



Task-effects in the L2 perception and production of English sentence types by L1 Spanish speakers

L. Colantoni, G. Klassen, M. Patience, M. Radu, O. Tararova

University of Toronto

Abstract

In the present paper, we investigated the acquisition of three English sentence types, statements (S), absolute yes-no questions (AQ), and declarative questions (DQ), by L1 Spanish-L2 English adult speakers. Learners of English must acquire not only the syntactic and intonational cues that distinguish AQs from DQs, but also the pragmatic distinction between the two. Participants completed two production and three perception tasks involving increasing levels of access to contextual meaning, in order to determine learners' ability to combine both syntactic and intonational cues in the correct pragmatic context. Results indicate that second language (L2) English speakers demonstrate difficulty acquiring the distinction between AQs and DQs. Learners incorrectly selected AQs in contexts where DQs should have been used in both perception and production, although error rates were much higher in production. Evidence of cross-linguistic influence (CLI) in the use of prosodic cues was observed in the production tasks, where learners tended to produce a higher pitch accent in interrogatives than statements, a characteristic of Spanish interrogatives. The results of this study provide further support for the claim that more open-ended tasks increase the degree of CLI.

Index Terms: intonation, sentence type, second language acquisition, Spanish, English, perception, production.

1. INTRODUCTION

The production and perception of contextually-appropriate utterances is a crucial aspect of L2 acquisition. However, most studies that have investigated the L2 acquisition of contextually-appropriate utterances have used controlled tasks, such as reading or imitation, to elicit data. Recent studies suggest that the exclusive use of controlled tasks to assess the degree of CLI may be problematic, since task effects in the perception and production of intonation have been reported for monolingual and bilingual speakers. Factors known to affect performance include the elicitation method (e.g., read vs. spontaneous speech) [1,2], the presence of linguistic vs. purely auditory information [3], and more generally, the availability of contextual information [4]. In particular, access to contextual information in perception and production has been reported to increase the degree of CLI [4]. These findings have methodological consequences, as well as implications for the teaching and assessment of an L2, yet our understanding of how the above factors influence the L2 acquisition of intonation is limited.

Our overall research objective is to investigate to what extent the perception and production of a given intonational contour may vary across tasks in L2 learners. To achieve this goal, we began by analysing the acquisition of sentence type prosody, given that there is a relative consensus that one of the primary uses of intonation is to mark sentence type [5]. In addition, while intonation can be the only cue used to mark sentence type, as in English declarative questions [DQs] (e.g., *John went home?*) vs. statements [Ss] (e.g., *John went home.*), it can also be one of multiple cues, as in English absolute yes-

no questions (AQs), where syntactic cues are also present (e.g., *Did John go home?*). To acquire the English sentence types, L2 speakers must learn both the intonational and syntactic cues, as well as the pragmatic context in which each sentence type can be used. This combination of primary and redundant cues can be difficult for speakers whose L1 may share some, but not all, characteristics. Consider Spanish and English, which have similar intonational profiles, yet a different syntactic order for questions and statements. For example, both English and Spanish interrogatives have rising boundary tones at the end of the utterance, but Spanish also cues interrogativity by means of a higher first pitch accent (PA) [6,7], whereas in English, the initial PA should not vary across sentence types [8]. Moreover, while English DQs (as opposed to AQs) do not require inversion and are context-specific (e.g., to express incredulity), the word order in Spanish AQs and DQs is identical. Thus, an L1 Spanish-L2 English speaker needs to acquire (i) a different syntactic order in AQs; (ii) the contextual restrictions of DQs (which are syntactically similar to AQs in Spanish); (iii) the phonetic differences in the realization of initial PAs in English as opposed to Spanish. In the present paper, we aim to answer the following questions:

(1) Do L1 Spanish-L2 English speakers differ from native English speakers in their ability to use and comprehend Ss, DQs, and AQs?

(2) Are differences between groups larger in tasks involving a greater degree of access to contextual meaning?

2. PREDICTIONS

Regarding the perception of English sentence types, learners are expected to demonstrate greater difficulty identifying DQs as questions than AQs, given that AQs are marked as interrogatives by both syntactic and intonational cues, as opposed to just intonational cues. Moreover, given the difficulty of acquiring native-like comprehension of form-meaning mappings reported in previous studies [4], we expect that learners will experience difficulty selecting the target context in which DQs should be used.

With respect to production, previous findings on the L2 acquisition of intonation have shown that the L2 production of intonation is often influenced by the learner's L1, especially in post-puberty learners [9,10]. Thus, we predict that in the production tasks, L2 Spanish speakers will produce a higher PA than controls. This difference is expected to be largest in the task involving the greatest amount of access to contextual meaning, which has been shown to facilitate CLI [4]. Participants are also expected to over-generalize the use of AQs in place of DQs, based on the fact that inverted questions in English are the most common (and the default) form and consist of salient syntactic markers (i.e., inversion).

3. METHODS

3.1. Participants

Table 1 illustrates the characteristics of the L1 Canadian English controls and L1 Spanish-L2 English learners.

Participants were all advanced to near-native speakers of English with several years of residency in an English-speaking country and with a reported use of English in their daily lives. Participants were all tested in Canada.

Table 1. *Participant profiles. N=number; Age; AoA= age of onset of acquisition; LoR = length of residency; LoE= length of experience (in years).*

Language	N	Age	AoA	LoR	LoE
English	7	25.0	n/a	n/a	n/a
Spanish	7	35.1	11.0	8.0	24.1

3.2. Tasks and stimuli

3.2.1. Perception

Participants performed three perception tasks. In the first (no context; henceforth ‘NC’) and second tasks (little context; henceforth ‘LC’), participants heard utterances and indicated whether they believed the utterance to be a statement, a question, or an exclamation. The utterances for the NC task were low-pass filtered (to remove segmental information, but maintain the pitch profile), whereas the utterances for the LC task were unmodified. A total of 30 target stimuli were included for both tasks (10 per sentence type: S, DQ, AQ), as well as 25 distractors. In the third task (context; henceforth ‘C’), participants heard a short scenario followed by three utterances (an S, a DQ, and an AQ), and had to select which of the three was the most logical continuation of the scenario. A total of 18 stimuli were included (six per sentence type; no distractors). Stimuli for all tasks were recorded by a female Canadian English speaker.

3.2.2. Production

Participants performed two production tasks. In the first task (NC), they were asked to repeat using the same prosody sentences that were presented to them aurally. A total of 30 target sentences were included (10 per sentence type: S, DQ, AQ), as well as 30 distractors. In the second task (C), participants heard a short scenario, and were required to produce a sentence that logically followed the scenario, either an S, a DQ, or an AQ. There were six scenarios for each sentence type, and no distractors. Stimuli for both tasks were recorded by the same speaker who recorded the perception stimuli.

3.3. Data analysis

For the perception data, the percentage correct was calculated, as were reaction times (reaction times are not reported here). Production data was extracted and annotated in Praat [11]. For each utterance, the first PA and the final PA + boundary tone (= nuclear contour) were marked and analysed acoustically. The F0 excursion over the stressed syllable (difference between max and min pitch) and the slope (F0 change/duration of the stressed syllable) were calculated and reported in ERB.

4. RESULTS

4.1. Perception

The proportion of errors found in the three tasks was low for both groups (see Figure 1), although clearly higher for the L1 Spanish group. To test our hypotheses, we ran a generalized mixed effects model with subject as a random effect and language, task type (NC, LC, C) and sentence type (S, AQ, DQ) as fixed effects. Results revealed a significant effect for language ($F_{(1,12)}=20.7$; $p=.001$), and sentence type ($F_{(2,12)}=6.3$; $p=.01$). The variable ‘task type’ and the random effect were not significant. As expected, significantly fewer errors were found for the control group than for the L1 Spanish speaking group ($\beta=-1.59$; $t=-4.5$; $p=.001$). As for sentence type, a significantly higher number of errors was found with DQs ($\beta=0.71$; $t=2.1$; $p=.05$), followed by Ss and AQs, respectively. In summary, of the three independent variables in our model, ‘language’ had the strongest effect (i.e., errors were mostly found among L2 speakers) followed by ‘sentence type’ (the largest proportion of errors was found with DQs across tasks). This is illustrated in Figure 1 (bottom panel): the total percentage of errors is similar across tasks (albeit slightly lower in LC) and the majority of errors were found for DQs.

