Mutual intelligibility of Dutch and German cognates by human and computers

Vincent van Heuven (Leiden), Charlotte Gooskens & Renée van Bezooijen (Groningen)

Many languages around the world are so closely related that they are mutually intelligible to a certain extent. Mutual intelligibility is not always symmetric. A number of investigations have shown that Danes understand Swedes better than Swedes understand Danes (Delsing & Lundin Åkesson, 2005) and that Brazilians understand (Argentinean) Spanish better than Argentineans understand (Brazilian) Portuguese (Jensen, 1989).

This project aims to objectively investigate factors determining intelligibility between Dutch and German. In the Dutch-German situation the fact that Dutch children learn German at school while Dutch is not a part of the curriculum for German children plays a role for intelligibility. The first aim of our investigation is to rule out the influence of education by testing the mutual intelligibility of Dutch and German children who have not yet learnt the neighboring language at school. We base our investigation on frequent Dutch-German cognates from the German and the Dutch Celex-corpora. These cognates were read aloud by a perfect bilingual speaker of German and Dutch and presented to the subjects in a translation task. If the percentages of correctly translated words is different for the two groups, we have evidence that the asymmetry is not caused by different school curricula in the two countries.

There is evidence that linguistic factors may explain asymmetric intelligibility (Gooskens & Van Bezooijen, 2006; Gooskens & Kürschner, 2009). However, in investigations of the role of linguistic factors it is problematic to isolate the influence of non-linguistic factors, such as difference in attitudes and contact, on intelligibility from the influence of linguistic factors. A possible solution is to use automatic speech recognition since attitudes and contact play no role for computers.

The second aim of our investigation, therefore, is to model the mutual Dutch-German intelligibility through automatic speech recognition. Dutch cognates that were also used in the intelligibility test described above were presented to a German speech recognizer and the German words to a Dutch speech recognizer of the same make (Dragon NaturallySpeaking Standard version 10). The speech recognizers were trained beforehand, in exactly the same way in both languages, by our bilingual speaker. The percentage of correctly recognized words per language is our measure of intelligibility. If the German-Dutch intelligibility is asymmetric it can be assumed that non-linguistic factors do not play a decisive role and that the asymmetry must be explained by linguistic factors only.

Finally, the results of the human-based experiment is compared to the results of the computer-based experiment.

References: