Detection and Analysis of Moralization Practices Across Languages and Domains

In this talk we propose a workflow for detecting and analyzing moralization practices in texts. Our **research question** is: Which linguistic features are characteristic for moralizing practices in different languages and different domains, and how can we detect and analyze those practices manually as well as automatically?

By **moralizing practices** we mean strategies in which moral values are utilized for describing controversial topics and demanding specific actions. Vocabulary that refers to positive or negative moral values (such as "freedom", "credibility", "cheating" or "inequality") is used to enforce a demand. This way, the demand is made to appear inescapable and requires no further explanation or justification. In the following example, which is taken from the protocols of the German parliament, the word "security" is used to support the demand for a cap for refugees: "We should introduce an upper limit for refugees to ensure the security of German citizens." [translation by the authors] Moralizing practices are widely used by many speakers and writers, e.g. in political speeches, online discussions or newspaper commentaries, and are an important discourse practice. Therefore, we propose an **approach for detecting and analyzing moralization practices** that is applicable to texts from different genres and domains as well as from different languages. It combines qualitative and qualitative methods and comprises the following steps:

I) Semi-automated creation of a multi-lingual dictionary with words that hint at moralizing practices. Our starting point is a manually created seed set of moral vocabulary in German. We expand this set by including words that frequently appear in similar contexts and are therefore likely to have similar meanings, using the co-occurrence database CCDB (Belica 2001) and manual post-filtering. Our final dictionary includes 2000 entries, which we automatically translate to English, French and Italian using DeepL and Google Translate. Afterwards we manually check and, if necessary, correct the translations.

II) Retrieval of potential moralization practices from texts. Next, we query large corpora and the web for text passages that include words from our dictionaries. We create parallel corpora for the four languages (EN, GER, FR, IT) with text passages from different domains such as political debates, newspaper articles, or online discussions.

III) Annotating texts and developing a model for detecting moralizing practices in texts. Using a word that expresses a moral value does not necessarily mean that the speaker/writer moralizes in the discourse-strategic sense as described above. We therefore manually categorize the retrieved text passages into instances that are moralizing practices and instances that refer to moral values in a descriptive manner. Subsequently, we use our annotated dataset as training data for BERT (Devlin et al. 2019), a neural transformer language model. We achieve high accuracy scores up to 75%, indicating the feasibility of automatically detecting moralizing practices in texts.

IV) Analyzing features of moralizing practices. Finally, we take a closer look at the instances that have been annotated as moralizing practices. We analyze them linguistically with respect to syntactic patterns, semantics and pragmatic functions. The result is a set of linguistic features that are helpful for characterizing, analyzing and deeply understanding moralizing practices.

Concluding, our **contributions** can be summarized as follows: (1) We provide a multi-lingual dictionary which can be used for detecting moralizing practices; (2) we release parallel datasets with texts from different languages and domains in which we annotate moralization practices; (3) we train a neural classifier that automatically detects moralization practices; (4) we provide a list of linguistic features which are characteristic for moralization practices, and compare the results across domains and languages; and (5) we provide a workflow for detecting and analyzing a complex, heterogenous and vague linguistic phenomenon, which can be adapted for similar phenomena.

References

Belica, Cyril: Kookkurrenzdatenbank CCDB. Eine korpuslinguistische Denk- und Experimentierplattform für die Erforschung und theoretische Begründung von systemisch-strukturellen Eigenschaften von Kohäsionsrelationen zwischen den Konstituenten des Sprachgebrauchs. © 2001 Institut für Deutsche Sprache, Mannheim.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.