
Automatic Detection of Sentiment in Luxembourgish User Comments

Daniela Gierschek

Université du Luxembourg

daniela.gierschek@uni.lu

Luxembourgish is a language that is mainly spoken by approximately 400,000 speakers (Steiner et al.) in the Grand Duchy of Luxembourg. Having become the national language in 1984, it coexists with the two other official languages, French and German, and has become an important symbol for national identity (Gilles). This project aims to advance the research in natural language processing for the Luxembourgish language in which only little work has been done so far. Combining techniques and knowledge from machine learning and linguistics, the project will use the user comments part of a large news corpus to implement the first sentiment analysis for texts written in Luxembourgish. On the machine learning level, a supervised approach will be the starting point. For this method, a part of the corpus was preprocessed and annotated with “positive”, “negative” and “neutral” labels. Further steps are training, testing and evaluation of the supervised system. Subsequently, the possible usage of word embeddings to improve sentiment detection shall be looked at. Besides techniques from machine learning, linguistics play an important role for the project. Challenges for natural language processing of Luxembourgish texts are especially the relatively low number of linguistic resources available and the high language variation on lexical and grammatical level (Gilles). As Luxembourgish is structural close to German (Gilles), one possibility to overcome those shortcomings could be to take existing linguistic resources for German language processing and translate and adapt them to Luxembourgish. Additionally, the role of parts of speech, especially adjectives, in expressing sentiment need to be examined. Adjectives are important for transmitting the subjective content of a text (Taboada) and therefore essential for a successful detection of sentiment.

References: • Gilles, Peter (in press): Luxembourgish. Maitz, Péter, Boas, Hans C., Deumert, Anna and Loudon, Marc, *Varieties of German Worldwide*, Oxford: Oxford University Press. • Steiner, Ingmar, Le Maguer, Sébastien, Manzoni, Judith, Gilles, Peter and Trouvain, Jürgen (2017): Developing new language tools for MaryTTS: the case of Luxembourgish. *Proceedings 28th Conference Elektronische Sprachsignalverarbeitung*, pp. 186-192. • Taboada, Maitte (2016): Sentiment analysis: An overview from linguistics. *Annual Review of Linguistics* 2, 325-347.