Linguistic distance as a determinant of the mutual intelligibility between Netherlandic and Belgian Dutch language varieties

Research on the mutual intelligibility of closely related Germanic languages has shown that several linguistic and extra-linguistic factors determine intelligibility scores to a high degree. In this paper, we aim to pinpoint the precise role of the determinant phonetic distance. As for example Gooskens (2007) shows, aggregate Levenshtein distances turn out to be good predictors of the intelligibility of Scandinavian standard and dialect varieties among subjects from Copenhagen. Also in studies on other languages such as Afrikaans and West Frisian, aggregate Levenshtein distances - in addition to other factors such as lexical distances, familiarity and language attitudes – have indeed proven to play an important role in predicting intelligibility scores (cf. van Bezooijen/Gooskens 2007).

Whereas most accounts on mutual comprehensibility deal with interlingual variation, this study focuses on intralingual differences, namely those of Dutch varieties. To gain insights into the linguistic basis of mutual intelligibility amongst Dutch language varieties, we first carried out a large-scale intelligibility experiment, consisting of a lexical decision test and a word identification task. The stimulus words we used in the test consisted of existing and nonsense words which were chosen according to several word-structural and frequency-based criteria. All words were recorded into ten Dutch target varieties, viz. a Netherlandic and Belgian realisation of the standard language as well as four more regional varieties from both countries. During the task, the stimulus words were auditorally presented in isolation to test subjects from the same eight regions as the regions from which we selected the regional target varieties. The subjects had to decide as quickly as possible whether the words exist in Dutch or not. We measured reaction time latencies to determine the degree of intelligibility.

Following a presentation of the results of this experiment, we will show to what degree the obtained intelligibility scores correlate with the factor linguistic distance. For that reason we have computed, as a second step, aggregate Levenshtein distances for all ten language varieties involved in the experiment, based on transcriptions of all cognates used in the tests. Levenshtein distance is a string edit distance calculating the least necessary costs to map one string of tokens on another one (cf. Heeringa 2004), and has already proven to be a successful method for measuring phonetic distances between Dutch dialects (Nerbonne et al. 1996). In the concrete, we aligned the phonetic segments from the word representations and calculated the least number of differences necessary to map the strings. The distance was normalized for word length. We obtained aggregate distances for all varieties by averaging the distances between all words compared per variety pair.
In this paper, we will present the results of these distance calculations and link them to the results of the intelligibility experiment. By correlating intelligibility scores with the aggregate distances, we will discuss the relative contribution of the factor *linguistic distance* to the intelligibility of Dutch varieties. Moreover, we will also discuss the limits of Levenshtein distance in the prediction of intelligibility, since Levenshtein distance is symmetric, whereas intelligibility often proves to be asymmetric.

**References**