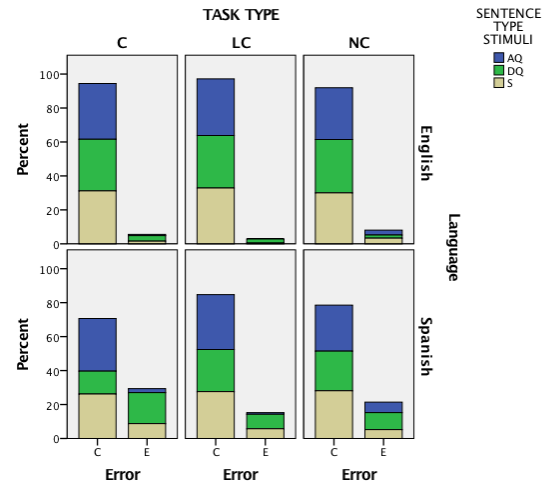


Figure 1. *Percentage of correct (C) and incorrect (E) responses in each task (C, LC, NC) organized by sentence type (AQ, DQ, S); top panel: English; bottom panel: Spanish.*

4.2. Production

In this section, we present the production results for the F0 excursion of the PA and the nuclear contours in the NC and C tasks. Whereas the same number of target responses for each language group was obtained in the NC task (given that participants were repeating a stimulus presented to them aurally), a smaller percentage of target responses was obtained in the C task for the L1 Spanish group, since in this task, participants had to produce the sentence that they considered most appropriate to the context. In contexts that prompted a DQ, 83% were non-target-like responses, produced primarily as AQs. Given the low number of DQ responses in the C task, two sets of statistics were calculated; the first set included only the correct responses, and the second contained all responses consisting of a question structure (i.e., DQs and

AQs, but not Ss). Since no differences were found between the statistics calculated over the total number responses and the statistics calculated only over the target-like responses, we will only present here the results of the statistical analysis on target-like responses.

Figure 2 displays the results of the F0 excursion in the first PA in the NC and C tasks. Results of a mixed effects model with the F0 excursion in the PA as the dependent variable; language, type of task (NC, C) and sentence type (S, AQ, DQ) as fixed effects and speaker as a random effect, revealed a significant effect only for type of task ($F_{(1,525)}=18.93$; $p=.000$), with a smaller F0 excursion in the first PA in the C task than in the NC task ($\beta=-0.26$; $t=-4.3$; $p=.000$). The variable ‘sentence type’ approached significance ($F_{(2,525)}=2.6$; $p=.07$), with DQs displaying a larger F0 excursion than the other two sentence types ($\beta=0.16$; $t=2.27$; $p=.02$), which was expected for our experimental group, but not for our control group.

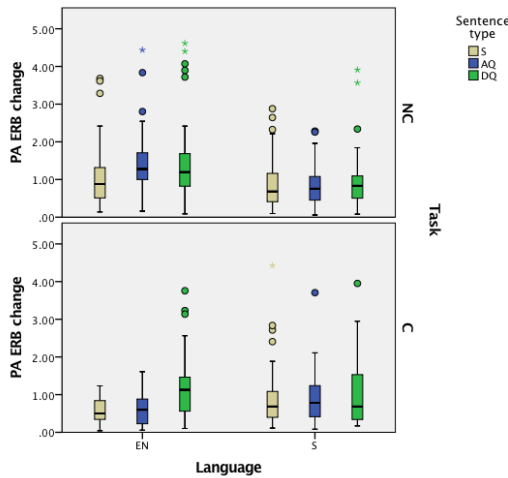


Figure 2. *F0 change in the first pitch accent (in ERB) organized by task, language, and sentence type.*

Differences in the slope of the PA were also analysed. A mixed effects model (with the same fixed and random effects as the model above) was calculated and again, only the variable ‘task’ turned out to be significant ($F_{(1,585)}=48.04$; $p=.000$), with a less steep slope observed in the C task than in the NC task.

