

## **How two Catalan dialects coexist in the same areas. An analysis of the use and distribution of some lexical forms in *twitter***

MARIA-PILAR PEREA  
*Universitat de Barcelona*

ANTONIO RUIZ-TINOCO  
*Sophia University*

The aim of this presentation is to analyze some lexical forms used in two dialects that coexist in the same Catalan speaking areas. During a few weeks we have collected a wide-ranging amount of data from *Twitter* utilizing its Streaming API ver. 1.1. Tweets are aggregated in a MySQL database for easy posterior analysis.

We also collected publicly available profiles such as location, sex, etc. from the *tweeters*. A small quantity of data are geotagged, allowing us to create maps based on exact geographical origin, using tools such as Google's Fusion Tables. We used this methodology to:

- a) analyze the survival of some lexical forms that are characteristic of Catalan-speaking areas with two dialects: Valencian (*vesprada* 'afternoon', *espill* 'mirror', *matalaf* 'mattress', etc.) and Balearic (*moix* 'cat', *nigul* 'cloud', *llinatge* 'surname', etc.), some of which are shared by the two dialects (*cementeri* 'cemetery'), and
- b) compare the results of these areas with the corresponding standard forms (*tarda* 'afternoon', *mirall* 'mirror', *matalàs* 'mattress', *gat* 'cat', *núvol* 'cloud', *cognom* 'surname', *cementiri* 'cemetery').

The final objective is to quantitatively assess the incidence of standard Catalan, the degree of usage in each area, and the effects of some sociolinguistic variables such as rural/urban locations, sex, and age.

Although data obtained from Twitter does not contain all of the attributes of its authors and is not fully representative of the studied areas, its analysis provides relevant quantitative and qualitative aspects of the distribution of lexical forms. We also consider the social network *Twitter* as a valuable corpus for the study of language variation.