend to appear on the posttonic syllable whereas English peaks are usually aligned with accented syllables ([5]–[7]). As for boundary tones, Spanish normally produces a falling pitch at the end of a declarative utterance ([8]), while it is not infrequent for speakers of some varieties of English to produce a rising contour in the same type of utterance. This phenomenon has been commonly referred to as uptalk (see [9] for a comprehensive description of the subject) or, more technically, high rising terminal (HRT). These rises have been traditionally observed in New Zealand and Australian English ([10]–[12]) and appear to be increasingly common in American English ([7], [13]) across genders and dialectal areas ([14]–[16]). The stylistic contexts for a HRT are normally narratives, invitations to acknowledgement from the listener, transaction openers, and answers to wh-questions ([7]). The handful of studies that addressed issues of acquisition and use of intonation in a second language have found that English speakers may indeed encounter difficulties perceiving or producing Spanish intonation. Results from [17] and [18] showed that learners of different proficiency levels perceived and interpreted the presence of the disambiguating intermediate tone (H–), though true beginners did so only in very limited cases. In [19], two out or four highly proficient learners studying Spanish in León, Spain, were found to decrease their production of rises in declaratives from approximately 45% at pretest to 25% at posttest.

Determining how a specific linguistic area develops as a function of overall proficiency in the L2 is a major interest in second language acquisition (SLA) research. For the case of intonation, some studies have identified improvements (that is, intonational patterns approach the standard patterns of native speakers of the L2) as learners receive more input in the L2 and their general competence improves ([4], [17]–[20]), others have found no effect for proficiency ([21]), and still others suggest that higher proficiency relates to lower performance in some areas of intonation ([22], [23]). Furthermore, recent research has shown that a very high proficiency level in the L2—learned during adulthood, as in [24], or as a result of balanced bilingualism, as in [25] and [26]—can reconfigure the intonational system of the native language (or languages, in the case of balanced bilinguals), with the result that the intonational systems in neither language function exactly as those of monolingual speakers of the languages in question.

In view of the remaining issues in our understanding of the acquisition of L2 Spanish intonation, the following two research questions were posed in this study:

- What kind of intonational contour do English-speaking learners of Spanish produce in unmarked declarative Spanish utterances?
- Do learners’ intonational patterns in unmarked declarative Spanish utterances change as a function of proficiency?
2. Methodology

2.1. Participants

Initial recruitment was conducted from third- and sixth-semester Spanish classes at two universities in the U.S. In addition, students pursuing a Master’s or Ph.D. in Spanish at one of the universities were also recruited. After initial recruitment, participants’ proficiency was measured with the Spanish Elicited Imitation Task (EIT) test ([27]), believed to be a reliable and valid tool to measure L2 oral proficiency (see [28] for discussion). For the sake of space, details about administration and scoring are omitted here, but followed the same procedures described in [27] and [28]. Final proficiency was hence operationalized according to two criteria: previous experience and scores in the EIT. From possible scores ranging from 0 to 120, means for the three groups were as follows: 57.96 ($SD = 12.24$) for the third-semester group, 93.33 ($SD = 11.65$) for the sixth-semester group, and 116.40 ($SD = 3.78$) for the group of graduate students. In order to maintain homogeneity in each group, outliers (two standard deviations above or below the group mean) were eliminated from the study. For this and other exclusion criteria (e.g., not completing all parts of the study, being highly proficient in an additional language, or having extended stays in Spanish-speaking countries), 15 participants were excluded from the original pool, resulting in 55 learners grouped into three levels: intermediate proficiency (IP, $n = 17$), comprised of third-semester Spanish students with EIT scores in the 42–82 range; high proficiency (HP, $n = 20$), comprised of sixth-semester students with EIT scores in the 83–107 range, and very high proficiency (VHP, $n = 18$), comprised of graduate students with EIT scores in the 109–120 range.

In addition, L2 speakers were compared with three groups: Spanish native speakers (SNS, $n = 17$), English (monolingual) native speakers (ENS, $n = 17$), and English-Spanish bilingual speakers (BS, $n = 16$). SNS speakers represented the various Spanish dialects to which learners in this study had likely been exposed. ENS participants were recruited from undergraduate classes at the same university. They were all monolingual native speakers of American English. The BS group was composed of heritage speakers, that is, early bilinguals who were raised in families that speak a minority language. This group was included to expand the comparison beyond typical monolingual norms ([29]).

