Supplementary materials

Traditional analyses

This section contains additional information regarding the response accuracy and response time analyses, as well as tables reported but not included in the main text.

Learner response accuracy.

The population effects of the response accuracy model were specified in the following manner:

$$\begin{aligned} is\_correct_{ij} &\sim Bernoulli(p_{ij}, m_{ij}) \\ logit(p_{ij}) &= \beta_0 + \beta_1 * question\_type + \beta_2 * LexTALE + \beta_3 * EQ + \\ &\quad \beta_4 * question\_type * LexTALE * EQ \end{aligned}$$

We employed the `0 + Intercept` syntax of `brms` and set weakly informative priors as follows:

$$\begin{aligned} \beta &\sim Normal(0,0.3) \\ \sigma &\sim Cauchy(0,0.1) \\ \rho &\sim LKJcorr(2) \end{aligned}$$

The summary of the response accuracy model is available in Table 2. The information provided in this table is equivalent to the left panel of Figure 2 in the manuscript.

Table 2: Summary of the posterior distribution modeling response accuracy as a function of utterance type, LexTALE, and Empathy quotient. The table includes posterior medians, the 95% HDI, the percentage of the HDI within the ROPE, and the maximum probability of effect (MPE).

Parameter	Median	HDI	% in ROPE	MPE	Rhat	ESS
Intercept	0.53	[0.23, 0.82]	0.00	1.00	1.00	2099
Wh- question	0.43	[0.17, 0.65]	0.00	1.00	1.00	2723
Narrow focus	2.13	[1.84, 2.37]	0.00	1.00	1.00	2608
Broad focus	2.34	[2.05, 2.59]	0.00	1.00	1.00	2576
LexTALE	0.28	[0.15, 0.41]	0.00	1.00	1.00	5245
Empathy quotient	−0.02	[−0.11, 0.09]	0.98	0.62	1.00	5351
Wh- question:LexTALE	0.12	[−0.06, 0.29]	0.42	0.90	1.00	4977
Narrow focus:LexTALE	0.02	[−0.17, 0.22]	0.72	0.58	1.00	8450
Broad focus:LexTALE	0.19	[−0.02, 0.42]	0.18	0.96	1.00	9482
Wh- question:EQ	0.20	[0.03, 0.36]	0.10	0.99	1.00	4912
Narrow focus:EQ	0.26	[0.08, 0.43]	0.02	1.00	1.00	8389
Broad focus:EQ	0.24	[0.05, 0.43]	0.05	0.99	1.00	8180
LexTALE:EQ	0.02	[−0.09, 0.14]	0.93	0.65	1.00	5954
Wh- question:LexTALE:EQ	0.19	[0.00, 0.39]	0.16	0.97	1.00	5604
Narrow focus:LexTALE:EQ	0.02	[−0.19, 0.23]	0.67	0.56	1.00	8344
Broad focus:LexTALE:EQ	0.08	[−0.17, 0.32]	0.51	0.74	1.00	8698

Drift diffusion models
Drift Diffusion Models (DDM), also referred to as Wiener Diffusion Models and Decision Diffusion Models, represent our preferred method for analyzing the data from our 2AFC task. DDMs are rarely used in SLA research, though they are commonplace in psychology. The primary selling point of using a DDM is related to the parameters the model estimates: boundary separation (α), drift rate (δ), bias (β), and non-decision time (τ). Together these parameters give rich information about the processes believed to underpin decision-making. Specifically, a DDM requires decision data, e.g., “left” or “right” choices, correct or incorrect responses, etc., and response times associated with said decisions. In linguistics, particularly in psycholinguistics, data of this nature derived from 2AFC tasks are often analyzed using separate models, one for responses, and another for response times (as we have done in our so-called ‘traditional analyses’).[7] As mentioned, a DDM uses both of these dependent variables—responses and response times—to estimate the 4 aforementioned parameters. The estimates can then be scrutinized in subsequent models, if one estimates the parameters for each participant (i.e., the approach taken in the present work), and/or used for simulations. For our purposes, we employ the Bayesian implementation of the DDM, thus we sample from a posterior distribution of plausible estimates for α, δ, β, and τ for each participant. We then summarize and report these posterior distributions for statistical inferences.

The no-pooling models were fit using the following specification in `brms`:

```
  rt_raw | dec(is_correct) ~ 0 + sentence_type,
  bs ~ 0 + sentence_type,
  ndt ~ 0 + sentence_type,
  bias ~ 0 + sentence_type
```

and the priors were:

```
prior("normal(0, 1)", class = "b"),
prior("normal(0, 5)", class = "b", dpar = "bs"),
prior("normal(0.2, 1)", class = "b", dpar = "ndt"),
prior("normal(0.5, 1)", class = "b", dpar = "bias")
```

The complete code used to fit the models are available in `09_ddm.R` in the r scripts directory.

Measurement-error models. The measurement error models fit to the boundary separation and drift rate data were specified to include the standard error around each posterior median for α and δ:

$$\begin{aligned} \alpha &\sim Normal(\alpha_{n,TRUE}, SE_{\alpha}) \\ \delta &\sim Normal(\delta_{n,TRUE}, SE_{\delta}) \end{aligned}$$

---

[7] Given how relatively uncommon DDMs are in linguistics, the present work includes both approaches, though it is reasonable to assume that this practice will diminish as DDMs become more well-known and the resources for implementing them become more user-friendly.

The priors for the drift rate model were:

$$
\begin{aligned}
\alpha &\sim \mathit{Normal}(1,0.5) \\
\beta &\sim \mathit{Normal}(0,0.3) \\
\tau &\sim \mathit{Cauchy}(0,0.3) \\
\sigma &\sim \mathit{Cauchy}(0,0.1) \\
\rho &\sim \mathit{LKJcorr}(2)
\end{aligned}
$$

and the priors for the boundary separation model were:

$$
\begin{aligned}
\alpha &\sim \mathit{Normal}(2,0.5) \\
\beta &\sim \mathit{Normal}(0,0.5) \\
\tau &\sim \mathit{Cauchy}(0,0.3) \\
\sigma &\sim \mathit{Cauchy}(0,0.1) \\
\rho &\sim \mathit{LKJcorr}(2)
\end{aligned}
$$

To specify this type of model in `brms` we use the `resp_se` function, as follows:

```
estimate | resp_se(se, sigma = TRUE) ~ 1 + # Criterion
  q_sum * lextale_std * eq_std +          # Population-level effects
  (1 + q_sum * lextale_std * eq_std | participant) # Group-level effects
```

The model summary is available in Table 3, which is equivalent to Figure 6 in the main document.

Table 3: Summary of the posterior distribution modeling boundary separation and drift rate as a function of question type, LexTALE, and Empathy quotient. The table includes posterior medians, the 95% HDI, the percentage of the HDI within the ROPE, and the maximum probability of effect (MPE).

Model	Parameter	Median	HDI	MPE	Rhat	ESS
Boundary	Intercept	1.77	[1.70, 1.83]	1.00	1.00	3407
separation	Question type	−0.04	[−0.08, −0.01]	0.99	1.00	3676
	LexTALE	0.14	[0.06, 0.22]	1.00	1.00	3460
	EQ	0.04	[−0.02, 0.11]	0.91	1.00	3585
	Question type:LexTALE	−0.05	[−0.09, 0.00]	0.97	1.00	3594
	Question type:EQ	−0.01	[−0.04, 0.03]	0.73	1.00	3993
	LexTALE:EQ	0.12	[0.03, 0.20]	1.00	1.00	3912
	Question type:LexTALE:EQ	−0.02	[−0.07, 0.03]	0.77	1.00	3580
Drift rate	Intercept	1.23	[1.20, 1.26]	1.00	1.00	3814
	Question type	0.08	[0.06, 0.10]	1.00	1.00	3584
	LexTALE	0.02	[−0.02, 0.05]	0.83	1.00	3276
	EQ	0.00	[−0.03, 0.03]	0.59	1.00	4063
	Question type:LexTALE	0.01	[−0.02, 0.05]	0.70	1.00	3846
	Question type:EQ	0.00	[−0.02, 0.02]	0.53	1.00	4123
	LexTALE:EQ	−0.06	[−0.11, −0.02]	1.00	1.00	4114
	Question type:LexTALE:EQ	0.01	[−0.03, 0.05]	0.66	1.00	3733

Supplementary analyses
In this section we present supplementary analyses, all of which are exploratory in nature.

D'. Figure 9 and Table 4 represent an exploratory analysis of d' scores as a function of utterance type and speaker variety. One observes similar patterns to those from the accuracy analysis presented in the manuscript. The primary takeaway is that the analysis of learners' sensitivity to Spanish prosody mirrors that of their accuracy. That is to say, learners are more sensitive to (and accurate with) statements (broad focus, narrow focus) than questions (wh-, yes/no) (left panel of Figure 9. Learner sensitivity to the different Spanish varieties represented in the stimuli pattern in the same manner, i.e., more sensitivity to the Peninsular variety and less sensitivity to the Cuban variety (right panel of Figure 9). Table 4 summarizes the posterior of these exploratory analyses.

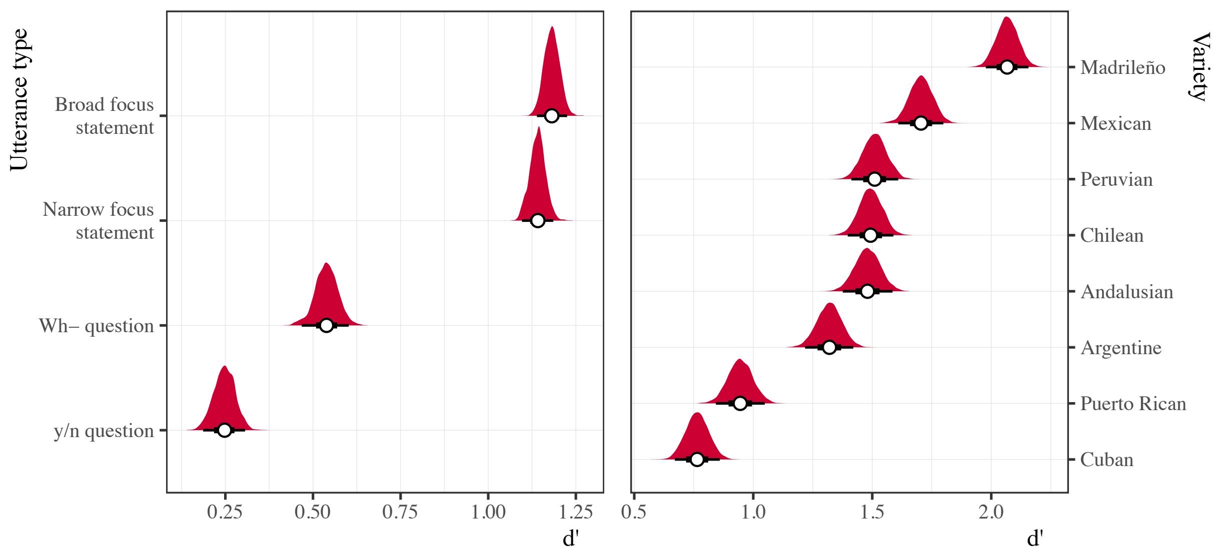

Figure 9. Exploratory analysis of d' as a function of utterance type and speaker variety. Points represent posterior medians with 66% and 95% credible intervals.

Table 4: Summary of the posterior distribution modeling d' as a function of question type or speaker variety. The table includes posterior medians, the 95% HDI, and the maximum probability of effect (MPE).

