

A database of syntactic contrastive contexts and constructions in Catalan and Spanish

Xavier Villalba

Center for Linguistic Research, Universitat Autònoma de Barcelona

Xavier.Villalba@uab.cat

1 Introduction

The database of syntactic constructions in Catalan and Spanish (HISPACAT) forms part of a bigger funded research project of the Center for Linguistic Research of the Universitat Autònoma de Barcelona, which involves 16 senior and 10 junior researchers. HISPACAT is conceived as a major empirical source for detecting syntactic microvariation (Kayne, 1996), and testing hypotheses regarding bilingualism, interference and attrition, but which may prove helpful for L2 learning, for it is designed as a dynamic comparative grammar of two closely related languages capable of offering a snapshot of basic (in)variance patterns. HISPACAT is, thus, a theory-oriented project, but with clearly-defined applied goals.

2 Design

HISPACAT fills a gap in the range of available linguistic databases, which are typically text-based, and monolingual. Instead, HISPACAT is a relational concept-oriented DB running under ORACLE, and it is based on comparative data. Its relational and concept oriented character reflects our interest in the grammatical concepts underlying the linguistic utterances, not in the utterances themselves. Furthermore, the design ensures a robust data-retrieval system based on a complex ontology of 176 linguistic concepts divided into properties and relations, and subdivided in turn into lexical-grammatical and semantic-pragmatic properties and syntactic and semantic relations. Thus, each construction in the DB –actually, HISPACAT counts 476 files– is decomposed in a series of primitive grammatical concepts that will serve as pointers for searches (hence, we don't consider constructions as primitives à la Goldberg, 1995, but rather as convenient descriptive devices arising from the interaction of simpler properties as in Chomsky's, 1981 classical formulation).

HISPACAT is also innovative in presenting contrastive and negative data – obviously lacking in textual DB–, to show both the symmetries and the asymmetries in the grammar systems of Catalan and Spanish, helping both the theorist *and* the learner to set the limits of the construction. For instance, in the file corresponding to *negación expletiva en el contexto de marcadores comparativos* ('expletive negation in the context of comparative markers'), we include (and comment) examples like: *Juan era antes más simpático que (no) ahora* ('Juan was more friendly then than now'); *Juan era antes más simpático de lo que (*no) es ahora* ('Juan was more friendly then than now'). This, in spite of its BD form, HISPACAT is a true dynamic comparative grammar, and attempts to fill the gap between the two major descriptive collective works Bosque and Demonte (1999) and Solà et al. (2002).

3 Architecture: The Conceptual Ontology

The conceptual ontology of HISPACAT is build on 175 concepts grouped into two major categories, properties and relations, which are subdivided in turn into lexical-grammatical and semantic-pragmatic properties, and into syntactic and semantic relations, respectively. As the GOLD linguistic ontology (Farrar and Langendoen 2003), our ontology is designed to allow a detailed description of constructions –it is instrumental in obtaining a robust data retrieval system–, but, more important, to permit reasoning about linguistic structure, and the building of far-reaching linguistic generalizations.

4 Applications

Even though HISPACAT is a theory-oriented project aimed at detecting the atoms of syntactic microvariation, it fulfills a series of applied goals as well. First of all, it proves useful for the language teacher, which can obtain in seconds a full set of commented examples grouped conceptually. Let us see an example. Suppose you are interested in the way locative prepositional pronouns are realized, and search for the (Spanish) keywords "locativo" ('locative') and "pronombre" ('pronoun'). You will be given the following file (irrelevant fields are omitted and the analysis field is translated for convenience):

```
NUM: 9
DEN_CAT: possessiu amb funció de complement locatiu
DEN_CAST: posesivo con función de complemento locativo
EX_CAT: Era sota meu
REF_CAT: GCC 7.5.2.8
EX_CAST:
REF_CAST:
ANÁL: The possessive locative pronoun is only possible in Catalan, and it surfaces as the complement of simple and complex locative prepositions replacing a genitive PP (cf. prop
```

teu/prop de tu). This construction sounds odd with inanimate referents (cf. *sota de la taula* '??sota seu').

Note how the analysis field includes negative evidence to show the lexical restrictions involved.

Consider now a case of syntactic interference stemming from the comparative treatment of the data:

NUM: S26-1

DEN_CAT: quantificador *cada* en expressions temporals aparentment no distributives

DEN_CAST: cuantificador *cada* en expresiones temporales aparentemente no distributivas

EX_CAT: Cada diumenge vaig al cinema

REF_CAT: Solà i Pujols (2002-26.1.2.3:2896)

EX_CAST:

REF_CAST:

ANÀL: [...] With temporal expressions, the distributive use is clear in *Cada dia va a escola a una hora diferent* 'Every day (s)he goes to school at a different time', but not in *Cada dia va a escola* '(S)he goes to school everyday', where Spanish resorts to the nondistributive quantifier (*Todos los días va a la escuela*). [...]

COMENT: Indeed, maybe Spanish allows such a construction (it is clearly found in the Spanish spoken by Catalan speakers).

Here one finds the description of a Catalan interference on Spanish that would have gone unnoticed unless the strong comparative spirit of HISPACAT.

Finally, consider a complex search case. You are looking for examples involving licensing of negative polarity items. You have the option to make a conceptual search for negation, which will return 37 files, or may search for a certain word—say *ningú* 'nobody'—just on the two fields devoted to examples, namely HISPACAT can be consulted as a textual DB.

References

- Bosque, I., and V. Demonte eds. (1999). *Gramática descriptiva de la lengua española*. Espasa, Madrid, Spain.
- Chomsky, N. (1981). *Lectures on Government and Binding*. Foris Publications, Dordrecht, Holland.
- Farrar, S., and T. Langendoen (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*. University of Chicago Press, Chicago.
- Kayne, R.S. 1996. *Microparametric Syntax. Some Introductory Remarks*. In *Microparametric Syntax and Dialectal Variation*, eds. J.R. Black and V. Montapanyane, ix-xvii. Amsterdam/Philadelphia: John Benjamins.
- Solà, Joan, Lloret, M. Rosa, Mascaró, Joan, and Pérez Saldanya, Manuel eds. 2002. *Gramàtica del català contemporani*. Barcelona: Empúries.