

The HISPACAT comparative database of syntactic constructions and its applications to
syntactic variation research

Xavier Villalba

Centre de Lingüística Teòrica

Universitat Autònoma de Barcelona

Spain

Xavier.Villalba@uab.cat

In this communication the HISPACAT database of syntactic contrastive constructions in Catalan and Spanish will be presented. HISPACAT actually counts with around 500 files, each containing a construction interesting from a comparative viewpoint, and it is designed as a dynamic comparative grammar of two closely related languages, which, from a theoretical point of view, offers us a privileged standpoint to advance in finding the “atoms” of linguistic microvariation (Baker 2001), and offering a snapshot of microparametric (in)variance (Kayne 1996, 2005), which will help us to predict which parts of the grammar are less stable, and hence more sensitive to syntactic change or attrition and interference phenomena (Silva-Corbalán 1994, Sorace 1998). Hence, this tool, which is conceived as a major empirical source for testing syntactic microvariation, may also prove helpful for researchers in bilingualism and language contact studies.

The first part of the talk will be devoted to discuss the goals of the project, stressing the theoretical and methodological differences existing between our enterprise and current textual databases. We will stress that HISPACAT is theory-driven, so that it assumes (micro)parameters (Baker 2001, Kayne 1996, 2005), and treats constructions as sets of properties (Chomsky 1981; cf. Goldberg 1995). As a direct consequence, HISPACAT is conceptually-based, for the interest is in linguistic concepts not in expressions themselves, in sharp contrast with textual databases. Yet, besides its theoretical goals, HISPACAT is a useful tool for linguists, for it offers a major relational database for researchers in bilingualism and L2 learning, and a comparative grammar and a catalog of explained examples for L2 teachers/learners.

In the second part of the talk, a general presentation will be offered of the main architecture of the database, namely its conceptual ontology, and the internal structure of the files. As for the ontology, it is build on 175 concepts grouped into two major categories, properties and relations, which are subdivided in lexical-grammatical and semantic-pragmatic properties, and syntactic and semantic relations, respectively.

Then, we will have a quick glance at the internal structure of the files of the database, with a special stress on the EXAMPLES and CONCEPTS fields, for, besides offering an exhaustive description of the basic features of the construction, they will prove instrumental for allowing multimodal information queries, particularly in the development of empirical generalizations.

Finally, we will close this second part of the talk taking a closer insight into the main technical features of the relational database, focusing on the different query systems, namely boolean searches on the concepts of the ontology, boolean searches on the ANALYSIS and EXAMPLES fields, and a novel search mode based on the ontology tree.

The last part of the talk will be devoted to analyze several cases related to microvariation, and attrition/interference phenomena, where HISPACAT has proved to be a reliable source for testing hypothesis, and establishing robust linguistic generalizations. For instance, it will be shown that the comparative and conceptual-based nature of HISPACAT has been capable of describing the unattested big differences in productivity between Catalan and Spanish regarding right-dislocation. In order to test the accuracy of the resultant generalization, a comparison will be conducted with an independent corpus-based study (see Villalba 2007). Eventually, the essence of the description in HISPACAT finds a parallel in the corpus-based study, which confirmed the reliability of the dynamic comparative method.

Baker, M. (2001). *The atoms of language*. New York: Basic Books.

Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Foris.

Farrar, S., and Langedoen, T. (2003). A linguistic ontology for the semantic web. *Glott International*, 7, 97–100.

Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.

Kayne, R. S. (1996). Microparametric syntax. some introductory remarks. J. Black and V. Montapanyane (Eds.), *Microparametric syntax and dialectal variation* (p. ix-xvii). Amsterdam/Filadelfia: John Benjamins.

Kayne, R. S. (2000). *Parameters and universals*. New York: Oxford University Press.

Kayne, R. S. (2005). Some notes on comparative syntax, with special reference to english and french. G. Cinque and R. S. Kayne (Eds.), *Handbook of comparative syntax* (p. 3-69). New York: Oxford University Press.

Miller, G. (1995). Wordnet: A lexical database for english. *Communications of ACM*, 2, 39–41.

- Silva-Corbalán, C. (1994). *Language contact and change: Spanish in los Ángeles*. Oxford: Oxford University Press.
- Sorace, A. (1998). *Near-nativeness, optionality and L1 attrition*. (Paper presented at 12th International Symposium of Theoretical and Applied Linguistics, Thessaloniki)
- Villalba, X. (2007). La dislocació a la dreta en català i castellà: microvariació en la interfície sintaxi/pragmàtica. *Caplletra* 42: 273-302.