Although differences between languages were not expected in the realization of nuclear contours, we inspected their realizations to determine whether there were differences between tasks and groups, in addition to the expected differences between Ss and questions. As illustrated in Figure 3, the main patterns observed in the realization of nuclear contours were not different from those obtained for PAs (i.e., smaller F0 excursions in the C than in the NC tasks for both languages). Statistics were calculated first on the F0 excursion over the nuclear contour, with the F0 excursion in the nuclear contour as the dependent variable; language, type of task and sentence type as fixed effects, and speaker as a random effect. Results revealed no significant effect for language and a significant effect for task type ($F_{(1,570)}=42.2$; $p=.000$) and sentence type $F_{(2,570)}=29.44$; $p=.000$), as well as for the random effect ($z=17.40$; $p=.000$). The F0 excursion was significantly larger in DQs ($\beta=0.95$; $t=6.8$; $p=.000$) and in AQs ($\beta=0.86$; $t=6.5$; $p=.000$) than in Ss. As was the case with PAs, the F0 excursion was found to be smaller in the C ($\beta=-0.75$;

$t=-6.4$; $p=.000$) than in the NC task for both question types. The analysis of the F0 slope revealed similar results to the analysis of the F0 excursion. Once again the variables ‘task’ ($F_{(1,570)}=65.2$; $p=.000$) and ‘sentence type’ ($F_{(2,570)}=31.09$; $p=.000$) turned out to be significant, whereas no significant differences were found between our two language groups. It is worth pointing out that a large degree of variability was observed in English statements in the NC task, which was indicative of the use of rising contours by some of our participants.

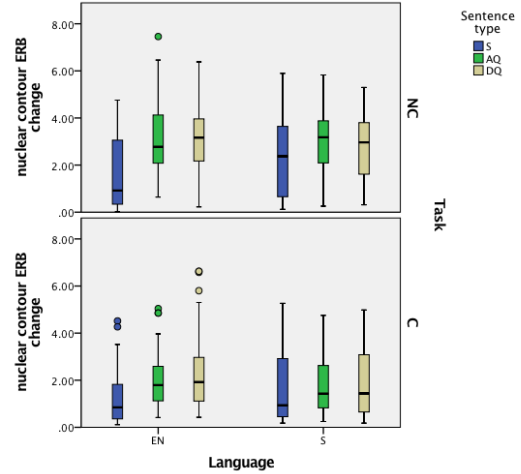


Figure 3. *F0 change (in ERB) in nuclear contour organized by task, language and sentence type.*

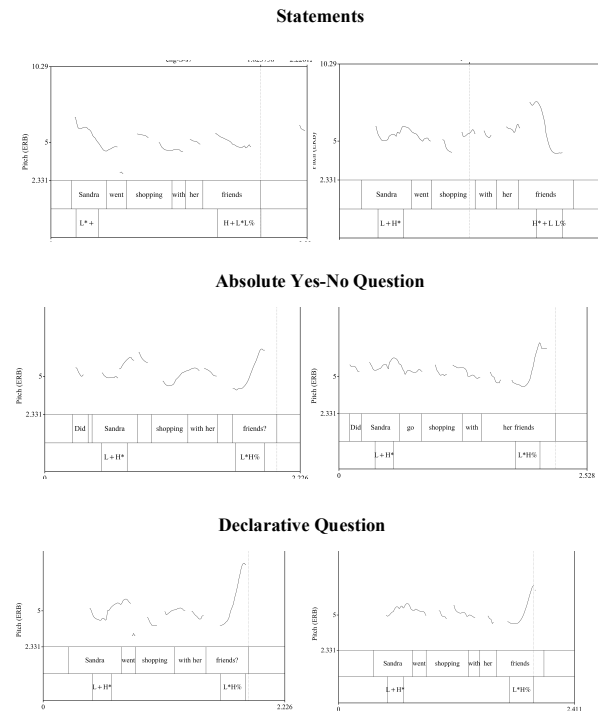


Figure 4. *Example pitch contours of L1 English (left) and L1Spanish-L2 English (right) speakers, for each sentence type in the NC task.*

An example of the realization of the pitch contours by the L1 English controls and L1 Spanish-L2 English speakers can be seen in Figure 4, for the sentence *Sandra went shopping with her friends*. Note that, in general, pitch movements are similar for both groups of speakers, which reflect the results obtained from our phonetic analysis.

4.3. Perception vs. Production

In this final section, we compare the error rates in the perception and production of DQs for the C task only, since this is the only task in which production errors can be calculated. To compute the errors, we determined how many times a given speaker produced a sentence other than a DQ (i.e., S or AQ). As seen in Figure 5, all speakers produced other types of sentences in place of DQs, with rates ranging from 35% to 100%. In general, the speakers with the largest error rates in production also had the largest error rates in perception. However, one speaker (S13) did not make any errors in the perception task, yet produced the incorrect sentence type 100% of the time. These findings suggest that this particular speaker was able to determine in which context DQs were more appropriate in perception, but had not yet acquired the complementary distribution of DQs vs. AQs in production. A future analysis of the individual results will need to determine whether this participant distinguishes the two sentence types prosodically (e.g., a larger F0 excursion in DQs than in AQs).

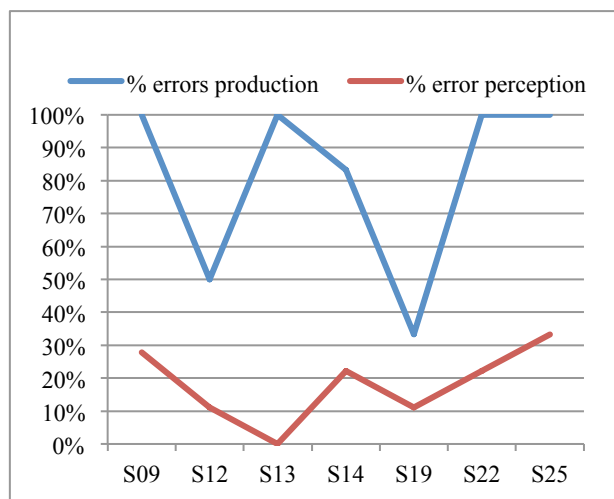


Figure 5. Error rates for declarative questions in the contextual task by speaker, in both perception and production.

5. DISCUSSION AND CONCLUSION

Going back to our first question, we conclude that L1 Spanish speakers do differ from controls across tasks and sentence types, specifically in production, and to a lesser extent, in perception. As predicted, the largest proportion of errors in perception was obtained for DQs in the C task, demonstrating that learners have not fully acquired the distribution of English DQs and AQs. An analysis of the individual L2 speakers reveals that all but one speaker had difficulty selecting DQs in the target context, and that the most common error was the selection of AQs in place of DQs. Thus, even though the participants of this study were advanced or near-native

speakers of English with several years of experience living in an English-speaking environment, they still had not acquired the distinction between AQs and DQs.

Differences in production were also observed in DQs. In particular, 83% of the expected DQs were produced as AQs in the C task. This is not surprising, as our perception results indicate that the AQ vs. DQ distinction had not been fully acquired. The analysis of perception vs. production revealed that overall, the speakers who demonstrated the most difficulty distinguishing between AQs and DQs in the perception tasks faced the same difficulty distinguishing between AQs and DQs in the production tasks. Nevertheless, the most accurate L2 speaker in the perception task did not produce a single DQ, which demonstrates that the comprehension of the DQ/AQ distinction does not necessarily mean that the learner will use that knowledge in production. Although the differences in the realization of PAs and nuclear contours by L1 vs. L2 speakers were not found to be significant, within-group differences were larger in the realization of PAs than in the realization of nuclear contours. Spanish speakers did produce a higher PA in Qs than in Ss (especially in the C task), but the differences between groups were obscured by the fact that controls also produced a higher PA in Qs than in Ss, which was not expected. Future research on the realization of English sentence types should investigate this in more detail.

