

Tracking linguistic and extra-linguistic features underlying lexical variation patterns: A case study on Tuscan dialects

SIMONETTA MONTEMAGNI
ILC-CNR

MARTIJN WIELING
University of Groningen

The hierarchical bipartite spectral graph partitioning (Dhillon, 2001) is a promising method used in the dialectometric literature to cluster dialectal varieties and simultaneously determine the underlying linguistic basis (Wieling and Nerbonne, 2010, 2011). This approach has been successfully used to obtain the linguistic basis of phonetic variation patterns observed for Dutch (Wieling and Nerbonne, 2010), English (Wieling et al., 2013a) and Tuscan (Montemagni et al., 2012) dialect datasets. Recently, the method has also been used to investigate lexical variation in contemporary English dialects on the basis of BBC Voices data (Wieling et al., 2013b).

This contribution will focus on lexical variation with some important novelties with respect to the study carried out on the BBC Voices data. First, the dataset used is represented by a dialectal atlas, i.e. the *Atlante Lessicale Toscano* (Giacomelli et al., 2000) and in particular the normalized results of onomasiological questions provided by 2060 informants distributed in 213 Tuscan-speaking locations. Second, the same type of analysis is carried out against the whole set of onomasiological questions as well as individual sections corresponding to the investigated semantic domains (e.g., agriculture, food, etc.), with the aim of investigating to what extent emerging patterns of lexical variation vary across semantic domains (preliminary results in this direction are reported in Montemagni, 2010). Third and finally, we adapted the importance measure used to determine the most important lexical variants associated with the identified dialectal areas multiplying distinctiveness and representativeness instead of averaging them, with the result of obtaining a measure reminiscent of tf-idf. The results of these analyses show interesting patterns (see e.g., Figure 1) and will be discussed together with the most important underlying features (see e.g., Figure 2).

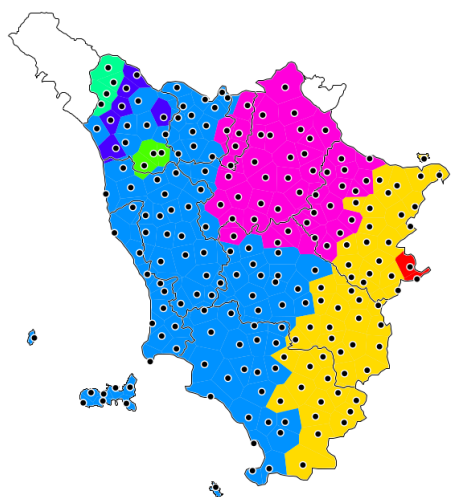


Figure 1:

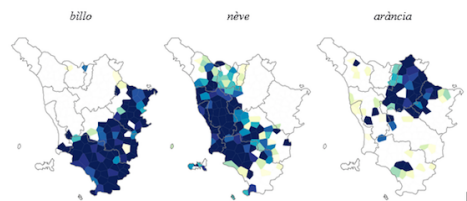


Figure 2: