Towards a Verb Class-based Semantic Analysis of German Literary Texts

Analyzing themes in literary texts is a challenge in traditional as well as in computational literary studies. Traditional approaches to thematic structures such as "Stoff" or "Motiv" (Frenzel, 1980; Daemmrich & Daemmrich, 1995) focus on conventionalized themes and lack modern methodological rigor. Topic modeling in computational approaches, on the other hand, does not provide a general thematic structure and is therefore heavily interpretation dependent. Therefore, we explore the use of semantic verb classes to create thematic profiles of literary texts. While resources like BookNLP (Bamman 2021) with its semantic role labeling exist for English, German still lacks a comparable implementation for semantic analysis.

We therefore are developing a classification of verbs into semantic classes based on GermaNet (Hamp & Feldweg, 1997). GermaNet is a lexical-semantic net for German, inspired by WordNet (Fellbaum, 1998) and categorizes verbs into 15 broad semantic fields derived from Levin's taxonomy (Levin, 1993). These fields are abstract, avoiding the sparse coverage issues seen in fine-grained lexicons like FrameNet (Ruppenhofer et al., 2016). Instead, broad categories like "Körperfunktion" (verbs of the body) and "Schöpfung" (verbs of creation) allow more flexible thematic analysis.

1. Annotation of Semantic Verb Classes

We conducted manual annotations of the EvENT-Korpus (Vauth & Gius 2022) using GermaNet's web interface. Based on the definitions provided by GermaNet, literary studies students classified verbs in verbal phrases according to their main verb while excluding modal verbs. This resulted in 10,547 annotations (cf. Table 1).

Category	Translation	Count
Allgemein	Stative Verbs	2561
Lokation	Verbs of Location	1911
Kognition	Verbs of Cognition	1151
Kommunikation	Verbs of Communication	1023
Veränderung	Verbs of Change	924
Perzeption	Verbs of Perception	727
Gefühl	Verbs of Emotion	516
Gesellschaft	Social Verbs	503
Besitz	Verbs of Possession	352
Kontakt	Verbs of Contact	322
Körperfunktion	Verbs of Body	270
Schöpfung	Verbs of Creation	91
Naturphänomen	Verbs of natural Phenomenon	80
Verbrauch	Verbs of Consumption	76
Konkurrenz	Verbs of Competition	40

Table 1. Occurrences of Verb Classes.

Cohen's Kappa scores across texts averaged 0.59, indicating moderate agreement (cf. Table 2). However, the scores varied significantly across texts, possibly reflecting the high dependence of semantic meaning from their specific context in literary texts. For example, "Die Verwandlung" (Kafka) and "Die Judenbuche" (Droste-Hülshoff) achieved scores above 0.6, while "Der blonde Eckbert" (Tieck) scored only 0.33 due to a single annotator's outlier ratings.

Moreover, the specific verb classes revealed disparities in annotation difficulty. For instance, "Verbrauch" (verbs of consumption) achieved near-perfect agreement, while "Konkurrenz" (verbs of competition) often overlapped with "Kommunikation" (verbs of communication) or "Lokation" (verbs of location). Additionally, the "Allgemein" class (stative verbs) dominated the dataset, accounting for 25% of the 10,547 annotations. This prevalence reduced specificity, raising questions about whether "Allgemein" functions as a catch-all for ambiguous cases.

Text Title	Number of Annotations			Fleiss' Kappa
	Annotator 1	Annotator 2	Annotator 3	
Das Erdbeben in Chili	542	725	677	0.6938
Der blonde Eckbert	898	715	54	0.3160
Die Judenbuche	52	0	2269	0.6627
Die Verwandlung	2296	1646	53	0.7621
Krambambuli	135	0	485	0.6151

Table 2. Annotations and agreement scores.

