On distributional patterns of verbs in English and German

This paper addresses the distribution of verbs in English and German with a particular focus on infinitival constructions. One frequently mentioned area of contrast between English and German is a difference in nominal versus verbal patterns. Kortmann and Meyer (1992, 165) discuss a tendency of English to maximise the verb phrase, whereas German is said to expand the noun phrase, resulting in English being primarily verb-oriented and German being noun-oriented. However, such tendencies do not translate to an increased frequency of nouns when accounting for spelling differences of compounds (Berg et al. 2012, 280; Neumann 2020; also implicitly Gast and Borges 2023). Against this background, Neumann (2020, 150) suggests that a difference in verb usage might be responsible for the perceived preference for nominal versus verbal patterns in the two languages. She speculates that German does not draw on non-finite verbs to the same extent in comparable grammatical contexts which involve similar numbers of nouns, explaining this by the limited additional information a non-finite verb may contribute. English, in turn, seems to express the procedural aspect of an event or state more congruently with the help of a verb.

This study sets out to test this claim with the help of a quantitative corpus analysis in English and German. It draws on the original parts of the CroCo Corpus (Hansen-Schirra, Neumann, and Steiner 2012). This balanced corpus contains original and translated texts in both translation directions from eight comparable registers. The original subcorpus comprises 500,697 words across 231 texts in the two languages and is annotated with part of speech (PoS) information and indexed with the IMS Open Corpus Workbench (CWB; Evert and Hardie 2011). Different forms of verbs are extracted with the help of complex queries in the CQP query syntax (Evert and The CWB Development Team 2020) supported by the CWB. A particular focus of the analysis is on *to/zu* infinitives. In German, these are extracted both in the form infinitive marker followed by verb (*zu verstehen*) and as single verbs with the infinitive marker integrated morphologically (*aufzulegen*). All occurrences are normalised using a reasonable unit of measurement, such as the number of finites or sentences per text. Normalised by number of words, English texts generally contain more verbs as well as more *to/zu* infinitives, but similar numbers of finites than German texts. When compared by number of sentences, all three frequencies are clearly higher in English than in German. Linear regression models with these three features as response variables and language and register as predictors (sum-coded) and including an interaction term for language and register yield significant main effects for language and various individual registers. Additionally, the model for *to/zu* infinitives also retrieves interactions between language and register, indicating that registers display specific distributions in the context of German.

These results corroborate claims about the stronger verb orientation of English. Inspection of the query hits suggest that many of the occurrences can be explained straightforwardly by constructions such as embedded clauses serving as postmodifiers within noun phrases, semi-modals such as *need to* and *have to* and the *going to* future. Frequent contexts also include phasal constructions with *continue* or *begin.* To explore possible equivalents for such constructions, the same query was applied to the aligned English-German translation pairs in the CroCo corpus. The results indicate that verbs expressing phase are often translated by adverbs such as *weiterhin* with the *to* infinitive translated by a finite verb. In such cases, the German clause contains only one instead of two verbs and the only remaining verb is finite, thus offering one explanation for the difference in frequency of verbs. The results suggest that German offers a wider range of distributional options corresponding to the English infinitival constructions. These other options are hinted at by the frequency difference in the formally corresponding construction and underline the importance of complementing system-based comparisons with usage-based data.

References

Berg, Thomas, Sabine Helmer, Marion Neubauer, and Arne Lohmann. 2012. 'Determinants of the Extent of Compound Use: A Contrastive Analysis'. *Ling* 50 (2): 269–303. https://doi.org/10.1515/ling-2012-0010.

Evert, Stefan, and Andrew Hardie. 2011. 'Twenty-First Century Corpus Workbench: Updating a Query Architecture for the New Millennium'. In *Proceedings of the Corpus Linguistics Conference 2011, University of Birmingham, UK, 20-22 July 2011*. Birmingham, UK: University of Birmingham. http://www.birmingham.ac.uk/documents/college-artslaw/corpus/conference-archives/2011/Paper-153.pdf.

Evert, Stefan and The CWB Development Team. 2020. 'The IMS Open Corpus Workbench (CWB) CQP Query Language Tutorial'. http://cwb.sourceforge.net/files/CQP_Tutorial/.

Garside, Roger, and Nicholas Smith. 1997. 'A Hybrid Grammatical Tagger: CLAWS4'. In *Corpus Annotation: Linguistic Information from Computer Text Corpora*, edited by Roger Garside, Geoffrey Leech, and Anthony McEnery, 102–21. London: Longman. http://ucrel.lancs.ac.uk/papers/HybridTaggerGS97.pdf.

Gast, Volker, and Robert Borges. 2023. 'Nouns, Verbs and Other Parts of Speech in Translation and Interpreting: Evidence from English Speeches Made in the European Parliament and Their German Translations and Interpretations'. *Languages* 8 (1): 39. https://doi.org/10.3390/languages8010039.

Hansen-Schirra, Silvia, Stella Neumann, and Erich Steiner. 2012. *Cross-Linguistic Corpora for the Study of Translations - Insights from the Language Pair English-German*. Berlin: de Gruyter Mouton.

Kortmann, Bernd, and Paul Georg Meyer. 1992. 'Is English Grammar More Explicit than German Grammar, after All?' In *New Departures in Contrastive Linguistics. Neue Ansätze in der Kontrastiven Linguistik: Proceedings of the Conference Held at the Leopold-Franzens-Universität Innsbruck, Austria, 10-12 May 1991*, edited by Christian Mair and Manfred Markus, 1:155–66. Innsbruck: Verlag des Instituts für Sprachwissenschaft.

Neumann, Stella. 2020. 'Is German More Nominal than English? Evidence from a Translation Corpus'. In *New Approaches to Contrastive Linguistics*, edited by Renata Enghels, Marlies Jansegers, and Bart Defrancq, 127–58. Berlin: De Gruyter Mouton. https://doi.org/10.1515/9783110682588-005.

Schiller, Anne, Simone Teufel, Christine Stöckert, and Christine Thielen. 1999. 'Guidelines für das Tagging Deutscher Textcorpora Mit STTS'. Universität Stuttgart, Universität Tübingen. http://www.ims.uni-stuttgart.de/ftp/pub/corpor⅜stts_guide.ps.gz.

Schmid, Helmut. 1994. 'Probabilistic Part-of-Speech Tagging Using Decision Trees'. In *Proceedings of the International Conference on New Methods in Language Processing*, 44–49. Manchester, UK.