Model	Parameter	Median	HDI	MPE
Utterance type	Broad focus statement	1.18	[1.14, 1.23]	1.00
	Narrow focus statement	1.14	[1.10, 1.19]	1.00
	Wh- question	0.54	[0.47, 0.61]	1.00
	y/n question	0.25	[0.19, 0.31]	1.00
Variety	Andalusian	1.48	[1.38, 1.59]	1.00
	Argentine	1.32	[1.22, 1.42]	1.00
	Chilean	1.49	[1.40, 1.59]	1.00
	Cuban	0.76	[0.67, 0.86]	1.00
	Mexican	1.70	[1.61, 1.80]	1.00
	Peninsular	2.07	[1.98, 2.15]	1.00
	Peruvian	1.51	[1.42, 1.61]	1.00
	Puerto Rican	0.95	[0.85, 1.05]	1.00

Familiar vs. unfamiliar varieties. At the beginning of the experimental session, participants responded to the prompt ‘I am most familiar with Spanish from’ and, using a drop-down window, made a selection from the following choices: I am not familiar with Spanish, Argentina, Bolivia, Chile, Colombia, Costa Rica, Cuba, Dominican Republic, Ecuador, El Salvador, Equatorial Guinea, Guatemala, Honduras, Mexico, Nicaragua, Panama, Paraguay, Peru, Philippines, Puerto Rico, Uruguay, Venezuela, Spain, and United States. These choices represent the countries/territories/commonwealths where Spanish is spoken as an official or co-official language with the exception of Andorra, Belize, Gibraltar, and the Philippines, which were not included by mistake. Table 5 below summarizes the participants’ responses to this question. Of note, participants overwhelmingly selected ‘U.S. Spanish’, followed by ‘Mexico’, ‘Peninsular’, and ‘Not familiar’.

Table 5: Summary of participant self-reported familiarity with Spanish.

Variety	n	Proportion
U.S. Spanish	78	0.35
Mexico	47	0.21
Peninsular	44	0.20
Not familiar	38	0.17
Colombia	4	0.02
Costa Rica	4	0.02
Puerto Rico	4	0.02
Dominican Republic	2	0.01
Honduras	2	0.01
Peru	2	0.01

We did not pre-register a hypothesis regarding participant familiarity with Spanish. The following analysis is exploratory in nature. We analyzed the data from the participants who claimed to be most familiar with a Spanish variety that was included in our speaker varieties: Peninsular and Mexican Spanish (note: we make the assumption that ‘Peninsular’ is most closely associated with the Madrileño speaker). We coded the participants’ responses to familiar versus unfamiliar varieties and fit a Bayesian logistic regression model to the data. The model was specified similar to previous models:

$$
\begin{aligned}
\text{is\_correct}_{ij} &\sim Bernoulli(p_{ij}, m_{ij}) \\
\text{logit}(p_{ij}) &= \beta_0 + \beta_1 * question\_type + \beta_2 * Familiarity \\
&\quad \beta_3 * question\_type * Familiarity
\end{aligned}
$$

We specified grouping variables for participant, speaker variety, and individual items. The models estimated varying slopes for the sentence type effect for each participant and the familiarity effect for each item. Again, we used the `0 + Intercept` syntax of `brms` and set weakly informative priors as follows:

$$
\begin{aligned}
\beta &\sim Normal(0,1) \\
\sigma &\sim Cauchy(0,0.2) \\
\rho &\sim LKJcorr(1)
\end{aligned}
$$

In short, we find that, marginalizing over proficiency and empathy, participants were more accurate when responding to a familiar variety. This is true for all utterance types to some degree, but more clearly the case for questions (likely because responses to declarative utterances were near ceiling). Figure 8 included in the manuscript illustrates the familiarity effect. Table 6 summarizes the model output. For convenience, we also provide the conditional effects of response accuracy in Table 7.

Table 6: Summary of the posterior distribution modeling response accuracy as a function of utterance type and familiarity. The model only includes data from participants who claimed to be familiar with Mexican (n = 47) and Peninsular (n = 44) Spanish. The table includes posterior medians, the 95% HDI, the percentage of the HDI within the ROPE, and the maximum probability of effect (MPE).

Parameter	Median	HDI	% in ROPE	MPE	Rhat	ESS
Intercept	0.25	[−0.28, 0.79]	0.19	0.84	1	1610
Wh- question	1.13	[0.71, 1.55]	0.00	1.00	1	2066
Narrow focus	3.19	[2.71, 3.71]	0.00	1.00	1	2561
Broad focus	3.32	[2.87, 3.82]	0.00	1.00	1	3217
Familiar	1.28	[0.26, 2.17]	0.00	0.99	1	2404
Wh- question:Familiar	−0.24	[−0.76, 0.32]	0.20	0.80	1	5970
Narrow focus:Familiar	−0.98	[−1.66, −0.24]	0.00	0.99	1	6516
Broad focus:Famliar	−1.24	[−1.94, −0.51]	0.00	1.00	1	5998

Table 7: Conditional effects of response accuracy as a function of sentence type and familiarity with the Spanish variety. Values represent posterior medians along with the 95% HDI for unfamiliar and familiar conditions, along with the posterior difference (familiar - unfamiliar). The posterior predictive distribution is based on data from participants who claimed to be familiar with Mexican (n = 47) and Peninsular (n = 44) Spanish.

Sentence type	Unfamiliar	Familiar	Difference
Broad focus statement	0.97 [0.95, 0.99]	0.97 [0.94, 0.99]	0.00 [-0.04, 0.03]
Narrow focus statement	0.97 [0.95, 0.99]	0.98 [0.95, 1.00]	0.01 [-0.03, 0.04]
Wh- question	0.80 [0.70, 0.88]	0.92 [0.83, 0.97]	0.12 [0.00, 0.22]
y/n question	0.56 [0.43, 0.69]	0.82 [0.67, 0.93]	0.26 [0.07, 0.41]

Rising vs. falling contours in yes/no questions. As pointed out by one of the anonymous reviewers, in our stimuli, the yes/no questions have rising intonation for all varieties except Cuban and Puerto Rican (see *Auditory stimuli* below). We calculated response accuracy as a function of intonational contour (falling: Cuban, Puerto Rican; rising: other) for yes/no questions using L2 and native listener data. Table 8 provides descriptive statistics. One can see that both listener types were least accurate when responding to the falling contours, though this is particularly true for the Cuban variety. This finding is illustrated in the previous section with regard to familiarity and is seen again in the following section regarding native listener data.

Table 8: L2 and native listener response accuracy to yes/no questions as a function of variety (Caribbean, Non-Caribbean). The stimuli from the Caribbean varieties include falling intonational contours.

Group	Cuban	Puerto Rican	Other
Native listeners	0.55 (0.5)	0.8 (0.4)	0.93 (0.25)
L2 learners	0.13 (0.33)	0.28 (0.45)	0.73 (0.44)

Similar to what was found in our primary analyses regarding yes/no questions, in this subset of the data, empathy had no effect on response accuracy regardless of whether the stimuli was produced with falling (i.e., Caribbean: Cuban, Puerto Rican) or rising (other) intonation (Figure 10).

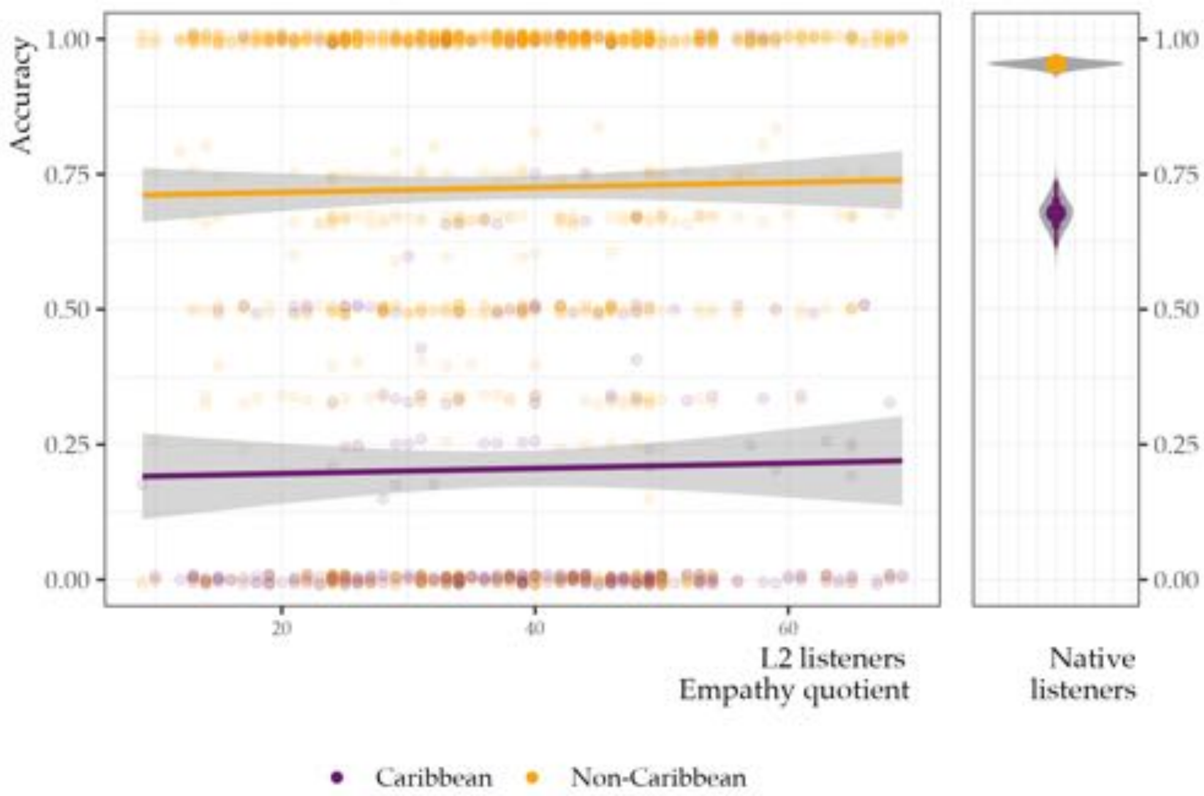

Figure 10. L2 and native listener response accuracy to Caribbean (Cuban, Puerto Rican) and non-Caribbean varieties for yes/no questions. Points represent by-participant means (learners) or posterior means (natives). For learners, the horizontal axis plots empathy quotient scores.

Monolingual response accuracy. In preparing our materials before official data collection, we piloted the 2AFC task and the auditory stimuli on monolingual Spanish speakers. The purpose of collecting this pilot data was to get an assessment of task difficulty—overall and as a function of speaker variety—and to have an idea what reasonable priors would be with regard to response times. From this pilot, we learned that, overall, monolingual speakers were least accuracy when responding to the Cuban variety, and to some degree the Puerto Rican variety as well. This finding led us to hypothesize that L2 learners would also have difficulties when responding to stimuli from the same varieties. Figure 11 plots the monolingual accuracy data (right panel) next to the learner accuracy data (left panel). The same information is also provided in numeric form in Table 9.