2. Semantic Verb Classes for Text Analysis

For the semantic exploration, we generated radar plots showing the ratio of the verb class occurrences in each of our texts to the expected occurrence of each class, taking into account their global frequency and the text length (cf. Fig. 1). The text-specific representation of the verb classes revealed discrete thematic profiles in at least four of the five texts. "Das Erdbeben in Chili" (Kleist) prominently featured "Naturphänomen" verbs, reflecting its focus on natural disasters. Meanwhile, Kafka's "Die Verwandlung" exhibited a more generic profile, and "Die Judenbuche" showed limited thematic distinctions, likely due to its length diluting verb class specificity. Shorter texts allowed rare classes, such as "Konkurrenz," to stand out, whereas longer texts emphasized common categories like "Allgemein" and "Kognition" (cognition).
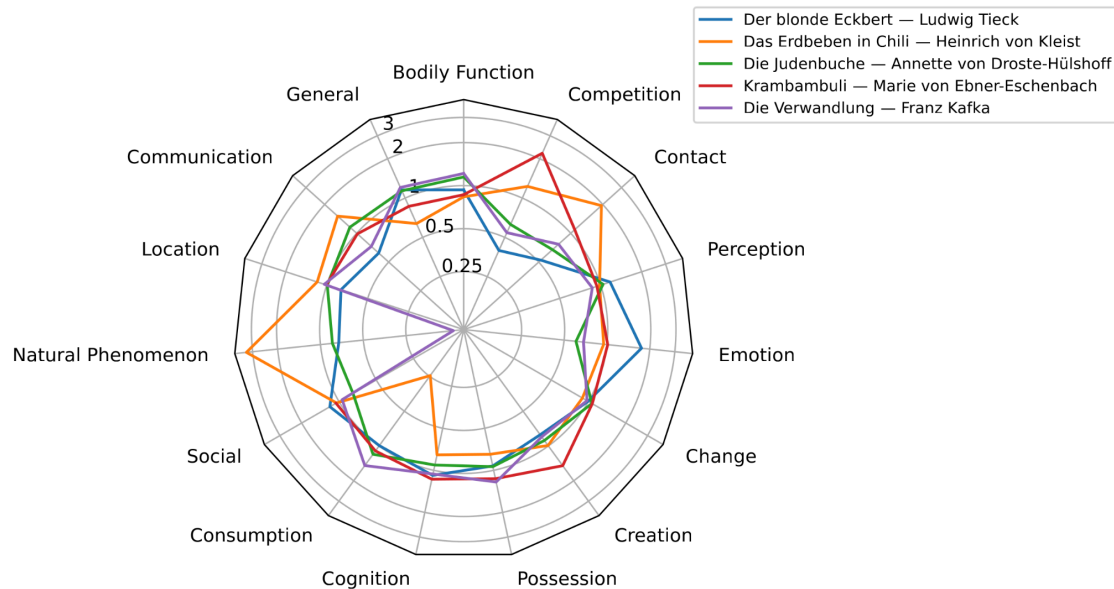
Figure 1: The comparative frequency of verb classes in our analyzed texts. Note that the radial axis is in log scale.

We further analyzed the relationship between semantic classes and with event types, following the framework by Vauth et al. (2021) in which types are categorized into non-events, stative-events, process-events, and changes of state based on their narrativity. We found that specific verb classes correlated strongly with event types. For example, "Kommunikation" (communication) verbs frequently represented process-events (89%), while "Naturphänomen" (natural phenomena) verbs aligned with stative-events (64%).

3. Next Steps

For the automation of this approach, we experimented with retrieval-augmented generation based on Llama-3 and achieved promising results with an accuracy of 91%, outperforming human annotators. For an exploration of the potential of our approach for literary analysis of bigger corpora, we are currently working on a comparative corpus study of German Modernism.

References

Bamman. D. (2021). BookNLP. https://github.com/booknlp/booknlp.

Daemmrich, H. S., & Daemmrich, I. G. (1995). *Themen und Motive in der Literatur. Ein Handbuch*. Francke.

Fellbaum, C. (1998). *WordNet: An Electronic Lexical Database*. MIT Press.

Frenzel, E. (1980). *Motive der Weltliteratur: Ein Lexikon dichtungsgeschichtlicher Längsschnitte*. Kröner.

Hamp, B., & Feldweg, H. (1997). GermaNet – A Lexical-Semantic Net for German.

Levin, B. (1993). *English Verb Classes and Alternations: A Preliminary Investigation*. University of Chicago Press.

Vauth, M., Hatzel, H.O., Gius, E. & Biemann, C. (2021). "Automated Event Annotation in Literary Texts." In *CHR 2021: Computational Humanities Research Conference*, edited by Maud Ehrmann, Folgert Karsdorp, Melvin Wevers, Tara Lee Andrews, Manuel Burghardt, Mike Kestemont, Enrique Manjavacas, Michael Piotrowski, and Joris van Zundert, 2989:333–45. CEUR Workshop Proceedings. Amsterdam, The Netherlands. http://ceur-ws.org/Vol-2989/short_paper18.pdf.

Vauth, M. & Gius, E. (2022). "Event Annotations of Prose." *Journal of Open Humanities Data* 8: 19. https://doi.org/10.5334/johd.83.