

## Quantifying dialect Dutch verb clusters

**1. SUMMARY** This paper presents a quantitative analysis of variation in verb cluster ordering in dialect Dutch. It uses dialectometric methods as a tool to test theoretical accounts of this phenomenon.

**2. THE DATA** Verb cluster ordering is a well-known area of microparametric variation within Germanic (Barbiers & Bennis 2010, Wurmbrand 2005). For example, out of the six theoretically possible orderings in the three-verb cluster in (1), four are attested in Dutch dialects (Barbiers e.a. 2008):

- (1) a. *Ik vind dat iedereen moet kunnen zwemmen.*  
 I find that everyone must can swim  
 ‘I think everyone should be able to swim.’  
 b. *Ik vind dat iedereen moet zwemmen kunnen.*  
 c. *Ik vind dat iedereen zwemmen moet kunnen.*  
 d. *Ik vind dat iedereen zwemmen kunnen moet.*  
 e. *\*Ik vind dat iedereen kunnen zwemmen moet.*  
 f. *\*Ik vind dat iedereen kunnen moet zwemmen.*

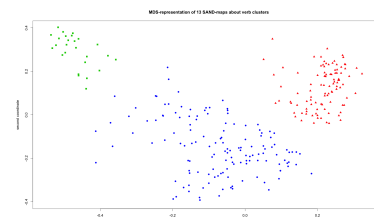


Figure 1: MDS-analysis of 13 SAND-maps

Missing from the literature, however, is a systematic study of the *correlations* between the various orderings. In this talk I examine the raw data from 13 maps of the *Syntactic Atlas of Dutch Dialects* (SAND, Barbiers e.a. 2008) to explore precisely this question.

**3. A DIALECTOMETRIC ANALYSIS** A technique that is often used in dialectometric research is Multidimensional Scaling (MDS, Nerbonne & Kretzschmar 2013), whereby a large number of linguistic differences between geographical locations is represented in a low-dimensional space. The goal is to detect—typically geographical—patterns in the data. When applied to the abovementioned data, three clusters can be detected (see Figure 1), and when these clusters are transposed to a map, they correspond nicely to three coherent geographical regions (see Figure 2). These results suggest that the distribution of verb cluster orderings across the Dutch-speaking area correlates with geography, but at the same time there are clear indications that a purely geography-based approach falls short: (1) when the data is split up per province, it becomes clear that there is no continuum going from north to south, (2) a statistical analysis in terms of a scree plot shows that the locations differ from each other in at least five or six dimensions, which makes an interpretation solely in terms of geography unlikely. In short, then, a deeper analysis of these facts is needed.

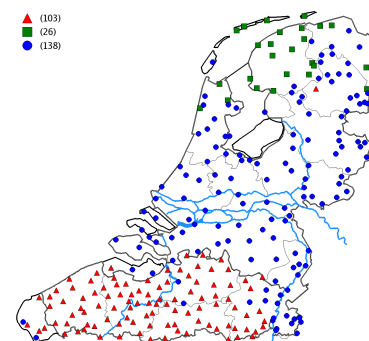


Figure 2: Geographical representation of Figure 1

**4. REVERSING THE PERSPECTIVE** What I propose to do is reverse the dialectometric perspective: rather than plot geographic locations based on their linguistic diversity, I plot linguistic constructions based on their geographical spread. This yields plots like the one in Figure 1, but with each point representing a verb cluster ordering rather than a location. Moreover, given that statistical analysis shows that there are four dimensions to the verb cluster data, the MDS-analysis yields the 4x4 scatterplot matrix in Figure 3. The main advantage of this approach is that it can serve as input for theoretical analyses of verb clusters: ideally the four dimensions of Figure 3 should correspond to four microparameters. I apply this line of reasoning by decomposing Barbiers’s (2008) analysis of verb clusters into four microparameters: (a) [ $\pm$ base-generation], (b) [ $\pm$ movement], (c) [ $\pm$  pied-piping], (d) [ $\pm$ feature-checking]. In Figure 3 I have color-coded the scatterplot according to Barbiers’s first parameter: pink signals [ $+$ base-generation] and blue [ $-$ base-generation]. In the interaction between the first two dimensions (first plot on the last line) we see a separation between the pink and the blue points, suggesting that Barbiers’s first microparameter does indeed play a role in accounting for the attested variation in verb cluster ordering.

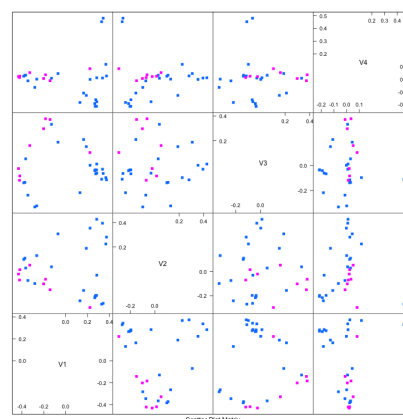


Figure 3: 4x4 scatterplot matrix of verb clusters color coded according to Barbiers’s (2008) base generation parameter