Corpus-based contrastive studies: beginnings, developments and directions

There is nothing new about the idea that languages can be compared, but field of corpus-based contrastive studies is still fairly young. This talk will outline the beginnings of corpus-based contrastive linguistics, take stock of recent developments and finally attempt to identify some prospects for the future of the field.

When multilingual corpora entered the scene in the early 1990s they represented a new development in corpus linguistics, which had been predominantly monolingual until then, as well as in contrastive studies, which got a better empirical basis, new methods and a general boost. The development of the English-Norwegian Parallel Corpus (ENPC) was a seminal contribution. The plans for this corpus were first presented as a project idea at the ICAME conference in 1993 (Johansson & Hofland 1994), and soon after the project was joined by other teams developing similar corpora of English/Swedish and English/Finnish (Aijmer et al. 1996). However, the revival of contrastive studies was not limited to the Scandinavian context. Other teams elsewhere started compiling parallel corpora of other language pairs, for example English and French in the PLECI corpus. Clearly, the time was right for such an enterprise, not only because of interest in contrastive studies, but also in terms of general technological advances (Ebeling 2016).

At present, corpus-based contrastive linguistics is a well-established field of research which can be distinguished from the neighbouring fields of translation studies and learner corpus studies. While the field has diversified to include comparisons of more languages, text types and linguistic features (as evidenced e.g. by papers published in Languages in Contrast), some challenges remain, for instance the availability of suitable data and reliable methods for making valid cross-linguistic comparisons based on parallel (translation) data as well as comparable data. In one of his last papers, Johansson (2012) still sees multilingual corpus linguistics “as a good means of doing language research and an important direction of research for the future” even though future challenges include a “need to learn more about how we can best exploit multilingual corpora”.

---

**References:**


