

## Dialect syntax beyond a corpus: A case study in European Portuguese

CARRILHO ERNESTINA  
FLUL/CLUL

SANDRA PEREIRA  
CLUL

Studies in Portuguese dialect syntax have been mainly driven by corpora-based data. The compilation of an annotated dialect corpus (CORDIAL-SIN *Syntax-Oriented Corpus of Portuguese Dialects*) has provided optimal access to precise morphological and syntactic information on spoken data of Portuguese regional varieties. CORDIAL-SIN compiles excerpts of spontaneous and semi-directed speech drawn from traditional geolinguistic fieldwork conducted in the Portuguese territory (materials from the Dialectal Sound Archive at CLUL). Accordingly, this dialectal resource has not only opened new empirical grounds to comparative syntax, but also enabled the development of new approaches to geolinguistic inquiry, brought about by the consideration of syntactic areas. To this respect, a major contribution has been the acknowledgment of the spatial distribution of morphosyntactic variants and their possible matching to well-defined dialectal areas (Carrilho & Pereira 2009, a.o.).

However, geolinguistic approaches to dialect syntax are to an important extent limited by the insufficiencies that corpus-based syntactic studies may raise. In this presentation, we discuss the limitations of CORDIAL-SIN as regards the geolinguistic representation of the nonstandard Portuguese construction in (1a):

- (1) a. Nunca se            tínhamos visto. (*MTM, CORDIAL-SIN*)  
      never RECIPR<sub>3</sub> had<sub>1PL</sub> seen
- b. Nunca nos            tínhamos visto. (*Standard European Portuguese*)  
      never RECIPR<sub>3PL</sub> had<sub>1PL</sub> seen
- ‘We had never seen each other (before)’

In the vein of recent methodological developments in dialect syntax (Barbiers 2005, Vangsnes & Johannessen 2011, a.o.), we propose a combined approach to dialect syntax data collection (in particular for the issues in connection with (1)), which couples together traditional geolinguistic concerns and advances in comparative syntax, corpus-driven and new methods of linguistic data collection.