Towards a Conflict Heuristic. Detecting Conflict in Literary Texts By Adapting Word Embedding Based Sentiment Analysis

Häußler, Julian
julian.haeussler@tu-darmstadt.de
Technische Universität Darmstadt, Germany

Gius, Evelyn
evelyn.gius@tu-darmstadt.de
Technische Universität Darmstadt, Germany

Conflict in the German Romantic Period

With the aim of detecting conflicts in literary texts, this short paper adapts a successful implementation of word embedding based sentiment analysis for assigning conflict values to texts. In literary studies, the concept of conflict is used in a variety of contexts ranging from analyzing conflicts between characters over describing epochs with regard to their dealing with conflicts in the society to discussing poetological issues. While working towards a broader understanding of conflict for literary texts, we here focus on conflict between characters according to Glasl's notion of social conflict (see Glasl 2011). More specifically, we aim at identifying texts and passages with comparatively high and low representations of conflictuality since in Glasl's conflict notion the very perception of a situation as conflictual is crucial. A corpus of 115 German novels from the Romantic period (see Schumacher et al. 2022) served as the base for developing the three approaches described here. The analysis was then performed on a sub-corpus of 26 novels from the core Romantic period (1792–1833). Of these, all verb phrases were extracted, as they constitute a coherent analysis unit, on which the conflict values were calculated.

Three Approaches to the Representation of Conflict

The first approach implements a word embedding based sentiment analysis as introduced by Jacobs in the *SentiArt* project (see e.g. Jacobs 2019). In *SentiArt*, emotion bearing label words are chosen to represent the extreme ends of a valence and an arousal spectrum. The sentiment value for word x in a text then is calculated by retrieving the cosine similarity values of word x to each label word in a word embedding model (we use Word2Vec) and subtracting the mean of similarities for the negative labels from the mean for the positive labels. Our approach follows the emotion model by Yu et al. 2016 (which is based on Russell 1980) and focuses on the intersections of the four sentiment poles (top/bottom valence/arousal). We translated its label words into German and localized them with regard to the language of the corpus. Finally, mean sentiment values for each verb phrase were calculated.

To tackle the notion of conflict directly, in the second approach a list of label words for the opposing concepts of conflict and harmony was established based on a German word field dictionary (Dornseiff 2004). Then, words from this preselection not in the Romantic corpus as well as words with an overall frequency lower than 100 (as low-frequency words make unreliable word vectors, see Řehůřek 2022) were filtered out. The mean conflict value was finally calculated in the same way as the sentiment values in the previous approach.

To also consider conflict representation from a literary source, a third approach selects label words from a conflict annotation effort of five German prose narratives. Based on this, a conflict dictionary, containing information on how indicative a word is for a tag was created by weighting the annotated words. The words that were most indicative of conflict were selected as label words for high conflict whereas the words that signaled conflict resolution for low conflict. All label words were lemmatized and those below the frequency threshold of 100 were filtered out. This conflict value finally was computed in the same way as the previous values. [1]

Comparison and First Results

For analyzing the conflictuality of the texts in our corpus, we determined the share of verb phrases that belong to the 10% phrases with the highest degree for each of the three approaches throughout the corpus. For each text, we then normalized the value according to the length of the text.

First results show that all three approaches identified verb phrases with a high/low relation to conflict and thus seem to be a viable heuristic for the representation of conflict in literary texts. Moreover, while there is a considerable amount of novels containing verb phrases at the intersection of both high arousal and negative valence ('angry texts') or both low arousal and positive valence ('relaxed texts'), there are only very few cases of verb phrases that combine low arousal and negative valence ('boring texts') or high arousal and positive valence ('delighted texts'). By combining the three approaches into one value, we also found that the frequency of verb phrases with high conflict representation seems to increase throughout the romantic period (see Figure 1) which might point to the poetological relevance of conflicts in early romanticism.

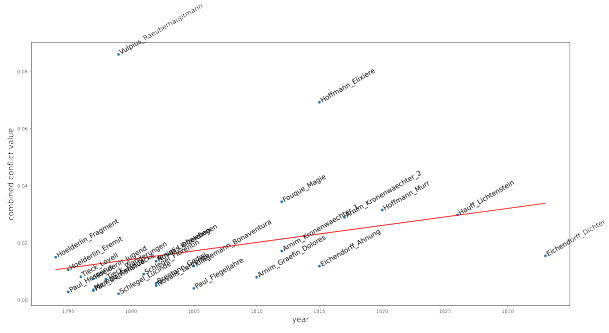

Figure 1: *Ratio of high conflict verb phrases in core corpus of romantic novels.*

Notes

1. The Jupyter Notebooks used for the analysis will be published under https://github.com/forTEXT.

Bibliography

Dornseiff, Franz / **Wiegand, Ernst** (2004): *Der Deutsche Wortschatz Nach Sachgruppen*. 8., Völlig neu bearbeitete und mit einem vollständigen alphabetischen Zugriffsregister versehenen Aufl. Berlin; New York: De Gruyter.

Glasl, Friedrich (2011): *Konfliktmanagement: Ein Handbuch für Führungskräfte, Beraterinnen und Berater*. Bern: Haupt.

Jacobs, Arthur M. (2019): "Sentiment Analysis for Words and Fiction Characters From the Perspective of Computational (Neuro-)Poetics", in: *Frontiers in Robotics and AI* 6, 53. DOI: 10.3389/frobt.2019.00053.

Řehůřek, Radim (2022): "Gensim: Topic Modelling for Humans", in: *Gensim* https://radimrehurek.com/gensim/auto_examples/tutorials/run_word2vec.html [27.10. 2022].

Russell, James A. (1980): "A Circumplex Model of Affect", in: *Journal of Personality and Social Psychology* 39, 6: 1161–78. DOI: 10.1037/h0077714.

Schumacher, Mareike / **Uglanova, Inna** / **Gius, Evelyn** (2022): "d-Romane-Romantik (d-RoRo)" in: *Zenodo*. DOI: 10.5281/ZENODO.7215170 [19.10.2022].

Yu, Liang-Chih / **Lee, Lung-Hao** / **Hao, Shuai** / **Wang, Jin** / **He, Yunchao** / **Hu, Jun** / **Lai, K. Robert** / **Xuejie, Zhang** (2016): "Building Chinese Affective Resources in Valence-Arousal Dimensions", in: Association for Computational Linguistics (ed.): *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, San Diego, California: 540–545. DOI: 10.18653/v1/N16-1066.