2.2. Elicitation task and procedure

In order to elicit quasispontaneous speech, participants were asked to tell the story in the book Frog. where are you? ([30]), which consists of a sequence of 24 wordless pictures. This commonly-used story in language research elicits naturally-occurring language under minimal instructions and has been used to examine a broad scope of issues in language acquisition (e.g., [31], [32]), including a study on intonation among Turkish-German bilinguals ([33]). Participants in the ENS group completed the task in English, while all other groups did so in Spanish.

Participants were asked to tell the story to the researcher, who was supposedly unfamiliar with it. Prior to recording, participants were given a few minutes to peruse the entire book. A list with key vocabulary (e.g., rana ‘frog’, ciervo ‘deer’, abeja ‘bee’) was also provided. Participants were encouraged to tell a coherent story but were also allowed to skip pages if they did not understand certain scenes or thought they lacked sufficient vocabulary to describe them. The researcher asked questions to elicit further language in cases of extended hesitations. Five participants were asked to stop after speaking for five minutes, as this was considered sufficient time. Recordings were made in a sound-proof room, with a high-sensitivity, head-mounted microphone attached to a personal computer and saved individually as WAV files, at a sampling rate of 44 kHz with the software Audacity.

2.3. Data analysis

Ninety seconds were extracted from each recording and used for analysis, beginning after the first 60 seconds, when participants usually felt more relaxed and comfortable with the task. The selection of utterances ultimately used in the analysis (henceforth referred to as analyzable utterances) was carried out in three phases. The goal of the data-trimming techniques was to obtain eight analyzable utterances per participant. These utterances varied in their nuclear and prenuclear composition, as they were used for a larger project on L2 intonation (see [34]), including prenuclear pitch accents. This paper reports on the analysis of utterance-final pitch only.

First, the 90-second speech sample was transcribed orthographically in order to select the sentences with the following syntactic and morphological features: (1) the sentence was syntactically complete, that is, contained a subject (overt or tacit) and a predicate, and any obligatory arguments within the predicate, (2) it was syntactically simple (no subordination or coordination), and (3) it was a declarative sentence. Second, utterances were examined with Praat to check for general acoustic and auditory properties and excluded from analysis if: (1) any part was emphasized or narrowly focused, (2) the utterance contained filled or unfilled pauses or hesitations, (3) was produced in a very soft voice and/or contained excessive creaky voice in the critical areas for analysis, (4) was uttered unnaturally slow or fast, and (5) the last word was not paroxytone. In the case of 18 participants, the researcher had to go back to the original recordings (beyond the 90-second sample) to find additional utterances that met the above criteria.

After the first two stages of data-trimming criteria, each participant averaged 13 utterances. Unlike a sentence-reading task, a quasispontaneous task frequently yields interruptions in the pitch tracks caused by non-sonorant segments, which led to methodological compromises when selecting the final eight analyzable utterances for each participant. To minimize any potential bias in the selection of these utterances, a trained phonetician naïve to the specific goals of the study was asked to examine acoustically the remaining utterances and select the eight that had the least amount of disruptions and detection errors in the pitch track. In total, 823 utterances were analyzed (131 for IP group, 156 for HP, 144 for VHP, 136 for SNS, 132 for ENS, and 124 for BS). The original objective was to obtain and analyze 840 utterances (105 participants x 8 utterances), but data for nine participants fell short of this goal, even after examination of their entire recordings.

Final boundary tone in the analyzable utterances was defined as the last non-spurious $f_0$ point in the pitch track generated by Praat (see [35] and [36] for similar procedures). First, an interval tier was created to delimit the analysis to a five-minute proof room, with a high-sensitivity, head-mounted microphone attached to a personal computer and saved individually as WAV files, at a sampling rate of 44 kHz with the software Audacity. Segmentiation was done manually via simultaneous inspection of spectrogram and waveform. Values for end pitch were extracted automatically.
using a script with instructions to search for end of pitch in the interval and to return a value in Hz. Extracted $f_0$ values were converted to Equivalent Rectangular Bandwidth (ERB) units to normalize for differences among speakers, in particular between males and females ([25], [35]). Normalized values were averaged so that each participant was represented in the statistical analysis by one aggregate score.

