

## A LOOK AT MICROVARIATION IN THEORETICAL SYNTAX: DIALECT DISTANCE AND DATA ASSESSMENT

**1. Goal:** to inform the study of microvariation in Spanish by (i.) comparing Chilean, Venezuelan and Puerto Rican Spanish in light of theoretical syntax; (ii.) evaluating the reliability of the theoretical syntax data for Spanish in these dialects; (iii.) studying the treatment of microvariation in the current theoretical literature.

**2. Background:** Traditional dialectological work has put a limited emphasis on syntax, whereas the theoretical literature has limited empirical basis in theoretical work, as compared to the comprehensive linguistic data samples available in dialectology (e.g., Craenenboeck 2016). The current research reports on an ongoing research program intended to bridge the gap between dialectology and theoretical syntax while addressing these issues.

**3. Methodology:** A random sample of 100 sentences (50 minimal pairs) from Spanish syntax papers published in *Probus: International Journal of Romance Linguistics* in the 2010-2016 period, complemented with 6 sentences from the Spanish syntax literature other than *Probus* (16 articles total) were studied in Chilean, Puerto Rican and Venezuelan Spanish, while controlling for differences in verbal morphology and the lexicon (e.g., (1)). 40 speakers of each dialect evaluated the sentences using a 7-point Likert Scale. The following is a partial list of topics covered: double-complementizer constructions, the syntax of object clitics and Differential Object Marking, null objects, spurious SE, wh-in situ, comitative constructions, clitic duplication, CLLD, degree operators, infinitival subjects and VS-inversion in wh-questions.

**4. Results:** Data were analyzed using Mixed ANOVAS. The predicted direction of the grammaticality scale (mean grammaticality judgment for each minimal pair showing a contrast in the direction claimed in the literature) shows a convergence rate of 76% among the three dialects (where 52% out of the 76% reaches statistical significance). Still another 13% of the results correspond to cases where the three dialects converge though the directionality of the grammaticality contrast is opposite what the theoretical literature predicts (where 9% out of the 13% reaches statistical significance). The three dialects converge in 89% of the data (irrespective of whether the directionality of the prediction is fulfilled or whether the results reach statistical significance). A discussion of shared data points and those that are not is included below.

**5. Further benefits of the design:** By matching the dialect of the data with the dialect of the participants, the design lends itself to a data assessment component focusing on the convergence rate of the data published in the theoretical literature (collected without formal data collection protocols and not subject to statistical scrutiny) and experimentally gathered data (see Sprouse et al. 2013 for an overview of the underlying debate). The results are included in Table 1, where the descriptive directionality in the predicted direction of the grammaticality scale is 80%, 91%, and 92% for Chile, Puerto Rico, and Venezuela, respectively. The percentage of statistically significant results in the predicted direction also reveals a non-trivial convergence rate between the theoretical literature and data gathered experimentally in this work, thus providing support for the reliability of theoretical syntax data even for languages subject to less scrutiny than English, contra Haider (2016) – a significant result given the relevance of scientific replication. A discussion of replicated data points and those that were not is included as well. Furthermore, the results suggest that the theoretical literature is less representative of Chilean

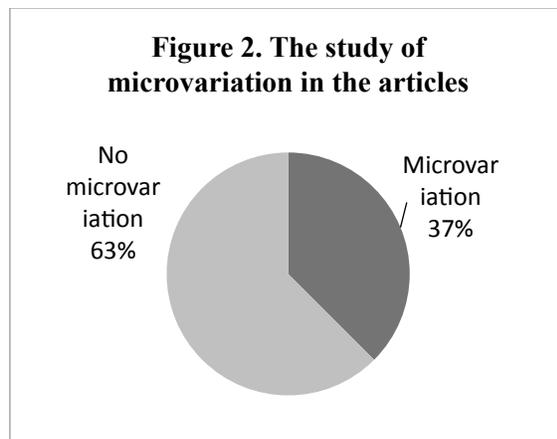
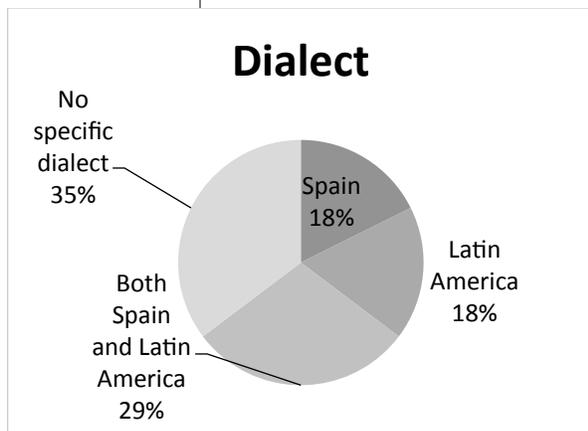
**A LOOK AT MICROVARIATION IN THEORETICAL SYNTAX:  
DIALECT DISTANCE AND DATA ASSESSMENT**

Spanish than of the rest of the dialects, hinting at the need for future research on this dialect. Finally, this research unveils current tendencies in the literature by analyzing the presence of each dialect in the corpus, see Figure 1, as well as the percentage of articles that focus on microvariation (37%) as opposed to dealing with one single variety of Spanish (Figure 2).

- (1) a. Juana leyó durante una hora. (Armstrong 2016)  
       Juana read for an hour  
    b. \*Juana leyó en una hora.  
       Juana read in an hour

**Table 2. Convergence rate between data published in the theoretical literature and experimentally gathered data**

Dialect	Descriptive directionality in the predicted direction	Significant in the predicted direction	Non-significant in the predicted direction	Significant in the opposite direction	Non-significant in the opposite direction	N (minimal pairs)
<b>Chile</b>	80%	69%	11%	10%	10%	41
<b>Puerto</b>	91%	79%	12%	6%	3%	37
<b>Venezuela</b>	92%	76%	16%	5%	3%	40
<b>Total sample</b>	88%	75%	13%	7%	5%	118



**REFERENCES:** Armstrong, G. 2016. "Spanish unspecified objects as null incorporated nouns". *Probus* 28: 165–229. Craenenboeck, J van. 2016. "Quantity and quality in linguistics: reverse dialectometry". Ms. Haider, H. 2016. *Incredible syntax – between Cognitive Science and imposture*. Book ms. Sprouse, J., C. Schütze & D. Almeida. 2013. "Assessing the reliability of journal data in syntax: Linguistic Inquiry 2001-2010". *Lingua* 134, 219-248.