As concerns our second question, we conclude that the type of task plays an important role, in line with previous findings [1,2,3,4]. The proportion of errors in the C task was higher for both groups in the perception task. In production, L1 Spanish speakers overextended the use of AQs to contexts where DQs were expected. They also displayed, similar to the controls, a smaller pitch excursion both in PAs and nuclear contours in contextualized than in non-contextualized tasks. Thus, we can tentatively conclude that the syntactic differences are larger than the prosodic differences. This was indeed expected considering that the two languages, as concerns the realization of sentence type, are more similar prosodically than syntactically.

To conclude, our results are consistent with previous research that has reported differences across tasks in first and second language speech. Research on monolingual speakers has consistently shown a larger degree of variability in more spontaneous tasks when compared to results obtained in controlled experiments [1,12]. Variability across tasks was also observed in the group of advanced and near-native L2 learners tested here. In particular, our results suggest that CLI is larger in tasks that require a higher processing demand [3,4]. Interestingly, L2 speakers seem to pattern in this respect with early bilinguals, who show different patterns of prosodic [2] and segmental [13] convergence depending on the formality of the task.

6. ACKNOWLEDGMENT

This research was supported by a Social Sciences and Humanities research –SSHRC– grant (#890-2011-0049) from the government of Canada.

7. REFERENCES

- [1] N. Hedberg and J. Sosa, "The prosody of topic and focus in spontaneous English dialogues," in *Topic and focus: Cross-linguistic perspective on meaning and intonation*, pp. 101-120, 2008.
- [2] L. Colantoni, A. Cuza, and N. Mazzaro. "Task-related effects in the Spanish of heritage speakers and long-term immigrants," in *Intonational grammar in Ibero-Romance: Approaches across linguistic subfields*, to appear.
- [3] E. Grabe, B. S. Rosner, J. Garcia-Albea, and X. Zhou, "Perception of English intonation by English, Spanish, and Chinese listeners," *Language and Speech*, vol. 46, no. 4, pp. 375-401, 2003.
- [4] M. Ortega Llebaria and L. Colantoni, "L2 English intonation: Relations between form-meaning associations, access to meaning and L1 transfer," *Studies in second language acquisition*, vol. 36, pp. 331-353, 2014.
- [5] C. Gussenhoven, *The phonology of tone and intonation*, Cambridge: Cambridge University Press, 2004.
- [6] J. M. Sosa, *La entonación del español: su estructura fónica, variabilidad y dialectología*, Madrid: Cátedra, 1999.
- [7] J. I. Hualde, *The sounds of Spanish*, Cambridge: Cambridge University Press, 2005.
- [8] C. Bartels, *Towards a compositional interpretation of English statement and question intonation*. PhD dissertation, University of Massachusetts Amherst, 1997.
- [9] I. Mennen, "Bi-directional interferences in the intonation of Dutch speakers of Greek", *Journal of Phonetics*, vol. 32, pp. 543-563, 2004.
- [10] P. Trofimovich and W. Baker, "Learning second language suprasegmentals: effects of L2 experience on prosody and fluency characteristics of L2 speech", *Studies in Second language acquisition*, vol. 28, pp. 1-30, 2006.
- [11] P. Boersma and D. Weenink, *Praat: Doing phonetics by computer* [Computer program], 2015.
- [12] T. Face, "Intonation in Spanish declaratives: Differences between lab speech and spontaneous speech", *Catalan Journal of Linguistics*, vol. 2, pp. 115-131, 2003.
- [13] R. Rao, "Manifestations of /b, d, g/ in two groups of Heritage Speakers of Spanish," paper presented at the *2012 UIC Bilingualism Forum*, University of Illinois-Chicago, Chicago, IL, October 2012.