Since the 1990s, research in the field of translation studies has not only been focusing on the relation between source and target texts, but also on the relation between translated language and non-translated language (Baker 1993, 2004). Corpus-based research in translation studies has shown that translated texts differ in a surprisingly systematic manner from non-translated texts resulting in, for example, lexical (Laviosa 1998, Kemppanen 2004, Tirkkonen-Condit 2004) and grammatical differences (Puurtinen 2003, Olohan & Baker 2000). It is, however, remarkable that little research has been carried out to examine the differences between registers in translated versus non-translated language.

We want to investigate (i) to what extent the linguistic choices made in translations are register dependent and (ii) whether the linguistic differences between registers in translated texts are identical to those between these same registers in non-translated texts. To this end, we use the Dutch Parallel Corpus (DPC), a 10-million-word, high-quality, sentence-aligned parallel corpus for the language pairs Dutch-English and Dutch-French, with Dutch as the central language. It consists of the following registers: fiction, non-fiction, journalistic texts, instructive texts, administrative texts and external communication. (Macken et al. 2011).

We selected a set of linguistic profiles, viz. sets of alternatives with the same meaning and function such as jij kunt vs. jij kan, u heeft vs. u hebt (you have) en heeft gedaan vs. gedaan heeft (have done). On the basis of the frequency of these alternatives in the DPC, we used profile-based correspondence analysis (Plevoets 2008) in order to measure and plot the linguistic distances between the various registers. The general idea behind this technique is that the linguistic distances between registers and translated vs. non-translated texts grow as a function of the extent to which linguistic choices differ.

Preliminary results show that there is a significant distance between translated and non-translated Dutch. More specifically, we observe a tendency towards the use of rather formal language for translated texts and a tendency towards the use of rather informal language for non-translated texts. Furthermore, we see that the linguistic choices made by authors or translators are register dependent, and, for external communication and journalistic texts, these choices are significantly different for translated versus non-translated texts.