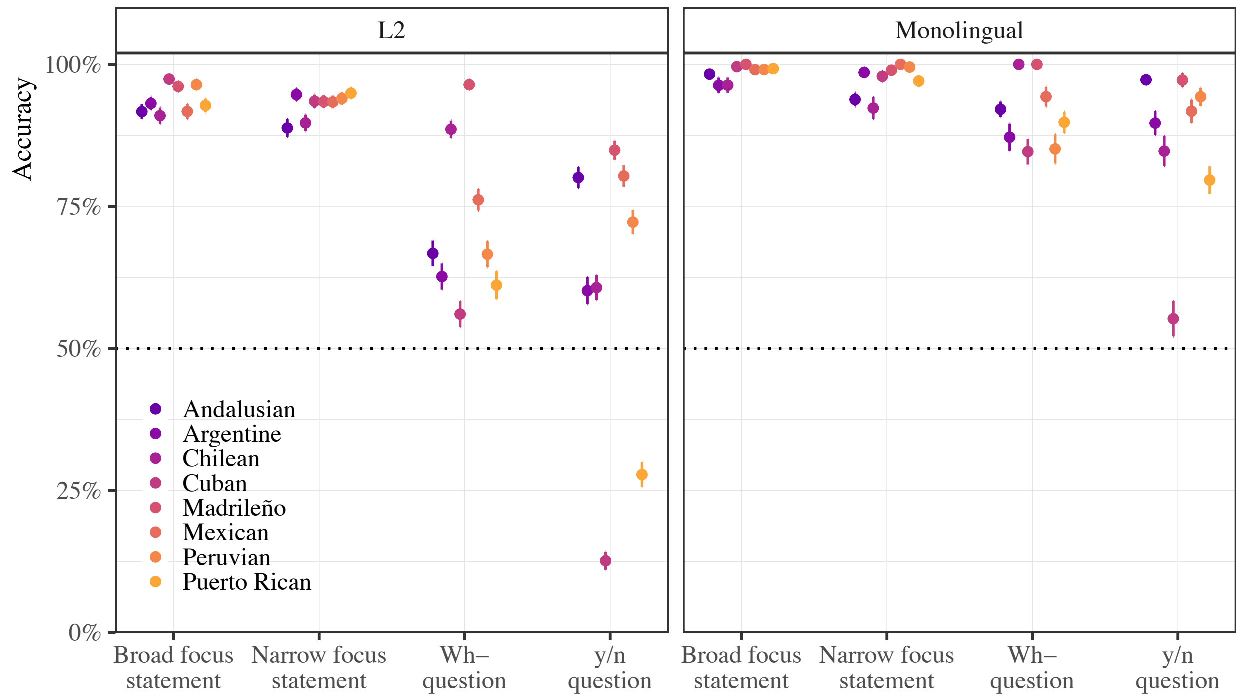

Figure 11. Response accuracy as a function of group (L2 learner, monolingual Spanish speaker), speaker variety (Andalusian, Argentine, Chilean, Cuban, Madrileño, Mexican, Peruvian, Puerto Rican), and utterance type (broad focus statement, narrow focus statement, wh- question, y/n question). Points represent means of the raw data surrounded by the standard error of the mean.

Table 9: Response accuracy as a function of group (L2 learner, monolingual Spanish speaker), speaker variety (Andalusian, Argentine, Chilean, Cuban, Madrileño, Mexican, Peruvian, Puerto Rican), and utterance type (broad focus statement, narrow focus statement, wh- question, y/n question). Each column provides the mean and standard error.

Type	Variety	L2	Monolingual
Broad focus statement	Andalusian	0.92 [0.90, 0.93]	0.98 [0.98, 0.99]
	Argentine	0.93 [0.92, 0.94]	0.96 [0.95, 0.98]
	Chilean	0.91 [0.90, 0.92]	0.96 [0.95, 0.98]
	Cuban	0.97 [0.97, 0.98]	1.00 [0.99, 1.00]
	Madrileño	0.96 [0.95, 0.97]	1.00 [1.00, 1.00]
	Mexican	0.92 [0.90, 0.93]	0.99 [0.98, 1.00]
	Peruvian	0.96 [0.96, 0.97]	0.99 [0.98, 1.00]
	Puerto Rican	0.93 [0.92, 0.94]	0.99 [0.99, 1.00]
Narrow focus statement	Andalusian	0.89 [0.87, 0.90]	0.94 [0.93, 0.95]
	Argentine	0.95 [0.94, 0.96]	0.99 [0.98, 0.99]
	Chilean	0.90 [0.88, 0.91]	0.92 [0.90, 0.94]
	Cuban	0.94 [0.92, 0.95]	0.98 [0.97, 0.99]
	Madrileño	0.93 [0.92, 0.95]	0.99 [0.98, 1.00]
	Mexican	0.93 [0.92, 0.95]	1.00 [1.00, 1.00]
	Peruvian	0.94 [0.93, 0.95]	1.00 [0.99, 1.00]
	Puerto Rican	0.95 [0.94, 0.96]	0.97 [0.96, 0.98]
Wh- question	Andalusian	0.67 [0.64, 0.69]	0.92 [0.91, 0.93]
	Argentine	0.63 [0.60, 0.65]	0.87 [0.85, 0.90]
	Chilean	0.89 [0.87, 0.90]	1.00 [1.00, 1.00]
	Cuban	0.56 [0.54, 0.58]	0.85 [0.82, 0.87]
	Madrileño	0.96 [0.96, 0.97]	1.00 [1.00, 1.00]
	Mexican	0.76 [0.74, 0.78]	0.94 [0.93, 0.96]
	Peruvian	0.67 [0.64, 0.69]	0.85 [0.83, 0.88]
	Puerto Rican	0.61 [0.59, 0.64]	0.90 [0.88, 0.92]
y/n question	Andalusian	0.80 [0.78, 0.82]	0.97 [0.96, 0.98]
	Argentine	0.60 [0.58, 0.63]	0.90 [0.88, 0.92]
	Chilean	0.61 [0.59, 0.63]	0.85 [0.82, 0.87]
	Cuban	0.13 [0.11, 0.14]	0.55 [0.52, 0.58]
	Madrileño	0.85 [0.83, 0.87]	0.97 [0.96, 0.98]
	Mexican	0.80 [0.78, 0.82]	0.92 [0.90, 0.94]
	Peruvian	0.72 [0.70, 0.74]	0.94 [0.93, 0.96]
	Puerto Rican	0.28 [0.26, 0.30]	0.80 [0.77, 0.82]

As a check on the low accuracy with the Cuban and Puerto Rican varieties, we decided to explore monolingual response accuracy further. To this end, we looked at the monolinguals' responses when they were presented with stimuli from their own variety, e.g., an Andalusian listener responding to stimuli from the Andalusian speaker. In our data, this implied a subset of Andalusian, Chilean, Cuban, Madrileño, Mexican, and Puerto Rican listeners. Figure 12 plots the variety-matched raw accuracy scores as a function of utterance type and Table 10 provides the same information in numeric form. Of note, all the monolinguals in our sample were at ceiling for all utterance types when responding to speakers from their own variety. This is taken as evidence that the auditory stimuli are accurate representations of questions and statements for these varieties.

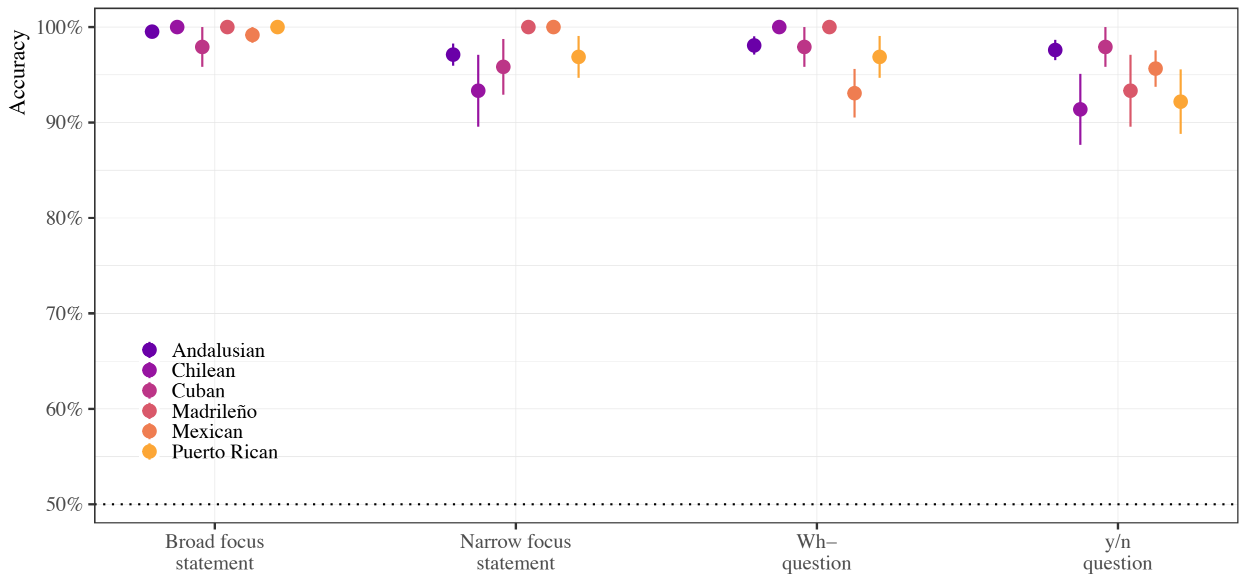

Figure 12. Variety-matched accuracy of monolingual listeners as a function of utterance type (broad focus statement, narrow focus statement, wh- question, y/n question). Points represent means of the raw data surrounded by the standard error of the mean.

Table 10: Variety-matched response accuracy as a function of utterance type. Accuracy refers to the proportion of correct responses along with the standard error of the mean.

Variety	Broad focus statement	Narrow focus statement	Wh- question	y/n question
Andalusian	1.00 [0.99, 1.00]	0.97 [0.96, 0.98]	0.98 [0.97, 0.99]	0.98 [0.97, 0.99]
Chilean	1.00 [1.00, 1.00]	0.93 [0.90, 0.97]	1.00 [1.00, 1.00]	0.91 [0.88, 0.95]
Cuban	0.98 [0.96, 1.00]	0.96 [0.93, 0.99]	0.98 [0.96, 1.00]	0.98 [0.96, 1.00]
Madrileño	1.00 [1.00, 1.00]	1.00 [1.00, 1.00]	1.00 [1.00, 1.00]	0.93 [0.90, 0.97]
Mexican	0.99 [0.98, 1.00]	1.00 [1.00, 1.00]	0.93 [0.91, 0.96]	0.96 [0.94, 0.98]
Puerto Rican	1.00 [1.00, 1.00]	0.97 [0.95, 0.99]	0.97 [0.95, 0.99]	0.92 [0.89, 0.96]

Auditory stimuli
Speakers. The auditory stimuli consisted of eight varieties of Spanish: Cuban, Peninsular-Madrileño, Peninsular-Andalusian, Puerto Rican, Chilean, Argentine, Mexican, and Peruvian. The speakers from Argentina, Chile, Peru, and Spain are linguists. Table 11 contains demographic information about the speakers.

Table 11: Demographic information for the eight varieties of Spanish represented in the auditory stimuli.

Country	City/Variety	Gender	Age
Argentina	Buenos Aires	Male	27
Chile	Valparaíso	Female	42
Cuba	Havana	Female	55
Mexico	Mexico City	Female	30
Peru	Lima	Male	30
Puerto Rico	Ponce	Female	35
Spain	Cádiz (Andalusia)	Female	35
Spain	Madrid	Female	29

Speech rate. In order to evaluate the possibility that the speech rate of the talkers in our stimuli may have affected response accuracy, we calculated the articulation rate (syllables spoken per second during phonation time), average syllable duration (in milliseconds), and speech rate (number of syllables divided by total time) for all items (64 items × 8 speakers = 512 utterances). These values are provided in Table 12.

Table 12: Average articulation rate (number of syllables divided by total phonation time), syllables duration (in milliseconds), and speech rate (number of syllable divided by total time) for each variety of the acoustic stimuli presented to listeners.

Variety	Articulation rate	Syllable duration	Speech rate
Andalusian	3.75	288	3.72
Argentine	3.59	300	3.59
Madrileño	3.72	282	3.72
Chilean	3.95	264	3.95
Cuban	3.99	272	3.99
Mexican	4.51	230	4.51
Peruvian	3.70	286	3.70
Puerto Rican	4.02	277	4.02

Figure 13 plots the posterior medians along with 66% and 95% HDI of standardized articulation rate for each variety. The plot shows that, generally, there was not a lot of variability between varieties. Though some of the slower varieties are also those in which we see higher response accuracy (compare with Figure 5), that was not always the case. That is, some of the faster varieties also had high response accuracy, e.g., Mexican Spanish.

