Stella Neumann (Saarbrücken, Germany)

Contrasting frequency variation of grammatical features

Abstract

This paper presents a quantitative approach to contrasting grammatical structures in English and German as structured by different registers of the two languages. In contrastive linguistics the comparison of grammatical structures is typically carried out on the level of the language system. Direct comparisons of lexico-grammatical structures across languages can be misleading: They may be distorted due to different functions of the features in the respective language. As Teich (2003) shows, the options of two language systems are increasingly divergent the more fine-grained the comparison is. Therefore, the contrastive analysis of grammatical features is typically carried out in relation to some basis of comparison. Underlying functions are often assumed to provide a suitable tertium comparationis. König & Gast (2007:5f.) employ general semantic notions such as 'temporal relation'.

These problems do not only apply to contrasting language systems but also to quantitative studies of language in use. Comparing the frequency of occurrence of a given feature directly may be a questionable procedure since it may fail to recognise the different quantitative roles of the feature in the two languages. A feature may be comparable in terms of the language system as shown by a similar function, but may be more, or less, common as expressed by diverging frequencies of occurrence.

Quantitative studies offer the opportunity to factor out language-specific frequencies provided that the corpus under investigation is stratified in terms of register and that a reference corpus balanced for register is included. This reference corpus should cover a wide range of different registers consequently serving as an indication of average language use. In order to avoid a confounding influence of the specific characteristics of the languages involved in terms of usage, an indirect method of comparison can then be adopted. The impact of peculiarities of the languages can be reduced by comparing only the magnitude of difference between the reference corpus serving as the baseline and the register-controlled corpus.

The present study approaches the contrastive comparison organised by registers assuming that the options provided by the language systems are not equally distributed and are used in different situational contexts. The functional categories of the register analysis (see e.g. Halliday & Hasan 1989) at the same time serve as a basis of comparison. This approach allows comparing the register-specific frequency of a given grammatical feature to the register-neutral reference corpus and only contrasts the magnitude of the difference between these two intralingual corpora to the corresponding magnitude in the other language. The study uses a subset of the CroCo Corpus (Vela et al. 2007), namely English-German original texts in 8 different registers totalling appr. 500,000 words plus reference corpora in both languages covering samples from 17 different registers. All subcorpora are annotated with metadata, part-of-speech and morphology information and phrase types and grammatical functions on the highest level in the sentence. Among the features analysed are part-of-speech frequencies, density of various grammatical units, modality, mood, tense and voice. On the basis of quantitative findings for these features, the paper will discuss the advantages of this method for the quantitative contrastive comparison and will also address its problems.

References

Halliday, M. A.K. & Ruqaiya Hasan (1989): Language, context, and text: aspects of language in a social-semiotic perspective. Oxford: Oxford University Press.

König, Ekkehard & Volker Gast (2007): Understanding English-German Contrasts. Berlin: Erich Schmidt.

Teich, Elke (2003): Cross-Linguistic Variation in System and Text. A Methodology for the Investigation of Translations and Comparable Texts. Berlin, New York: de Gruyter.

Vela, Mihaela, Stella Neumann & Silvia Hansen-Schirra (2007): "Querying multi-layer annotation and alignment in translation corpora." In: Proceedings of Corpus Linguistics 2007, Birmingham, UK, 27.-30. Juli 2007, http://ucrel.lancs.ac.uk/publications/CL2007/, article # 97.