**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:

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

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’).[[1]](#footnote-1) 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:

The priors for the drift rate model were:

and the priors for the boundary separation model were:

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 ±66% and 95% credible intervals.

*Figure* *9.*  Exploratory analysis of d’ as a function of utterance type and speaker variety. Points represent posterior medians ±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:

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:

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] |

*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 10 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 8.

Figure 10.  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.

*Figure* *10.*  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 8: 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 11 plots the variety-matched raw accuracy scores as a function of utterance type and Table 9 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 11.  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.

*Figure* *11.*  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 9: 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] |

1. 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. [↑](#footnote-ref-1)