3. Results

This section presents results for final boundary tone in the storytelling task for the six groups under examination. Descriptive statistics are followed by inferential analyses using a one-way analysis of variance (ANOVA) and post-hoc comparisons.

Table 1: Descriptive statistics for final boundary tone production in storytelling task.

<table>
<thead>
<tr>
<th>Group</th>
<th>n</th>
<th>Mean</th>
<th>SD</th>
<th>Min.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>IP</td>
<td>17</td>
<td>3.74</td>
<td>0.41</td>
<td>2.95</td>
<td>4.47</td>
</tr>
<tr>
<td>HP</td>
<td>20</td>
<td>3.61</td>
<td>0.38</td>
<td>2.98</td>
<td>4.30</td>
</tr>
<tr>
<td>VHP</td>
<td>18</td>
<td>3.60</td>
<td>0.47</td>
<td>2.80</td>
<td>4.30</td>
</tr>
<tr>
<td>SNS</td>
<td>17</td>
<td>3.31</td>
<td>0.43</td>
<td>2.56</td>
<td>4.30</td>
</tr>
<tr>
<td>ENS</td>
<td>17</td>
<td>3.84</td>
<td>0.39</td>
<td>3.20</td>
<td>4.60</td>
</tr>
<tr>
<td>BS</td>
<td>16</td>
<td>3.55</td>
<td>0.40</td>
<td>2.75</td>
<td>4.18</td>
</tr>
</tbody>
</table>

Note. Means represent ERB Units.

IP = Intermediate Proficiency, HP = High Proficiency, VHP = Very High Proficiency, SNS = Spanish Native Speakers, ENS = English Native Speakers, BS = Bilingual Speakers.

Results of descriptive statistics show ENS and SNS groups at different extremes, with the former obtaining the highest and the latter the lowest mean for final boundary pitch scaling. The other four groups appeared in the middle of the monolingual groups.

Table 2: Post-hoc results for final boundary tone height in storytelling task.

<table>
<thead>
<tr>
<th>Group</th>
<th>n</th>
<th>Subset</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNS</td>
<td>17</td>
<td>3.31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BS</td>
<td>16</td>
<td>3.55</td>
<td>3.55</td>
<td></td>
</tr>
<tr>
<td>VHP</td>
<td>18</td>
<td>3.60</td>
<td>3.60</td>
<td></td>
</tr>
<tr>
<td>HP</td>
<td>20</td>
<td>3.61</td>
<td>3.61</td>
<td></td>
</tr>
<tr>
<td>IP</td>
<td>17</td>
<td>3.74</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENS</td>
<td>17</td>
<td>3.84</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Sig. ($p$) = .25 .32

Note. Means represent ERB Units.

Subsets indicate groups of homogenous means ($p < .05$). Since sample sizes are unequal, harmonic mean is used.

In order to test differences between groups statistically, a one-way ANOVA was run. Lavene's test showed that group variances could be assumed to be equal, $F(5, 99) = .243$, $p = .942$. The one-way ANOVA revealed that group means were statistically different: $F(5, 99) = 3.248$, $p < .01$. Table 2 summarizes post-hoc comparisons (Tukey’s).

Results of the post-hoc test revealed that participants can be divided into two statistically different groups. Final boundary tone values for native monolingual speakers of Spanish are statistically lower than values for native monolingual speakers of English and learners with intermediate proficiency; these last two groups, in turn, produced statistically equal values. The remaining three groups (HP, VHP, and BS) are not statistically different between themselves or from any other group.

Finally, a Pearson correlation test between scores on the EIT and ERB units was performed for the three learner groups ($n = 55$) in order to further address the research question regarding the role of proficiency in learners’ intonational patterns. This test treats proficiency as a continuous variable and therefore complements the results of the by-group analysis of variance. The result of the test barely reached significance ($r = -.268$, $p = .048$), suggesting a moderate inverse relationship between proficiency level and final boundary tone height: as the former increases, the latter decreases. This relationship is presented graphically in Figure 1.

![Figure 1: Scatterplot for EIT scores (X axis) and ERB Units (Y axis) for IP, HP, and VHP groups (n = 55), with regression line.](image)

4. Discussion and conclusion

The present study set out to determine what pattern of final intonation English-speaking learners produce in unmarked declarative Spanish sentences and how proficiency affects this pattern. As far as the first goal, scores for the three groups of learners fell between scores from both groups of monolingual speakers, who in turn produced final pitch scaling at two extremes: Spanish native speakers produced the lowest boundary tone, while English native speakers produced the highest. Even though this study posed two broad research questions and did not set out to test the production of high rising terminals, these results are consistent with the presence of these tones. Furthermore, the narrative nature of the task matched a typical context where uptalk normally occurs ([7]). A follow-up analysis would need to examine the frequency of HRTs as a percentage of total boundary tones produced (e.g., [15]). During analysis of utterances in this study, however, clear cases of
uptalk did appear in the groups of L2 learners. Figure 2 presents an example of a HRT from a male speaker in the IP group, which was the only group that statistically patterned with the group of English native speakers, while Figure 3 contains an example of a HRT produced by a female speaker in the HP group.

![Figure 2: Example of utterance with HRT, produced by male speaker in IP group. Utterance: Cayó de la ventana. ‘(He) fell from the window.’](image)

![Figure 3: Example of utterance with HRT, produced by female speaker in HP group. Utterance: Hay un ave. ‘There’s a bird.’](image)

The finding that intermediate-proficiency speakers produced tones at the same level of English monolingual speakers corroborates previous studies that found a strong effect of transfer from English to Spanish in the production of final tones. [19], for example, also observed a predominance for final rises among participants who were beginning a study abroad program in Spain. That being said, the other groups of L2 learners in the current study produced final pitch height at the same level as Spanish native and bilingual speakers did. In fact, as shown in post-hoc analyses (Table 2), boundary tone values for HP, VHP, and BS participants did not differ statistically from either the ENS or the SNS groups. However, comparisons also showed that values for ENS and SNS groups did differ statistically. Results from a correlation analysis also showed a modest but significant relationship between proficiency scores and production of final boundary tones: as proficiency increases, final tones are produced progressively lower. Taken together, these results suggest that end pitch height might prove challenging only at initial stages of acquisition. Learners’ production seems to attune to monolingual norms beyond an intermediate proficiency level. The perceptual salience of utterance-final tones and the frequency of declarative utterances in the input might facilitate acquisition of this tone in the L2, even if the first language produces different contours under certain circumstances.

The results in this study, therefore, revealed two important facts: (1) advanced L2 learners produced intonational patterns equivalent to those of Spanish-English bilinguals (heritage speakers) and (2) these patterns appeared in between values of English and Spanish monolingual speakers. The first observation has implications for issues of development and ultimate attainment in L2 acquisition. There has been a growing impetus in the field of SLA to drop the monolingual native norm as a benchmark for learning (e.g., [29]) and adopt, instead, bilingualism (and early bilinguals) as the yardstick against which L2 development and use are measured. That being said, the data in this study do not rule out the possibility that production for L2 learners could change in the future to match monolingual native standards. However, the relatively stable and exceptionally high proficiency level of the most advanced learners in this study suggests that their intonational patterns should not change substantially beyond their current state.

The second observation provides strong support for the ‘merger’ category predicted by Flege’s Speech Learning Model (SLM, [37]), which posits that certain phonological aspects in bilingual speech are composed of categories taken from both languages, thus creating a sort of single, mixed resolution that falls in the middle of both languages. This is what the current study found in the tonal event under investigation. Very advanced L2 speakers and early bilinguals have also been found to behave differently from monolingual speakers in areas outside phonology (e.g., vocabulary knowledge and reading fluency in [38]). Though the evidence is scant for prosody, similar results have been obtained, for instance, by [1] in a study of pitch range among advanced German-speaking learners of English. Other studies, however, have tested this prediction and the merger effect was not found ([39]). The current exploratory study did not set out to test the merger prediction but the results still showed that English-speaking learners with a high proficiency level in Spanish can achieve bilingual- and native-like production of Spanish intonation.

5. Acknowledgements

I would like to thank Alfonso Morales-Front, Lee M. Penyak, Cristina Sanz, and Joan Carles Mora for their suggestions during the development of this study. The feedback from three anonymous reviewers also greatly improved the quality of this paper.

6. References


[34] G. Zárate-Sánchez, “Perception and production of intonation among English-Spanish bilingual speakers at different proficiency levels,” Georgetown University, 2015.


