Hilde Hasselgård

CORPUS-BASED CONTRASTIVE GRAMMAR STUDIES
Some challenges and insights from crosslinguistic studies of adverbials

Keywords Corpus linguistics; grammar; comparable corpora; parallel corpora; total accountability; tertium comparationis; adverbials

"Exploring grammatical phenomena in a multilingual corpus is a difficult and time-consuming task involving much manual intervention, unless there are clear lexical correlates or the corpus has been syntactically annotated". These words come from a pioneer in corpus-based contrastive studies, Stig Johansson (2007, p. 37). In my talk I will discuss some of the challenges involved in grammar studies based on multilingual corpora. I will use examples and insights from my own crosslinguistic studies of adverbials, especially comparing English and Norwegian. Even monolingual studies of adverbials are problematic for a corpus linguist because it is not obvious how adverbials can be retrieved with reasonable reliability. In a crosslinguistic study, even of closely related languages, the problems are compounded because equivalent syntactic functions may be realized by different lexicogrammatical means, as in this example, where the English PP corresponds to a German adverb.

> At last the litany came to an end. (PM1)

> Endlich fand die Litanei ein Ende. (PM1T)

It may thus be difficult to achieve the corpus-linguistic goal of total accountability, i.e., the wish to account "for all the corpus data in the relevant samples" (Leech 1992, p. 113).

Other challenges are related to the tertium comparationis (TC) – what is the common ground for the comparison? And what is the rationale for comparing set of expressions in one language to another set in a different language? Types of TC include etymological similarity (cognates), similarity of function (e.g., time adverbials), dictionary data and/or translation equivalence, but not all of these may be relevant in grammar studies.

In addition to these challenges, I will discuss different corpus types (parallel and comparable) and search methods (e.g., manual excerption, use of PoS tags, and N-gram extraction). An important consideration is whether different methods of investigation give results that are compatible with each other, and I will illustrate this with evidence from my own crosslinguistic explorations (e.g. Hasselgård 2014, 2017, 2021).

References

Hasselgård, Hilde (2014): Additive conjunction across languages: 'dessuten' and its correspondences in English and French. In: Oslo Studies in Language 6 (1). https://doi.org/10.5617/osla.691.

Hasselgård, Hilde (2017): Temporal expressions in English and Norwegian. In: Janebová, M./ Lapshinova-Koltunski, E./Martinková, M. (eds.): Contrasting English and other languages through corpora. Newcastle: Cambridge Scholars Publishing, pp. 75–101.

Hasselgård, Hilde (2021): Time adverbials in English and Norwegian news discourse. In: Cermáková, A./Egan, T./Hasselgård, H./Rørvik, S. (eds.): Time in languages, languages in time. Amsterdam: Benjamins, pp. 201–227. https://doi.org/10.1075/scl.101.09has.

Johansson, Stig (2007): Seeing through multilingual corpora. (= Studies in Corpus Linguistics). Amsterdam: Benjamins. https://doi.org/10.1075/scl.26.

Leech, Geoffrey (1992): Corpora and theories of linguistic performance. In: Svartvik, Jan (ed.): Directions in corpus linguistics. Proceedings of Nobel Symposium 82 Stockholm, 4–8 August 1991. (= Trends in Linguistics. Studies and Monographs 65). Berlin/New York: De Gruyter Mouton, pp. 105–123. https://doi.org/10.1515/9783110867275.105.