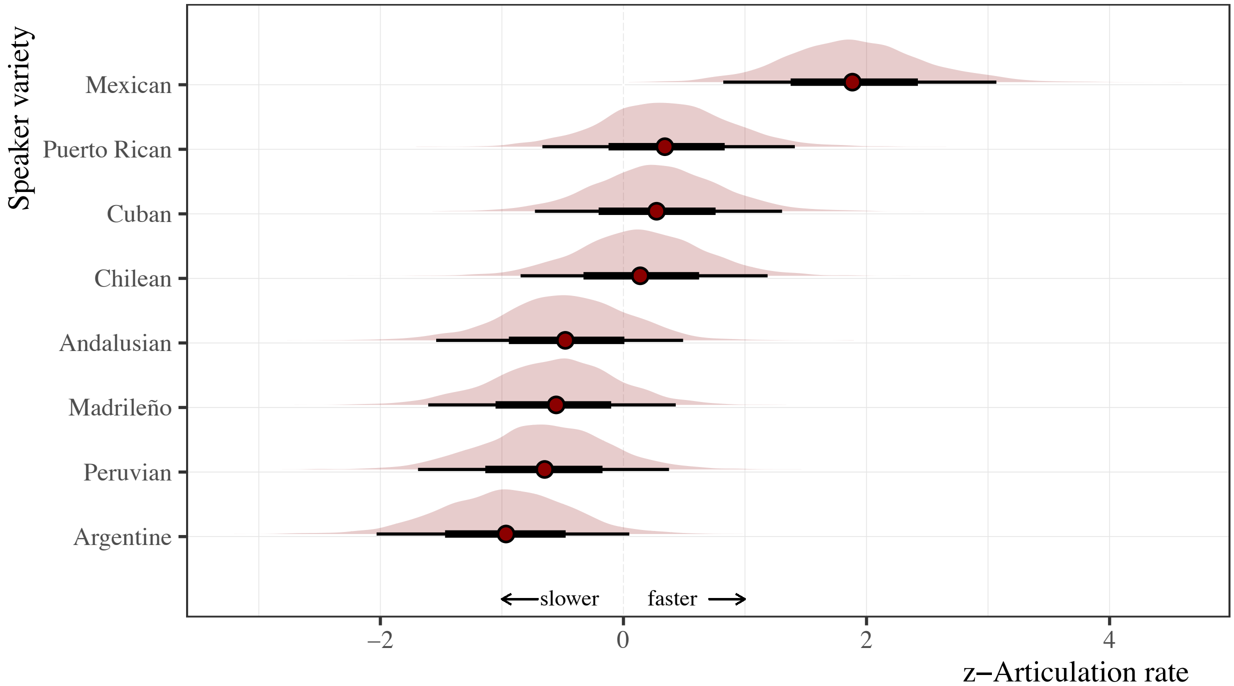

Figure 13. Standardized articulation rate as a function of speaker variety. Points represent posterior medians along with 66% and 95% HDI. The value ‘0’ on the horizontal axis represents the grand mean in the standardized space, which is equivalent to 3.90 syllables per second.

Acoustic description. In this subsection we provide examples of intonation contours for the four utterance types (broad focus state, narrow focus statement, wh- question, yes/no question) from the speakers of our eight varieties of Spanish (Cuban, Peninsular-Madrileño, Peninsular-Andalusian, Puerto Rican, Chilean, Argentine, Mexican, and Peruvian). Before proceeding to the acoustic descriptions, we will first provide an overview of the theoretical framework underpinning the descriptions.

The Autosegmental Metrical (AM) framework, developed by researchers like Pierrehumbert (1980) and Ladd (2008), aims to map the instrumentally-derived F0 values onto sequential, discrete and categorically-distinct intonational events. Within the AM framework, F0 is mapped onto pitch targets. Pitch targets can be monotonal, H(igh) and L(ow), or bitonal combinations of H and L, which represent rises (e.g., L+H*) or falls (e.g., H+L*). Pitch targets are identified at three points in an utterance: * indicates a pitch accent, which is associated and aligned with a stressed syllable; - indicates a phrase accent, which is associated with the boundary of an intermediate phrase; and % indicates a boundary tone or intonational phrase. It is important to note that an H at any prosodic level is the same H. The symbol that comes associated with it, whether *, -, or %, is merely giving information about how that tone is aligned with a specific event or boundary, not about the phonetic realization of that tone.

These abstract representations are mapped individually to instances of phonetic realizations specific to each language, similar to how /t/ has different acoustic properties in English and Spanish, as well as different phonetic realizations within a language depending on the context, similar to allophones. This is further complicated by dialect, as different dialects may present different relationship mappings.

The Tones and Breaks Indices (ToBI) labeling system is used within the AM framework to annotate utterances for intonation. ToBI has two obligatory tiers: an orthographic tier, on which the utterance is recorded in IPA; and a tone tier, on which the pitch targets are recorded. A third, optional tier is the break-index tier, on which prosodic breaks and their relative strengths are recorded on a scale of 0 (no break) to 4 (Intonation Phrase break). It is idealized that each variety of a language should have its own ToBI system due to different varieties having distinct, but similar, intonational inventories. For example, in Dominican and Puerto Rican Spanish varieties, intonational phonologists may identify a bitonal fall H+L*, but the phonetic realization and pragmatic contexts in which that tone is produced may differ. For the present work we adopt the Spanish Tones and Breaks Indices (Sp_ToBI) as described in Hualde and Prieto (2015).

Using the recordings, we generated figures that display the waveform, spectrogram, and fundamental frequency contours. The figures are accompanied with two annotation tiers: an orthographic tier and a tone tier. We use this information together to identify associations between tonal targets and specific structure within example utterances following the Autosegmental Metrical model (Pierrehumbert, 1980). We follow the conventions of the Spanish Tones and Break Indices (Sp_ToBI) and its revisions (Beckman, Díaz-Campos, McGory, & Morgan, 2002; Face & Prieto, 2007; Sosa, 2003a; Vilaplana Estebas & Prieto, 2008). Our description is not meant to be exhaustive, but rather to give the interested reader an idea of how some of the contours included in our stimuli looked. Here we provide a single example of each utterance type from each speaker (Figures 14-21), though we encourage the interested reader to take a look/listen to all items, which are freely available on the OSF in the `data/stimuli` directory.

The statement utterances from our speaker of Madrid Spanish included a nuclear configuration of L* L%. The examples provided in the top row of Figure 14 are primarily differentiated by a prenuclear L+>H* in the broad focus condition and L+H* in the narrow focus condition. Wh- and yes/no question had a similar prenuclear distinction, respectively. Our speaker also tended to use a final rise (L* ¡H%) in wh- questions, which differs from the more common final fall. This is often associated with politeness (Quilis, 1993; Sosa, 2003b), or possibly “[…] a nuance of interest and greater speaker involvement in the speech act” (Estebas-Vilaplana & Prieto, 2010, p. 35). We believe we may have obtained more formal utterances, or possibly ‘lab’ speech due to the context in which the stimuli were recorded, i.e., in a lab.[8] For more examples, see http://prosodia.upf.edu/atlasentonacion/enquestes/espanol/madrid/index-english.html#maptask (“¿Dónde está tu gasolina?”, minute: -3:18). For more description of Madrileño Spanish see also Quilis (1981), Quilis (1987), and Sosa (1999).

[8] For more information on lab speech, see Face (2003) and Xu (2010).

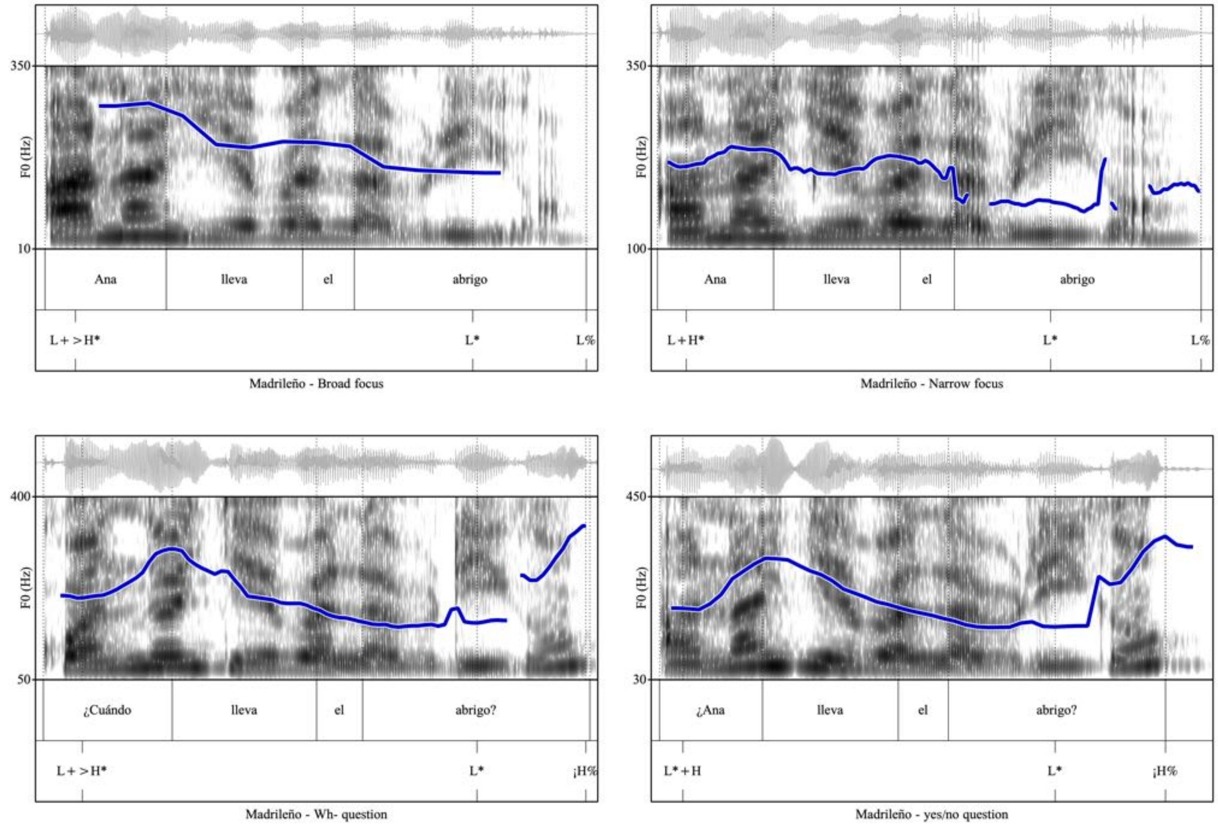

Figure 14. Waveform, spectrogram, and F0 trace exemplifying a broad focus (top left), narrow focus (top right), wh- question (bottom left), and yes/no question (bottom right) from a 29 year old female speaker of Madrileño Spanish.

Figure 15 plots the four utterance types as spoken by the Andalusian speaker. Similar to speaker of Madrileño Spanish, we do not see a major distinction between declaratives in nuclear position (H+L* L%). We found a tendency for this speaker to a use statement-of-the-obvious pattern where we elicited narrow focus. We show the nuclear configuration for wh- questions as L+¡H* L%, but this speaker also used L+H* L%. A final rise for wh- questions is also documented in this variety (Henriksen et al., 2010) and is associated with formality. For yes-no questions, the boundary tone was typlically ¡H%. More information on Western Andalusian intonation is available Sosa (1999) and Henriksen and García-Amaya (2012).

