

Hanna Mahler

# The Use of Verb Phrases in English and German

## A Quantitative Case Study Using Comparable Corpus Data

**Keywords:** English; German; Comparable Corpus; GECCo Corpus; Verb Phrase; Bayesian Statistics

In the existing literature on English-German contrasts, claims as to the more frequent use of verb phrases in English, compared to a more “nominal” style in German, abound (e.g. Kortmann/Meyer 1992, p. 163; Königs 2004, p. 1) The empirical evidence on this is, however, scarce. The few existing empirical studies either use translation data (e.g. Fischer 2013), which is not adequate, focus on nominal style (e.g. Neumann 2020), or rely purely on the count of automatically assigned parts-of-speech tags (e.g. Berg 2017).

A thorough empirical investigation of the “verbality” of English and German is needed to complement the existing research. For such an investigation, it is essential to pay due attention to differences associated with register and mode in each language. It is a possibility that language-internal register/mode differences might be overall more pronounced than cross-linguistic differences, or that cross-linguistic differences are especially pronounced in certain registers (e.g. Königs 2004, pp. 3–5).

The study at hand uses GECCo, a comparable corpus of spoken and written texts from English and German (Kunz et al. 2021). The German component contains around 488,000 tokens, the English component around 551,000 tokens (UPOS-version of the corpus).

Through a combination of automatic and manual processing, all verb phrases in the corpus are identified and annotated for finiteness, verb form, and grammatical function in the clause. The frequency of finite and non-finite verb phrases (in various functions) in English and German is then examined with the help of a Bayesian mixed-effects regression model. In total, 68,658 verb phrases are found in English (of which 15,508 are non-finite) and 50,289 in German (4,462 non-finite).

The results confirm existing assumptions that English uses more verb phrases overall, that in both languages verb phrases are used more often in spoken texts, that information density is negatively correlated with the frequency of verb phrases, and that there are considerable differences by register. The difference by mode is illustrated in Figure 1, which depicts the number of verb phrases (sum of finite and non-finite) per hundred words<sup>1</sup> for the two languages. One can see that the effect of mode is stronger in English, which is confirmed by the regression model.

---

<sup>1</sup> Counting verb phrases per sentence was not an option due to unreliable sentence segmentation in the corpus.

The data furthermore allows for a subsequent investigation of the contribution of non-finite verb phrases to these overall differences since the existing literature heavily suggests a central role of these structures (e.g. Fischer 2013, p. 169).

In sum, the study at hand

- provides new insights into verbal style through fine-grained measurements and statistical modelling based on non-translation data,
- offers empirical evidence for common claims about English-German contrasts,
- illustrates the value of advanced statistical procedures for contrastive linguistics,
- and contributes to the emerging branch of (corpus-based) “quantitative Contrastive Linguistics” (Gast 2015, p.5).

From a practical perspective, the results of this study will be useful for providing more nuanced advice for translation between the two languages, as well as for designing empirically grounded language acquisition material. From a theoretical perspective, the study provides insights on the cross-linguistic relevance of mode, information density, and register for the use of verb phrases in a given language.

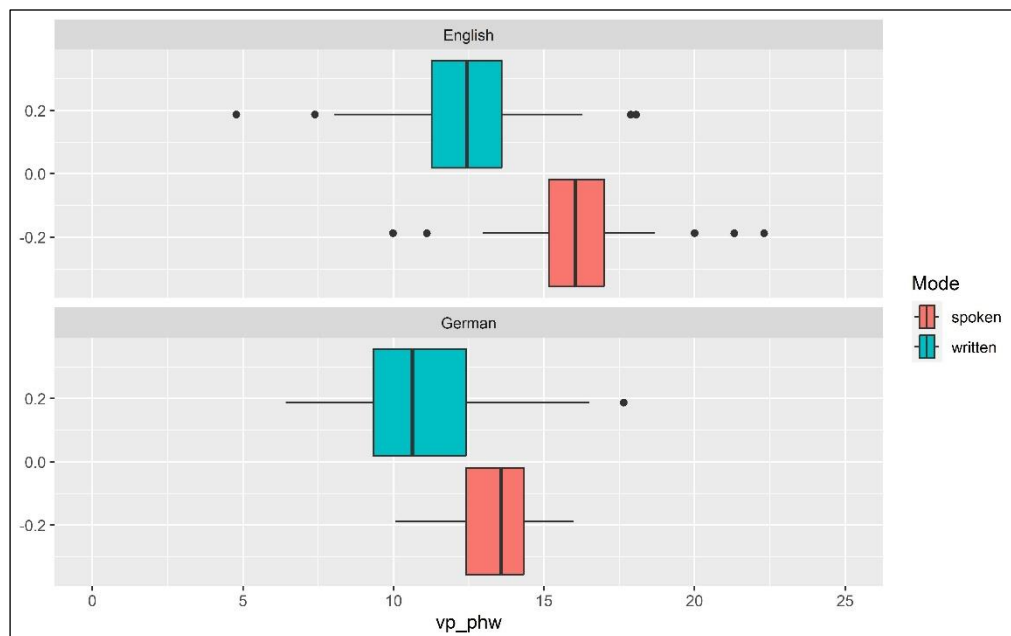


Fig. 1: Frequency of verb phrases per hundred words in English and German by mode (based on GECCO corpus)

## References

- Berg, Thomas (2017): Compounding in German and English: A Quantitative Translation Study. In *Languages in Contrast*, 17(1), pp. 43–68. DOI: 10.1075/lic.17.1.03ber.
- Fischer, Klaus (2013): *Satzstrukturen im Deutschen und Englischen: Typologie und Textrealisierung (Konvergenz und Divergenz 1)*. Berlin: Akademie Verlag.
- Gast, Volker (2015): On the Use of Translation Corpora in Contrastive Linguistics: A Case Study of Impersonalization in English and German. In *Languages in Contrast*, 15(1), pp. 4–33. DOI: 10.1075/lic.15.1.02gas.

- Königs, Karin (2004): Übersetzen Englisch-Deutsch: Ein Systematischer Ansatz. 2nd edition. München: Oldenbourg Wissenschaftsverlag (Lehr- und Handbücher zu Sprachen und Kulturen).
- Kortmann, Bernd/Meyer, Paul (1992): Is English Grammar More Explicit than German Grammar, after all? In: Mair, Christian/Markus, Manfred (eds.): *New Departures in Contrastive Linguistics: Proceedings of the Conference Held at the Leopold-Franzens-Universität Innsbruck, Austria, 10-12 May 1991* (Innsbrucker Beiträge zur Kulturwissenschaft / Anglistische Reihe 4), pp. 155–166. Innsbruck: Verlag des Institutes für Sprachwissenschaft.
- Kunz, Kerstin/Lapshinova-Koltunski, Ekaterina/Martínez Martínez, José M./Menzel, Katrin/Steiner, Erich (2021): *GECCo - German-English Contrasts in Cohesion: Insights from Corpus-Based Studies of Languages, Registers and Modes* (Trends in Linguistics. Studies and Monographs 355). De Gruyter Mouton.
- Neumann, Stella (2020): Is German more Nominal than English? Evidence from a Translation Corpus. In Enghels, Renate/Defrancq, Bart/Jansegers, Marlies (eds.): *New Approaches to Contrastive Linguistics: Empirical and Methodological Challenges*, pp. 127–158. Berlin, Boston: De Gruyter.

## Contact information

**Hanna Mahler**

Albert-Ludwigs-Universität Freiburg

[hanna.mahler@students.uni-freiburg.de](mailto:hanna.mahler@students.uni-freiburg.de)