Mixed-Effects Design and Power for Generalizability of Results: Copula + Adjective in the Spanish variety of Limon, Costa Rica

Jorge Aguilar-Sanchez, University of Dayton

Manuel Diaz-Campos, Indiana University

A typical concern in sociolinguistic studies relates to the design of a corpus and how these samples are representative of the speech community under study. In order to take into account some important considerations, the present study examines issues relate to the sample size and research design that improve the generalizability of results. For this purpose, we take the study of copula choice in contact Spanish as an example case. In recent years, we have seen a move from mono-level analysis of variance (Tagliamonte, 2006) to Multi-level analysis of variance (Johnson, 2009, 2014). These methods have been adopted to explain different linguistic phenomena. However, little has been done to address the issue of generalizability of results following Silva-Corvalán's (2001) suggestion that sample size planning should be based on theoretical and practical issues. We follow what is recommended in the field of statistics for the social sciences for the determination of a proper sample size and power of a study. In order to achieve this goal, we present the results of power calculations for the study on copula choice + adjective in the Spanish varieties of Limón, Costa Rica. We present a series of possible outcomes resulting from different research design decisions while following Author 1's design and suggestions (2009, 2012, 2014) and Author 1 & Kelly (forthcoming) to achieve generalizability of our results. We analyze, through a power analytic method, different models (DV + IVs) for the study of copula choice in this particular geographical context using a mixed-effect method and simulated data. For each model, we calculate sample size (speakers and tokens), power (% of generalizability), and effect size (strength of the phenomenon in the population). We compare the different parameters and their outcomes and how they affect the possible statistical results. Results of the present study present an overview of how important the a-priory exploration of the phenomenon is and they contribute to the design of future studies and the advancement of knowledge in this particular syntactic variable and other areas of linguistic inquiry. The analyses show that statistical results of a study, whether they are significant or not significant, are very sensitive to its design. When an independent variable is added or subtracted, the power of the study (generalizability of the results) changes and so does the need for more or less speakers or tokens (sample size). Design decisions made during early stages of a study may yield under- or over-power studies. These particularities will trigger time and economic expenses that sometimes are unnecessary. Our contribution is to address these particularities of research design in the context of linguistic variation and change to further strengthen our field with best practices that will help us build new and representative corpora that would, in turn, yield a more cohesive and comprehensive body of knowledge.

Work Cited

Aguilar-Sanchez, J. (2009). Syntactic Variation: The case of copula choice in the Spanish of Limón, Costa Rica. (Doctor), Indiana University, Bloomington, IN.

Aguilar-Sanchez, J. (2012). Formal Instruction and Language Contact in Language Variation: the Case of Ser and Estar + Adjective in the Spanishes of Limón, Costa Rica. In K. Geeslin and M. Díaz-Campos (eds), Selected Proceedings of the 14th Hispanic LinguisticSymposium (pp. 9-25). Sumerville, MA: Cascadilla Proceedings Project.

Aguilar-Sanchez, J. (2014). Replicability of (Socio)Linguistics studies. JRDS *1(1) pp. 1-23*. doi:http://www.equinoxpub.com/journals/index.php/JRDS/article/view/17428

Aguilar-Sanchez, J. & Kelly, K. (forthcoming). Dichotomous Variables in Repeated Measures Designs: A Primer on Multilevel Logistic Regression.

Johnson, D. E. (2009). Getting off the GoldVarb Standard: Introducing Rbrul for Mixed-Effects Variable Rule Analysis. Language and Linguistics Compass 3 (1): 350–383. http://dx.doi.org/10.1111/j.1749-818X.2008.00108.x

Johnson, D. E. (2014) Progress in regression: Why sociolinguistic data calls for mixed-effects models. Self published manuscript. http://www.danielezrajohnson.com/johnson_2014.pdf

Silva-Corvalán, C. (2001). *Sociolingüística y pragmática del español*. Washington, DC: Georgetown University Press.

Tagliamonte, S. A. (2006). *Analysing Sociolinguistic Variation*. Cambridge: Cambridge University Press.