

“To Think Under”: Behavioral and Computational Evaluation of Novel Complex Verb Interpretability in German

Moritz Wackerbarth (moritz.wackerbarth@uni-bielefeld.de)¹, Sina Zarrieß¹,
Sabine Schulte im Walde², Antje Lorenz¹

¹Faculty of Linguistics and Literary Studies, Bielefeld University

²Institute for Natural Language Processing, University of Stuttgart, Germany

German is a morphologically rich language and offers its speakers a multitude of ways to express new meaning through the formation of novel complex words. While a large field of psycholinguistic research concerns itself with the processing of familiar complex words, the interpretation and learning of novel complex words is a comparatively limited field. The present study aimed at identifying compositional restrictions in the formation of German particle and prefix verbs that arise in different combinations of morphemes and semantic verb classes [e.g., 1, 2]. To this end, we generated complex verbs by prepending 12 German particles and five true German prefixes to 15 German simplex verbs in each of five semantic classes, expanding on [3] (see Supplemental Information). This yielded 1,275 complex words, 861 of which were unattested in a large corpus of German [4] and were therefore treated as novel. An online study was conducted in which 104 native speakers of German participated. Each participant rated the interpretability of a subset of 80 novel complex verbs on a 7-point Likert scale. Each participant further provided text definitions and usage examples for an additional 8-10 items interspersed between rating trials. Likert ratings were analyzed with a cumulative link mixed model with fixed effects of semantic verb class, morpheme type, their interaction, and base-verb frequency (standardized and log-transformed), with random intercepts for participants, items and morphemes (nested within morpheme type). For both particles and prefixes, we found that ratings shifted significantly towards higher rating categories for *physical-process* verbs and towards lower rating categories for *achievement* and *state* verbs (see Figure 1). A particle-prefix difference only emerged for *physical-process* verbs, where particle verbs shifted towards higher ratings than prefix verbs. Further, ratings were found to significantly shift toward lower rating categories when base verb frequency increased, irrespective of morpheme type (see Figure 2). Our findings on *physical-process* type verbs may reflect greater compatibility of imageable concepts with various meaning modifications in German complex verb formation. This appears to be pronounced for combinations with particles, which are particularly amenable to spatial readings. This is an ongoing study, which currently aims to align its findings in behavioral data with interpretability predictions acquired from LMs trained with character-level tokenization. These types of models have been shown to be well-suited to process non-word data and thus lend themselves to investigate novel words [5].

References

- [1] Quante, L., et al. (2018). Dissociating predictability, plausibility and possibility of sentence continuations in reading: evidence from late-positivity ERPs. *PeerJ*, 6, e5717.
- [2] Springorum, S., et al. (2019). Cognitive and linguistic aspects of composition in German particle verbs. *Perspectives on Abstract Concepts: Cognition, Language and Communication*, 65, 185
- [3] Springorum, S., et al., (2013). Sentence generation and compositionality of systematic neologisms of German particle verbs, *Proceedings of the 5th Conference on Quantitative Investigations in Theoretical Linguistics*, 81-84.
- [4] Kliegl, R., Hanneforth, T., Geyken, A., Würzner, K.-M., Heister, J., Pohl, E., Bubenzer, J., & Wiegand, F. (2025). dlexDB – annotated lexical data [Data set]. Zenodo.
- [5] Bunzeck, B., & Zarriß, S. (2025, July). Subword models struggle with word learning, but surprisal hides it. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)* (pp. 286-300).

Figures and Tables

Figure 1: Model-estimated ratings across semantic verb classes

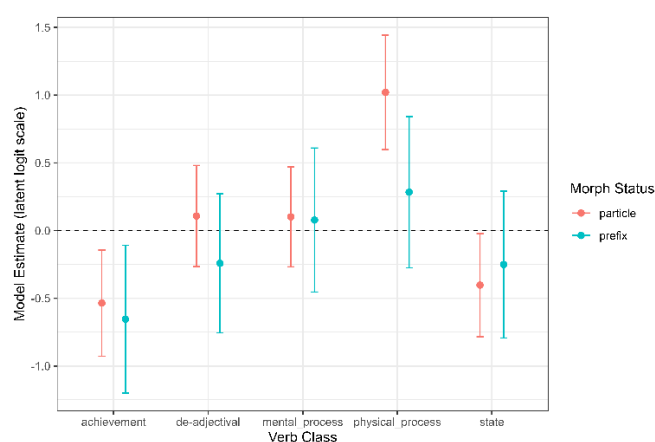
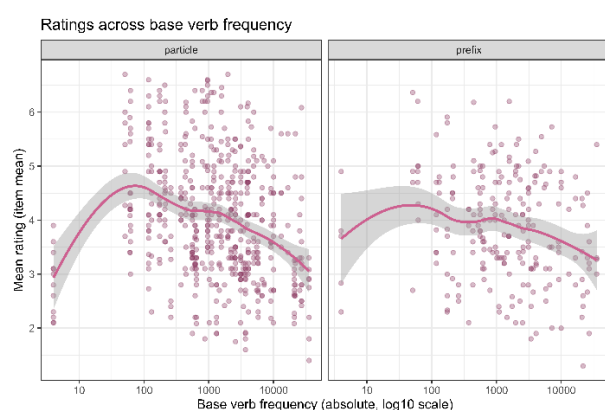


Figure 2: Mean ratings across base verb frequency by morpheme type



Optional Supplemental Information

Verb classes and base verbs

Verb Class	Simplex Verb (Gloss)
Achievement	brechen (to break), fallen (to fall), ..., schließen (to close / to shut)
De-adjectival	glätten (to smooth), härten (to harden), ..., weiten (to widen)
Mental process	ahnen (to suspect / to sense), denken (to think), ..., zweifeln (to doubt)
Physical process	graben (to dig), laufen (to run), ..., tanzen (to dance)
State	bleiben (to stay / to remain), gelten (to be valid / to apply), ..., wollen (to want / to intend)

Particles

Particle	Prototypical Gloss(es)
ab	off / down / away
an	on / at / to (contact)
auf	on top of / up / open
aus	out
bei	at / by / near / with
durch	through / through-and-through
hin	toward / thither (away from speaker)
los	off / start
nach	after / to (esp. cities, countries)
unter	under / among
vor	in front of / before
weg	away

Prefixes

Particle	Prototypical Gloss(es)	Typical Function / Tendency
be-	often untranslatable, ≈ affect/cover/provide	often transitivizes; marks an object as affected/covered/targeted
ent-	de- / un- / away	removal, reversal, undoing
miss-	mis-	error/incorrectness, negative evaluation
ver-	varies, ≈ mis-, away-, change-	change of state, loss/away, error, consumption/completion
zer-	apart / to pieces	disintegration/destruction