Douglas Biber, Mark Davies, James K. Jones, Nicole Tracy
Northern Arizona University and Brigham Young University
Douglas.Biber@nau.edu

Register variation in Spanish: A multi-dimensional comparison with English, Somali, and Korean

Multi-dimensional (MD) analysis is a methodological approach that applies multivariate statistical techniques (especially factor analysis) to the investigation of register variation in a language. The approach was originally developed to analyze the range of spoken and written registers in English (e.g., Biber 1986, 1988), while later studies have applied the same approach to Somali, Korean, and Tuvaluan (see Biber 1995). Interestingly, these studies have found some striking similarities in the underlying 'dimensions' that distinguish among registers in these diverse languages, raising the possibility of universal patterns of register variation.

To date, no comprehensive analysis of register variation in a European language other than English has been completed. Spanish provides an ideal test case for investigating the extent to which register variation patterns in similar ways across languages: Spanish is a major international language with a long social history of literacy which has been influenced by the role of an official language academy; it is a Romance language, with interesting linguistic similarities and differences to English; and there are important differences in the preferred conversational styles and rhetorical priorities in writing.

The present study begins to fill that gap by undertaking a comprehensive MD analysis of Spanish, based on analysis of a register-diversified corpus of Spanish (18.2 million words; 4049 texts; 19 registers). Six primary dimensions of variation are identified and interpreted in linguistic and functional terms. Some of these dimensions are specialized, without obvious counterparts in the MD analyses of other languages (e.g., a dimension related to discourse with a counterfactual focus). However, other Spanish dimensions correspond closely to dimensions identified for other languages, reflecting functional considerations such as interactiveness, personal stance, informational density, argumentation, and a narrative focus.