Tools in corpus analysis

Silvia Hansen-Schirra

This paper discusses the role of annotated corpora as works of reference for grammatical translation problems. Within this context, different corpus designs as well as (online) corpora will be introduced as well as techniques for multi-layer alignment and annotation. These techniques comprise part-of-speech tagging, parsing, coreference resolution, morphological annotation, terminology etc. as well as the linguistic enrichment of treebanks. In addition, corpus query techniques are shown in order to extract relevant information from the corpora.

On the one hand, these corpora are used to investigate typical translation strategies as well as translation properties – such as explicitation, simplification, normalization, etc. Translations from the language pair English-German are evaluated against their source texts and originals in the target language in order to test whether translated text is more explicit, more typical in terms of register features or easier to understand. The findings are interpreted against the backgrounds of language typology, text type and register as well as cognitive effort during translation or translation reception.

On the other hand, it will be described how the corpora are exploited as interactive resource to display translation solutions for typologically problematic constructions. We illustrate the use of these interactive reference grammars for solving the following typical translation problems that occur in translating from English into German: English seems to be far more productive concerning cleft sentences, raising constructions and deletions, while German is characterised through more word order freedom. In the process of translating from English into German, this means that compensations have to be found and the word order has to be adapted. The existing descriptions of these grammatical problems are all example-based. While this is a good way to describe and define a phenomenon, it is not ideal for practical applications because it cannot take into account all instances to be found in the day-to-day work of translators. On the basis of the corpus annotation, however, relevant structures are extracted and translation candidates are displayed and discussed. The advantage of this technique is that translation solutions are extracted from published translations, i.e. language in use. Consequently, they are more comprehensive and inventive than dictionary entries or descriptions in grammars are. The corpora can thus be used as an interactive reference grammar in translation education and practice.