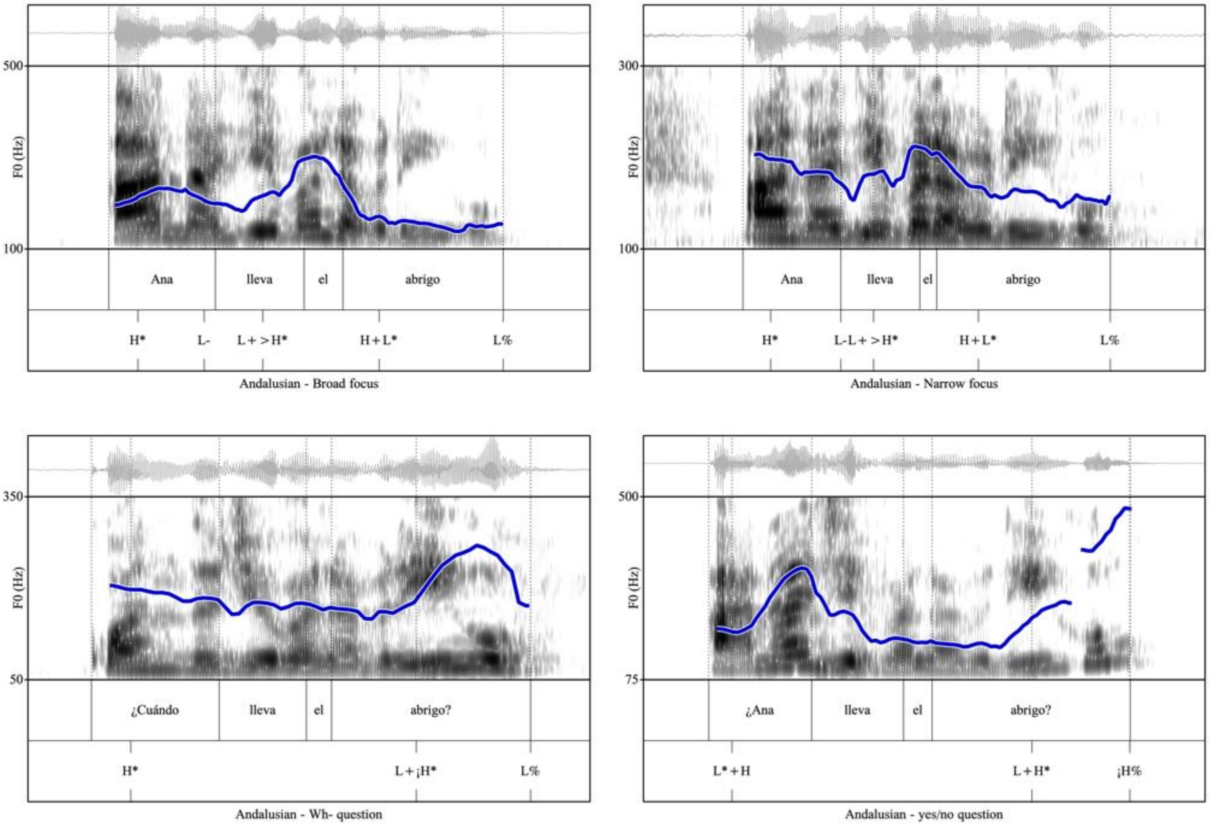

Figure 15. Waveform, spectrogram, and F0 trace exemplifying a broad focus (top left), narrow focus (top right), wh- question (bottom left), and yes/no question (bottom right) from a 35 year old female speaker of Andalusian (San Fernando, Cádiz) Spanish.

The Argentine speaker (Figure 16) generally produced declaratives with L* L% nuclear configurations. Notably absent from the narrow focus stimuli is the tritonal accent, L+H*+L (Labastía, 2006, among others; See Sosa, 1999). For wh- and yes/no questions, we see L% and HL% nuclear boundary tones, respectively. For more information on Argentine intonation we recommend Colantoni and Gurlekian (2004), Colantoni (2011), and Labastía (2011).

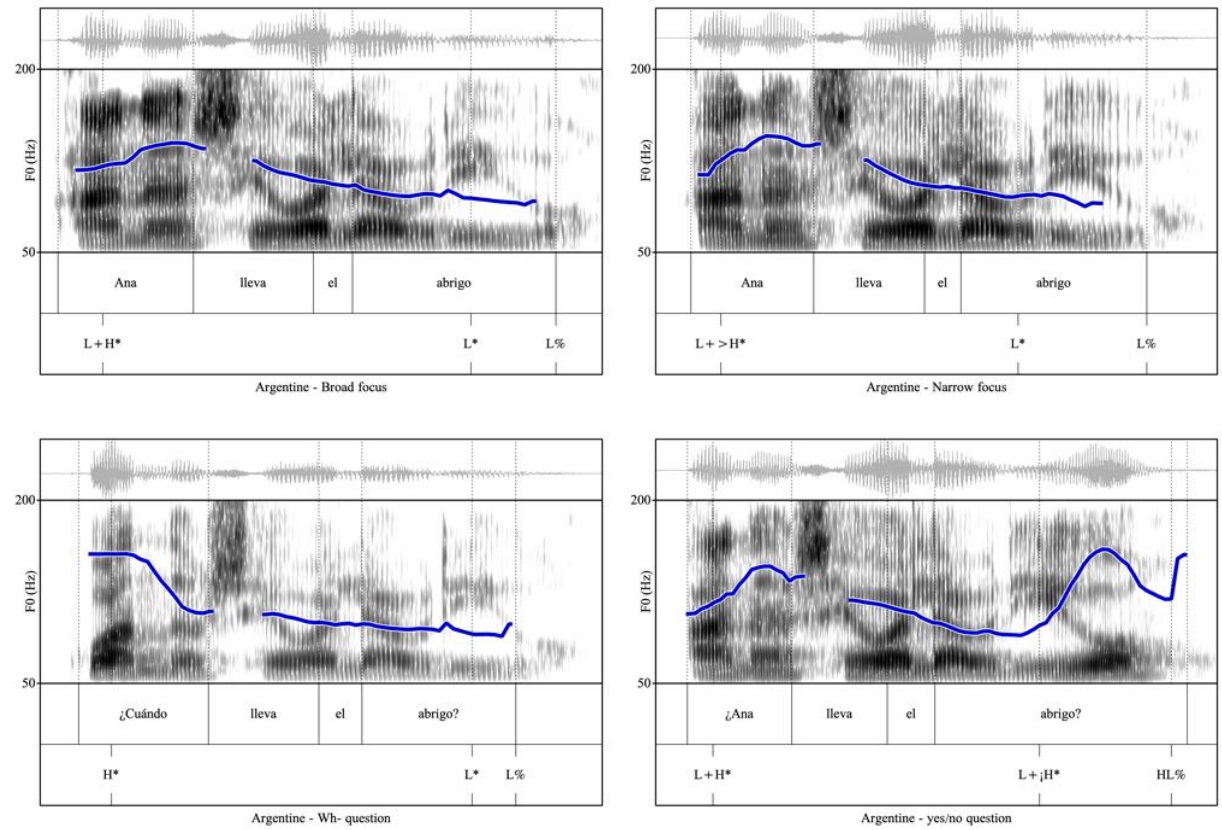

Figure 16. Waveform, spectrogram, and F0 trace exemplifying a broad focus (top left), narrow focus (top right), wh- question (bottom left), and yes/no question (bottom right) from a 27 year old male speaker of Argentine (Buenos Aires) Spanish.

Figure 17 illustrates example utterances from the Chilean speaker. We found that broad and narrow focus statements differed in their nuclear configuration (H+L* L% vs. L+H* L%, respectively). We also noted that wh- and yes/no questions differed in the nuclear pitch accents (L* vs. H+L*, respectively). For more information on Chilean intonation, we refer the reader to Ortiz-Lira and Cid-Uribe (2000) and Ortiz-Lira (2003).

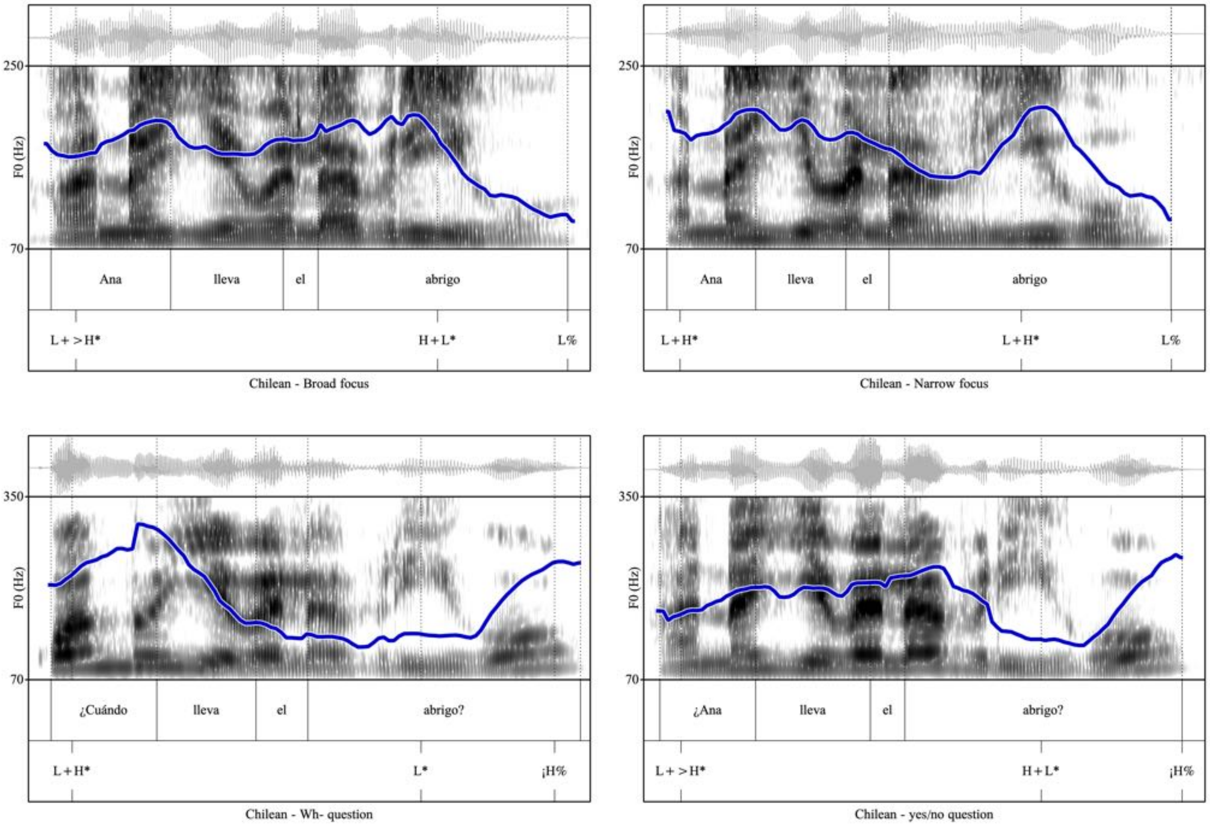

Figure 17. Waveform, spectrogram, and F0 trace exemplifying a broad focus (top left), narrow focus (top right), wh- question (bottom left), and yes/no question (bottom right) from a 42 year old female speaker of Chilean (Valparaíso) Spanish.

In Figure 18 we see example utterances from the Cuban speaker. We typically only found a distinction in broad and narrow focus items in the prenuclear pitch accents (L+H* and L+>H*, respectively). Both questions showed followed falling patterns in nuclear position. For example, we found L* L% in wh- questions and L+H* L% in yes/no questions. In prenuclear position, our speaker also used L+>H* in yes/no questions. For more information, the interested reader can consult Sosa (1999), or Alvord (2006) for an account of Cuban Spanish in contact with English in Miami, Florida.

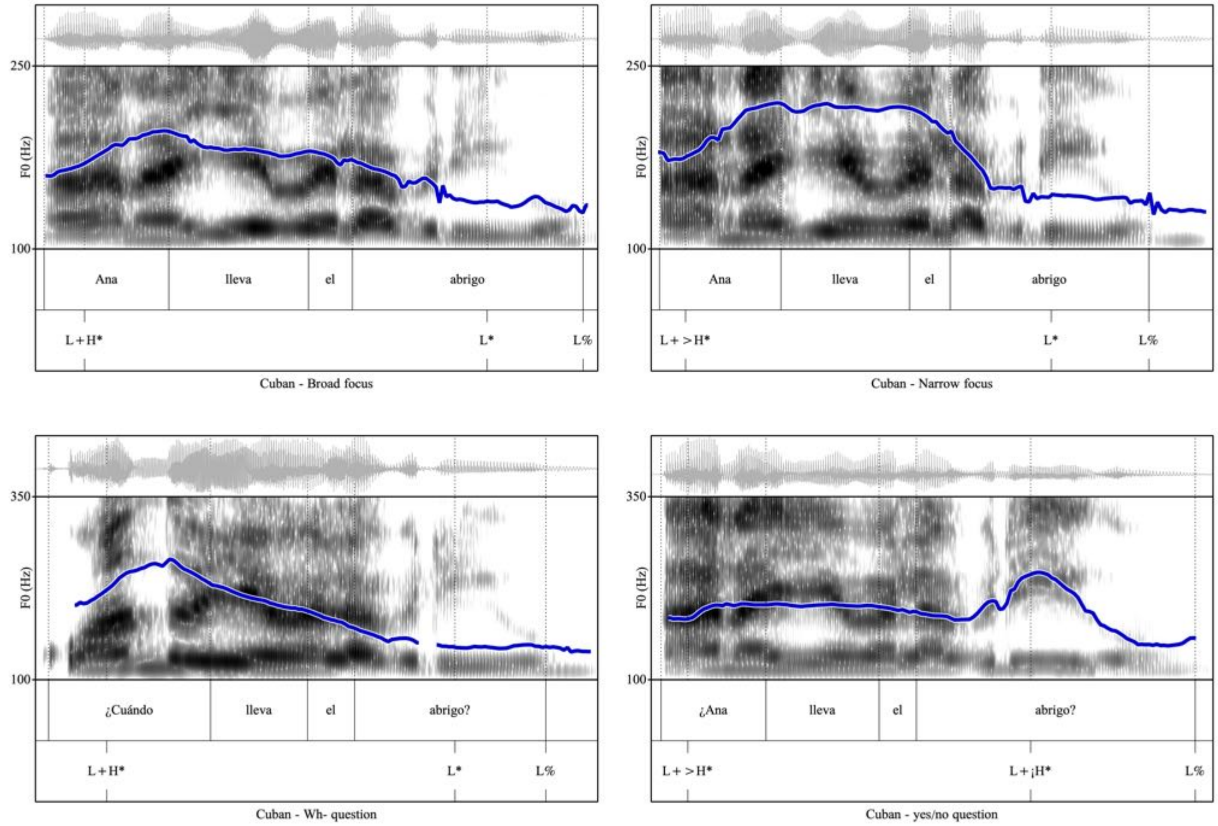

Figure 18. Waveform, spectrogram, and F0 trace exemplifying a broad focus (top left), narrow focus (top right), wh- question (bottom left), and yes/no question (bottom right) from a 55 year old female speaker of Cuban (Havana) Spanish.

The Mexican speaker distinguished between broad and narrow focus using a L* L% pattern in the former and L+H* H% in the latter. For wh- questions in this variety, final falls are considered the unmarked configuration (Sosa, 2003b), though we observed some rises, as in Figure 19 (bottom left plot). For yes/no questions this speaker produced a L* ¡H% nuclear configuration. Mexican Spanish is well-documented. We refer the reader to Quilis (1981), Quilis (1987), Quilis (1993), Sosa (1999), Butragueño (2003), and Butragueño (2004) for more information.

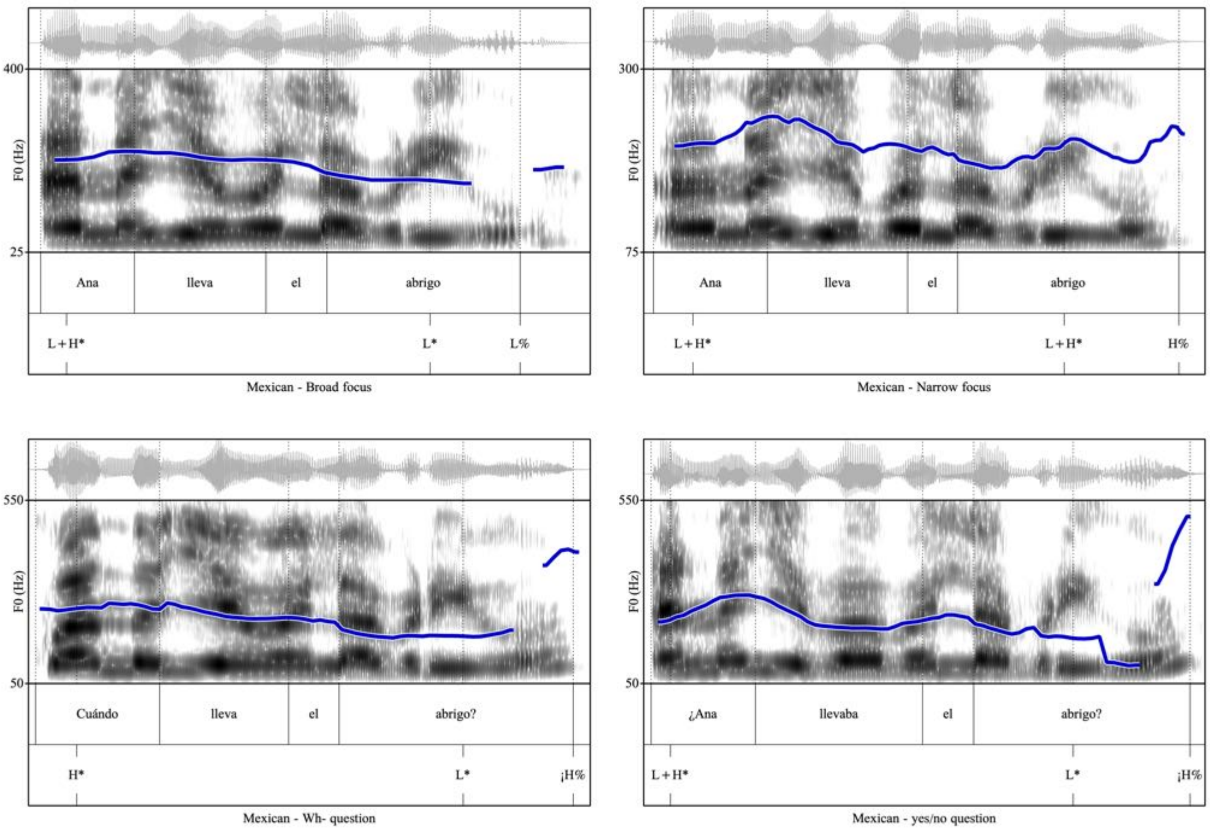

Figure 19. Waveform, spectrogram, and F0 trace exemplifying a broad focus (top left), narrow focus (top right), wh- question (bottom left), and yes/no question (bottom right) from a 30 year old female speaker of Mexican (Mexico City) Spanish.

The statement utterances from our speaker of Peruvian Spanish included a nuclear configuration of L* L%. The examples provided in the top row of Figure 20 are primarily differentiated by a prenuclear L+>H* in the broad focus condition and L+H* in the narrow focus condition. With regard to questions, this speaker often used a L+H* L% nuclear configuration for wh- questions, and L* ¡H% for yes-no questions. More information on Peruvian intonation is available in Sosa (1999) and O'Rourke (2005).

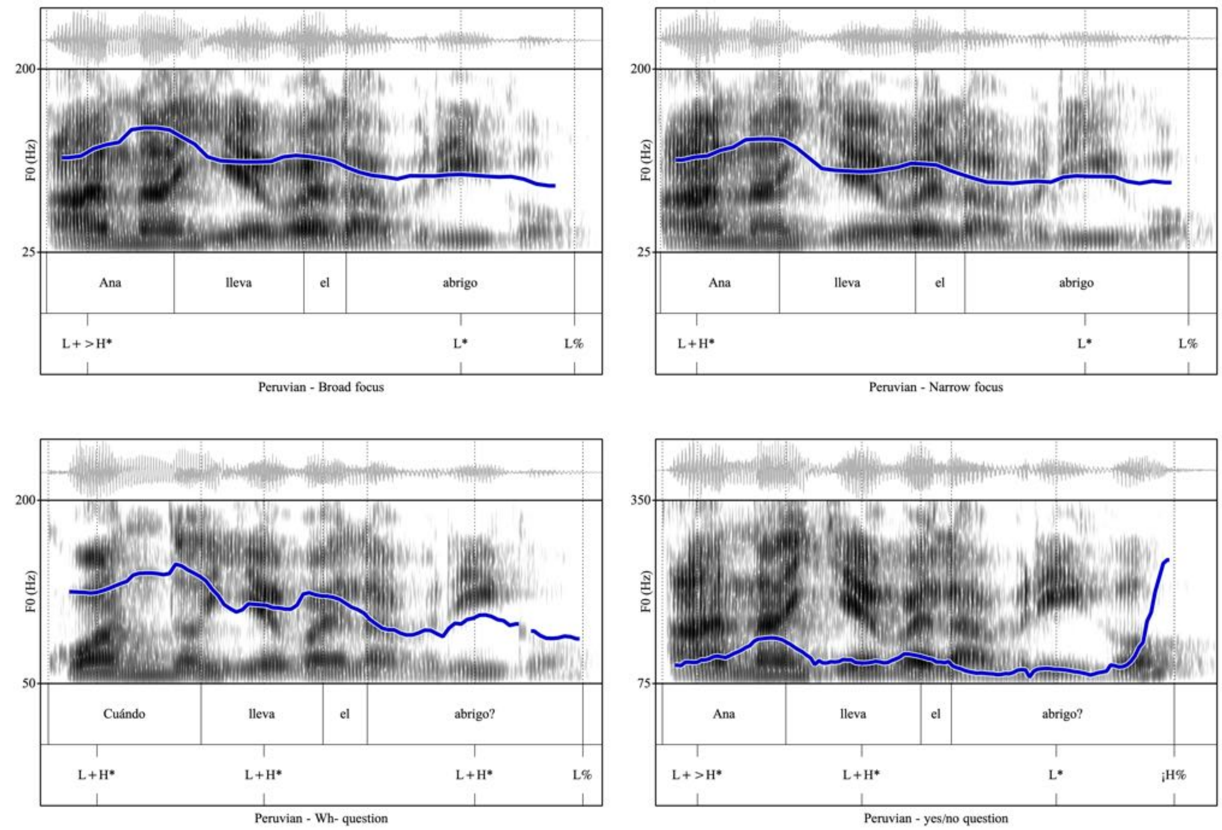

Figure 20. Waveform, spectrogram, and F0 trace exemplifying a broad focus (top left), narrow focus (top right), wh- question (bottom left), and yes/no question (bottom right) from a 30 year old male speaker of Peruvian (Lima) Spanish.

In Figure 21 we see example utterances from the Puerto Rican speaker. We typically only found a distinction in broad and narrow focus items in the prenuclear pitch accents (L+H* and H*, respectively). On occasion this speaker produced final rises for wh- questions (L* ¡H%), but also often used a falling contour. For yes/no questions, this speaker produced ¡H* L% configurations. More information on Puerto Rican intonation is available to the reader in Quilis (1981), Quilis (1987), Quilis (1993), Sosa (1999), Sosa (2003b), Armstrong (2010), and Armstrong (2012), among many others.

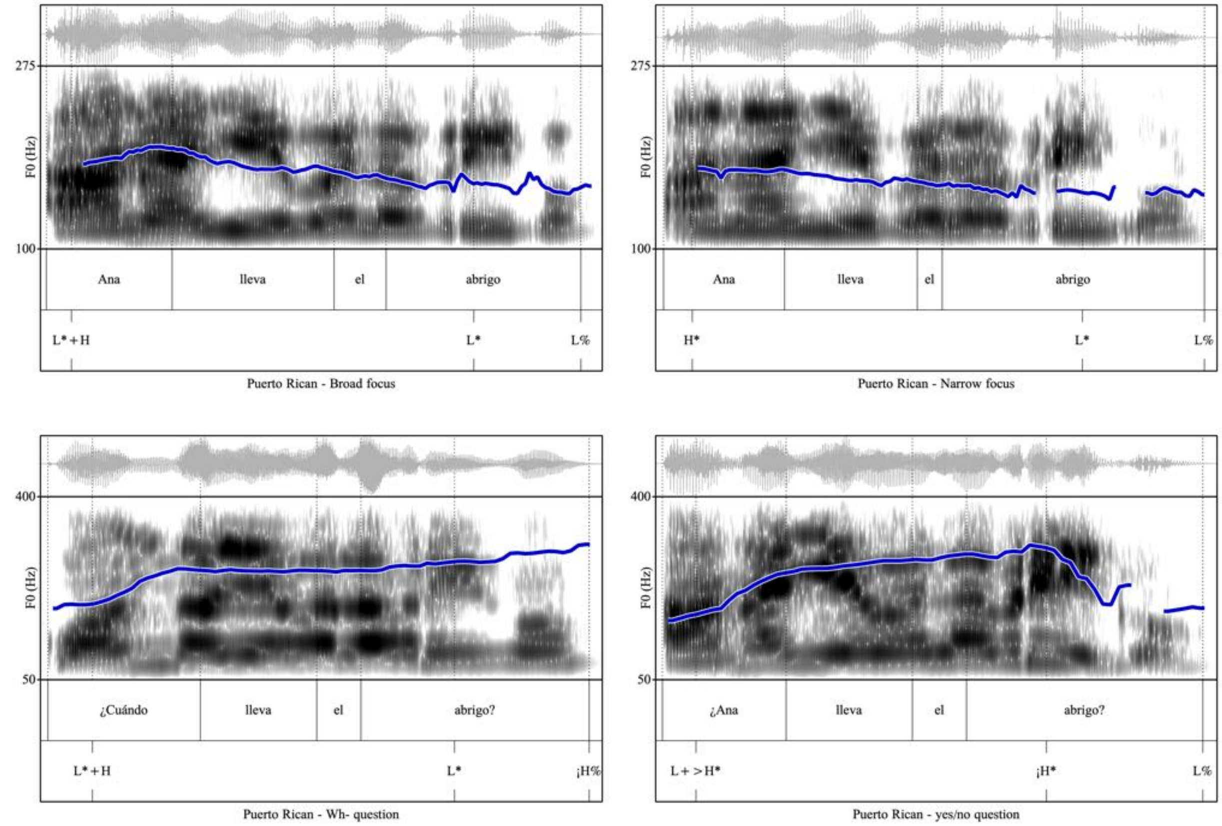

Figure 21. Waveform, spectrogram, and F0 trace exemplifying a broad focus (top left), narrow focus (top right), wh- question (bottom left), and yes/no question (bottom right) from a 35 year old female speaker of Puerto Rican (Ponce) Spanish.

In sum, we see that the acoustic stimuli used for the 2AFC task of the present project presents a large amount of between-variety and within-speaker variability. Most notably, our declarative utterances are often difficult to distinguish and we do not often see the typical L+H* L%) nuclear configuration that most commonly seen in the extant literature. We believe this is likely due to the fact that the manner in which we elicited narrow focus differs from other studies, which tend to use more specific pragmatic contexts (narrow focus correction, narrow focus contradiction), as opposed to a narrow focus utterance produced from answering an information seeking wh-question. That is, it may be the case that the use of a different pragmatic context and lab speech may result in contours different from those attested in other studies. This likely resulted in our narrow focus condition being easier than we initially intended. Our method of eliciting narrow focus follows that of Brandl et al. (2020), of which our study is a conceptual replication. We do not have access to their auditory stimuli, thus we cannot confirm which pitch contours were most common in their narrow focus condition. To illustrate the panhispanic and within-speaker variability of statements and questions, we plot the individual pitch contours from each utterance type by variety in Figure 22.

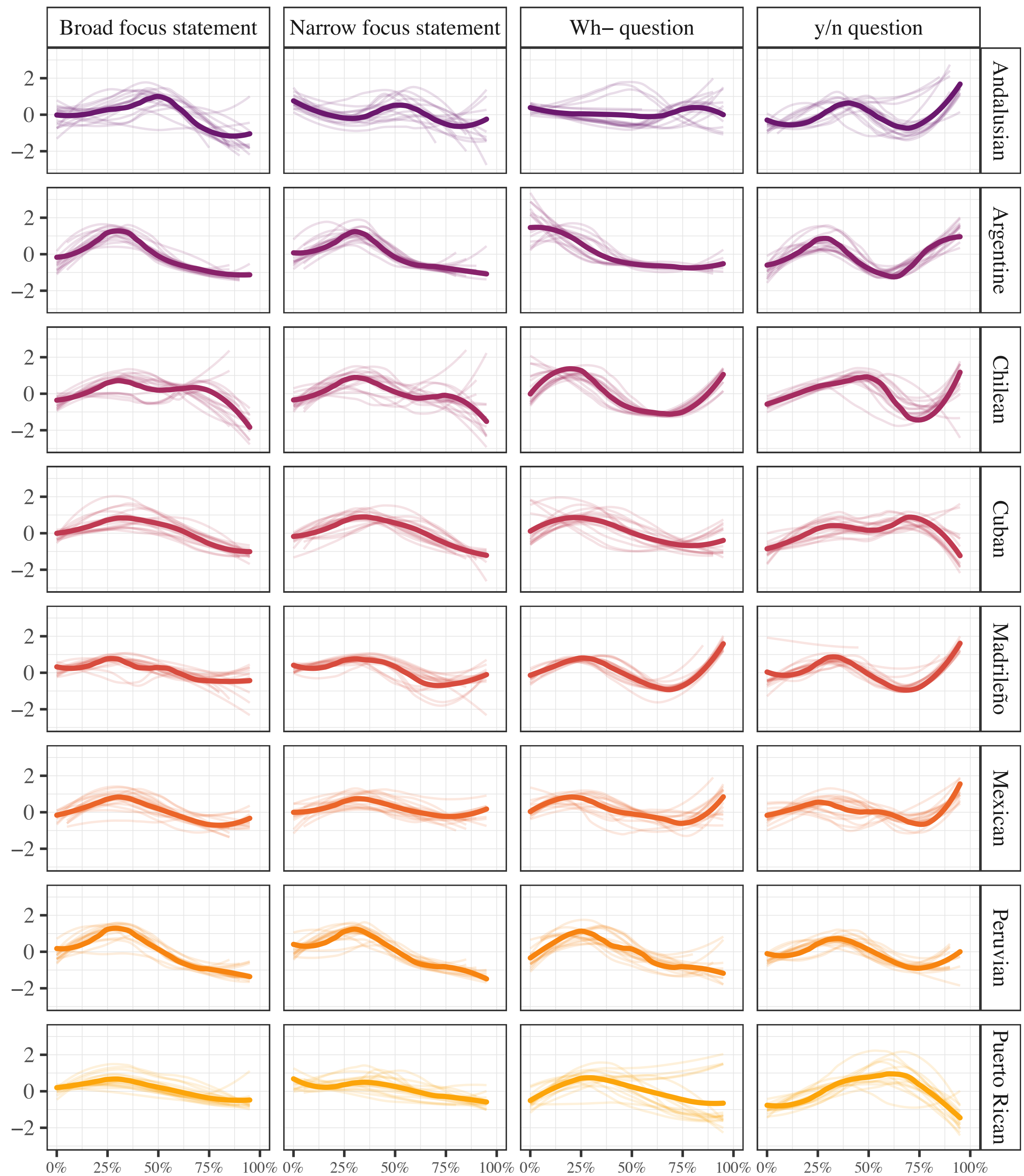

Figure 22. Pitch contours as function of time for utterance type and variety of Spanish. Thinner, opaque lines represent individual items, and thicker, dark lines represent the average contour trajectory. F0 values were log transformed and standardized for between speaker/variety comparability.

Randomization check across participants. For the purposes of our research questions, it was important that every participant be presented with stimuli from all of the Spanish varieties to which we had access. Recall that the 2AFC task contained 64 items, 16 of each utterance type. Using javascript we assigned each variety an equal probability of being selected in a given trial (0.125). To ensure that our randomization worked as planned (i.e., with each variety represented approximately equally across all trials and all participants), we calculated the average number of times each variety was presented in the data set (n = 225, and 14400 trials). One can observe in Figure 23 that this is indeed the case.

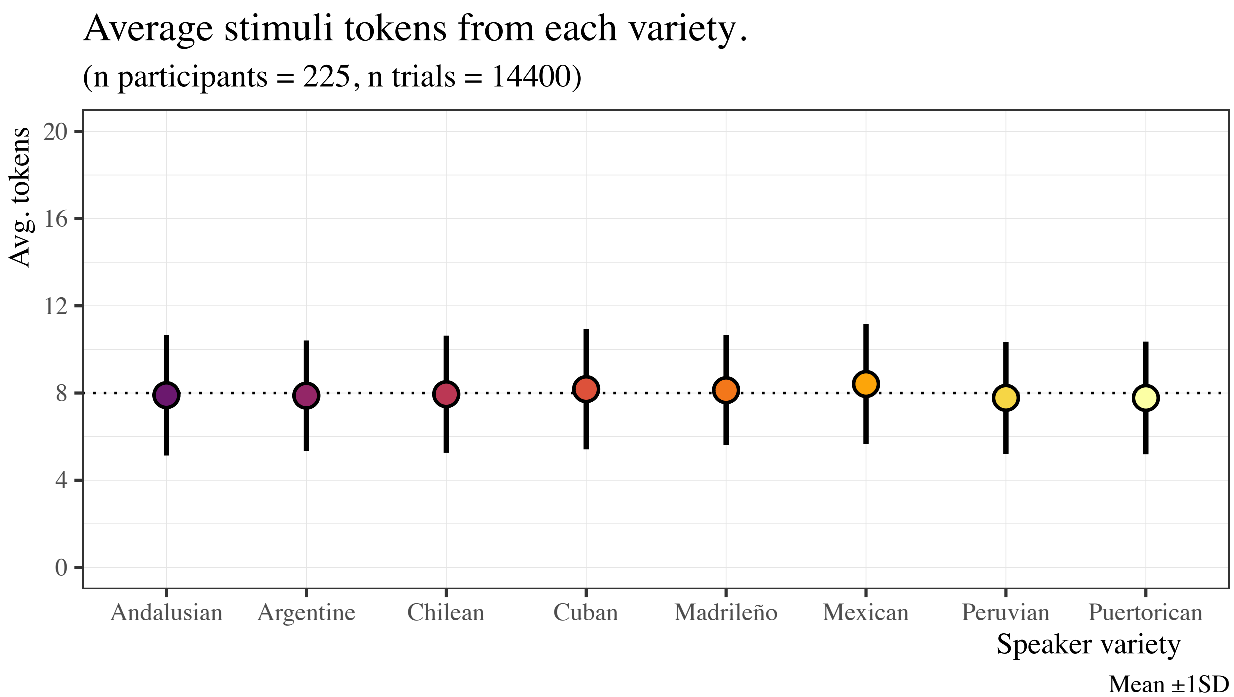

Figure 23. Average number of tokens (±1 SD) presented from each speaker variety across all 14,400 trials. The experiment was programmed such that each of the 8 varieties had an equal probability of being presented (12.50%) across 64 experimental trials.

Items. Table 13 provides a list of all of the target items used for each utterance type.

Table 13: Experimental items produced in auditory stimuli.

