Habían profesores: A usage-based analysis of the pluralization of *haber* in Spanish

Manuel Díaz-Campos (*Indiana University*), Mark Hoff (*The Ohio State University*),
& Rosa María Piqueres Gilabert (*Indiana University*)

The pluralization of *haber* is a commonly documented phenomenon in many varieties of Spanish. Several prescriptive grammars considered *haber* a transitive verb and the corresponding noun phrase (NP) is described as its direct object. Descriptive work and sociolinguistic perspectives on *haber* have argued that speakers use subject-verb agreement because the noun phrase is reanalyzed as the subject of the clause (e.g. Bentivoglio & Sedano 1989). More recently, Brown and Rivas (2012) present an analysis in which noun phrases used probabilistically more in the role of subject (a measure known as grammatical relation probability) favor *haber* pluralization. The present study takes Brown and Rivas' (2012) analysis as a point of departure to investigate the pluralization of *haber* in an oral Spanish corpus. Building on Erker and Guy (2012), this study contributes to provide independent evidence regarding the role of experience with linguistic structures and the frequency threshold that needs to be reached to find statistically significant effects.

One hundred sixty speakers were selected from the corpus *Estudio Sociolingüístico de Caracas*, 'Sociolinguistic Study of Caracas' (Bentivoglio & Sedano 1987). The speech samples from this corpus are conversations in a spontaneous style recorded using standard sociolinguistic interview techniques. The participants are divided according to age (19-29 years old, 30-45 years old, 46-60 years old, and 61+ years old), socioeconomic class (upper, middle, and lower class), and gender (female and male). All cases of *haber* (3,292 tokens) in all its grammatical persons and tenses were extracted. From this data, 689 tokens were of presentational *haber* + plural NP. The relative frequency as subjects of all plural NPs with pluralized *haber* was calculated by searching their occurrence in the whole corpus. Using this method, nearly 46,000 tokens were extracted. These NPs were divided into three frequency groups: low (1-10 occurrences), medium (11-100), and high (101+).

The findings of the statistical analysis reveal that the model proposed in Brown and Rivas (2012) is generalizable to a limited extent to Bentivoglio and Sedano's (1987) Venezuelan corpus. That is, grammatical relation probability has statistically significant predictive power related to the pluralization of presentative *haber* only among the most frequent NPs (i.e. more than 100 occurrences). For NPs of lower frequency, percentage of use as a subject does not have predictive power for *haber* pluralization. These results suggest that a frequency threshold exists for the Subject Model and for connections in cognitive representation to take effect as proposed by Erker and Guy (2012).

References

Bentivoglio, Paola and Mercedes Sedano. 1987. *Estudio sociolingüístico del habla de Caracas*. Caracas: Universidad Central de Venezuela.

Bentivoglio, Paola and Mercedes Sedano. 1989. 'Haber': ¿Un verbo impersonal?'. In *Estudios sobre español de América y lingüística afroamericana. Ponencias presentadas en el 45 Congreso Internacional de Americanistas*, 59-81. Bogotá: Instituto Caro y Cuervo.

Brown, Esther and Javier Rivas. 2012. Grammatical relation probability: How usage patterns shape analogy. *Language Variation and Change* 24(3), 317-341

Erker, Daniel and Gregory Guy. 2012. The role of lexical frequency in syntactic variability: Variable subject personal pronoun expression in Spanish. *Language* 88(3), 526-555