Utterance type	Item
Broad focus statement	Ana lleva el abrigo
	Daniel iba a Bolivia
	David leía el libro
	El bebe comía muy bien
	El hombre mira la luna
	El niño oye el río
	Emilio ama la marcha
	La maestra vive en Paris
	La niña lava el plato
	Manuela vende el carro
	María bebe el vino

Utterance type	Item
	Maríano habla del tiempo
	Marta abre el regalo
	Mi madre come la fruta
	Mi novio viene del lago
	Mi tía odia la lluvia
Narrow focus statement	(¿Qué lleva Ana?) Ana lleva el abrigo
	(¿A dónde iba Daniel?) Daniel iba a Bolivia
	(¿Qué leía David?) David leía el libro
	(¿Cómo comía el bebé?) El bebe comía muy bien
	(¿Qué mira el hombre?) El hombre mira la luna
	(¿Qué oye el niño?) El niño oye el río
	(¿Qué ama Emilio?) Emilio ama la marcha
	(¿Dónde vive la maestra?) La maestra vive en Paris
	(¿Qué lava la niña?) La niña lava el plato
	(¿Qué vende Manuela?) Manuela vende el carro
	(¿Qué bebe María?) María bebe el vino
	(¿De qué habla Mariano?) Maríano habla del tiempo
	(¿Qué abre Marta?) Marta abre el regalo
	(¿Qué come tu madre?) Mi madre come la fruta
	(¿De dónde viene tu novio?) Mi novio viene del lago
	(¿Qué odia tu tía?) Mi tía odia la lluvia
Wh- question	¿Cuándo bebía el vino?
	¿Cuándo comía la fruta?
	¿Cuándo lavaba el plato?
	¿Cuándo leía el libro?
	¿Cuándo lleva el abrigo?
	¿Cuándo miraba la luna?
	¿Cuándo vendia el carro?
	¿Por qué abre el regalo?
	¿Por qué ama la navidad?
	¿Por qué desayuna muy bien?
	¿Por qué hablaba del agua?
	¿Por qué iba a Bolivia?
	¿Por qué odiaba la lluvia?
	¿Por qué oia el río?
	¿Por qué venia del lago?
	¿Por qué vivia en Paris?
yes/no question	¿Ana lleva el abrigo?
	¿Daniel iba a Bolivia?
	¿David leía el libro?
	¿El bebe comía muy bien?
	¿El hombre mira la luna?
	¿El niño oye el río?
	¿Emilio ama la marcha?
	¿La maestra vive en Paris?
	¿La niña lava el plato?
	¿Manuela vende el carro?
	¿María bebe el vino?
	¿Maríano habla del tiempo?
	¿Marta abre el regalo?
	¿Mi madre come la fruta?

Utterance type	Item
	¿Mi novio viene del lago?
	¿Mi tía odia la lluvia?

Author contributions

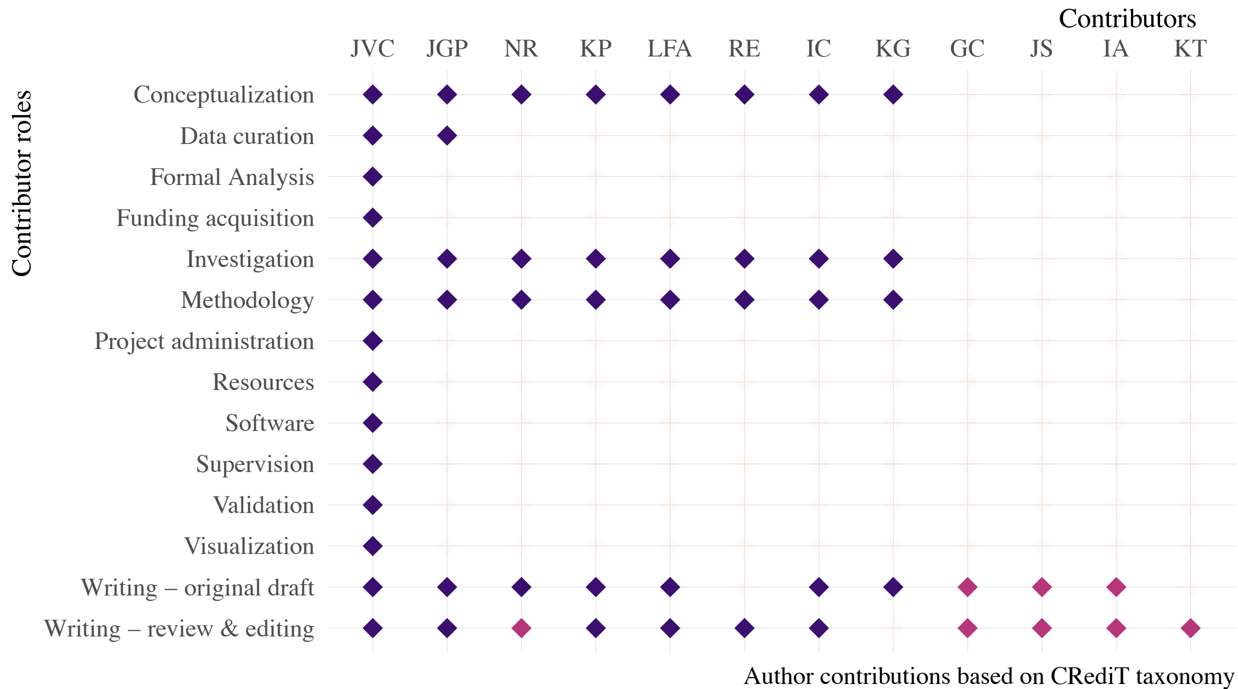

Figure 24. Author contributions according to the CREDiT author roles taxonomy. Contributions are indicated as being substantial (dark diamonds) or moderate (light diamonds).

Reproducibility information

About this document

This document was written in RMarkdown using `papaja` (Aust & Barth, 2018).

Session info

```
 setting  value
 version  R version 4.2.1 (2022-06-23)
 os       macOS Big Sur ... 10.16
 system   x86_64, darwin17.0
 ui       X11
 language (EN)
 collate  en_US.UTF-8
 ctype    en_US.UTF-8
 tz       America/New_York
 date     2023-02-17
 pandoc   2.19.2 @
/Applications/RStudio.app/Contents/Resources/app/quarto/bin/tools/ (via
rmarkdown)
               loadedversion       date
bayesplot             1.10.0 2022-11-16
bayestestR            0.13.0 2022-09-18
beeswarm               0.4.0 2021-06-01
bookdown                0.32 2023-01-17
brms                  2.18.0 2022-09-19
cmdstanr               0.5.3 2022-08-25
coda                  0.19-4 2020-09-30
contributoR            0.3.0 2022-10-12
devtools               2.4.5 2022-10-11
dplyr                  1.1.0 2023-01-29
effectsize             0.8.3 2023-01-28
emmeans              1.8.4-1 2023-01-17
forcats                1.0.0 2023-01-29
fs                     1.6.0 2023-01-23
ganttrify         0.0.0.9008 2022-10-12
ggbeeswarm             0.7.1 2022-12-16
ggdist                 3.2.1 2023-01-18
ggplot2                3.4.0 2022-11-04
ggrepel                0.9.3 2023-02-03
ggridges               0.5.4 2022-09-26
glue                   1.6.2 2022-02-24
here                   1.0.1 2020-12-13
insight               0.19.0 2023-01-30
knitr                   1.42 2023-01-25
loo                    2.5.1 2022-03-24
magrittr               2.0.3 2022-03-30
mvtnorm                1.1-3 2021-10-08
nlme                 3.1-162 2023-01-31
papaja                 0.1.1 2022-07-05
parameters            0.20.2 2023-01-27
patchwork              1.1.2 2022-08-19
```

```
posterior              1.3.1 2022-09-06
prettyunits            1.1.1 2020-01-24
printy            0.0.0.9003 2022-10-17
purrr                  1.0.1 2023-01-10
Rcpp                  1.0.10 2023-01-22
RcppParallel           5.1.6 2023-01-09
readr                  2.1.3 2022-10-01
readxl                 1.4.1 2022-08-17
remotes                2.4.2 2021-11-30
rlang                  1.0.6 2022-09-24
rmarkdown               2.20 2023-01-19
rprojroot              2.0.3 2022-04-02
rstan                 2.21.8 2023-01-17
rstantools             2.2.0 2022-04-08
scales                 1.2.1 2022-08-20
sessioninfo            1.2.2 2021-12-06
stringr                1.5.0 2022-12-02
tibble                 3.1.8 2022-07-22
tidybayes              3.0.3 2023-02-04
tidyr                  1.3.0 2023-01-24
```

References

Aust, F., & Barth, M. (2018). *papaja: Create APA manuscripts with R Markdown*. Retrieved from https://github.com/crsh/papaja

Beckman, M. E., Díaz-Campos, M., McGory, J. T., & Morgan, T. A. (2002). Intonation across Spanish in the Tones and Break Indices framework. *Probus*, 9–36. https://doi.org/10.1515/prbs.2002.008

Butragueño, P. M. (2003). Hacia una descripción prosódica de los marcadores discursivos. Datos del español de México. *La Tonía: Dimensiones Fonéticas y Fonológicas*, 375–402.

Butragueño, P. M. (2004). Configuraciones circunflejas en la entonación del español mexicano. *Revista de Filología Española*, *84*(2), 347–373.

Colantoni, L. (2011). *Broad-focus declaratives in Argentine Spanish contact and non-contact varieties*. John Benjamins.

Colantoni, L., & Gurlekian, J. (2004). Convergence and intonation: Historical evidence from Buenos Aires Spanish. *Bilingualism: Language and Cognition*, *7*(2), 107–119. https://doi.org/10.1017/S1366728904001488

Estebas-Vilaplana, E., & Prieto, P. (2010). Castilian Spanish intonation. In P. Prieto & P. Roseano (Eds.), *Transcription of intonation of the Spanish language* (pp. 17–48). Münich: Lincom Europa.

Face, T. L. (2003). Intonation in Spanish declaratives: Differences between lab speech and spontaneous speech. *Catalan Journal of Linguistics*, *2*, 115–131. https://doi.org/10.5565/rev/catjl.46

Face, T. L., & Prieto, P. (2007). Rising accents in Castilian Spanish: A revision of Sp_ToBI. *Journal of Portuguese Linguistics*, *6*, 117–146. https://doi.org/10.5334/jpl.147

Henriksen, N., & García-Amaya, L. J. (2012). Transcription of intonation of Jerezano Andalusian Spanish. *Estudios de Fonética Experimental*, *21*, 109–162.

Labastía, L. O. (2006). Prosodic prominence in Argentinian Spanish. *Journal of Pragmatics*, *38*(10), 1677–1705. https://doi.org/10.1016/j.pragma.2005.03.019

Labastía, L. O. (2011). Procedural encoding and tone choice in Buenos Aires Spanish. *Procedural Meaning: Problems and Perspectives, CRISPI*, *25*, 383–413.

O'Rourke, E. (2005). *Intonation and language contact: A case study of two varieties of Peruvian Spanish* (PhD thesis). University of Illinois at Urbana-Champaign.

Pierrehumbert, J. (1980). *The phonology and phonetics of English intonation* (PhD thesis). Massachusetts Institute of Technology.

Quilis, A. (1981). *Fonética acústica de la lengua española*. Madrid: Gredos.

Quilis, A. (1987). Entonación dialectal hispánica. In H. López Morales & M. Vaquero (Eds.), *Actas del I Congreso internacional sobre el español de América. San Juan, Puerto rico: Academia puertorriqueña de la lengua española* (pp. 117–163).

Quilis, A. (1993). *Tratado de fonología y fonética españolas*. Madrid: Gredos.

Sosa, J. M. (2003a). La notación tonal del español en el modelo Sp-ToBI. In P. Prieto (Ed.), *Teorías de la entonación* (pp. 185–208). Ariel.

Sosa, J. M. (2003b). Wh-questions in Spanish: Meanings and configuration variability. *Catalan Journal of Linguistics*, *2*, 229–247. https://doi.org/10.5565/rev/catjl.51

Vilaplana Estebas, E., & Prieto, P. (2008). La notación prosódica del español: Una revisión del Sp_ToBI. *Estudios de Fonética Experimental*, *17*, 264–283.

Xu, Y. (2010). In defense of lab speech. *Journal of Phonetics*, *38*(3), 329–336. https://doi.org/10.1016/j.wocn.2010.04.003