A Dialectometric Study of Low Saxon Syntactic Variation through Time

We present a corpus-based dialectometric study of synchronic and diachronic syntactic variation in literary Low Saxon, where we focus on aggregate similarity on the one hand and the occurrence of particular structures on the other. These results are then compared to our previous studies targeting other levels of representation as well as to findings from traditional dialectology. Our two major research questions for this study are: 1) Does the overall syntactic similarity of the dialect groups change over time and, if yes, how? 2) Do certain structures considered characteristic for Low Saxon decrease in frequency in the written language as well, as found in studies on spoken language?

The major part of our Modern Low Saxon data comes from the LSDC dataset [Siewert et al. (2020)] and is divided into two time periods (1800–1939 and 1980–2022) and six major dialect groups (Figure 1). Our findings will be compared to the Reference Corpus Middle Low German / Low Rhenish [REN-Team (2019)].

In addition to our own research, recent dialectometric studies of Low Saxon have appeared by, for instance, [Buurke et al. (2022)] and [Bartelds and Wieling (2022)]. A slightly older study is by [Lameli (2016)] who re-analysed the Wenker atlas data and found a north-south split in German Low Saxon.

In previous experiments, we have compared aggregate distances in Modern Low Saxon, Standard Dutch and Standard German at the levels of characters, PoS (Part-of-Speech) tags[^1] and morphological features from whole corpora. Here, we have found different trends at the different levels of representation. Whereas Dutch Low Saxon seems to approach Standard Dutch at all levels, the picture for German Low Saxon is more diverse: While we find a comparable trend of German Low Saxon approaching Standard German at the PoS level (Figure 2), when adding morphological information, the northern dialects appear to approach Standard Dutch (Figure 3). Furthermore, similar to Lameli, we find a north-south division in German Low Saxon to be more prominent than the traditionally assumed east-west division (compare, e.g., Schröder [2004]).

To complement our previous studies, we make use of syntactic relations that and lemmata to look at structures that PoS tags do not sufficiently differentiate. In addition to the aggregate similarity based on syntactic relations, we particularly want to investigate the occurrence of structures that according to [Elmentaler and Borchert (2012)] are often presented as characteristic for Low Saxon in textbooks and grammar books but which they have not found to be particularly frequent in the spoken language.

[^1]: PoS tags, morphol. features and syntactic relations explained here: https://universaldependencies.org/guidelines.html

Figure 1: Low Saxon dialects included in our study: NNS: Dutch North Saxon, NWF: Dutch Westphalian, DNS: German North Saxon, DWF: German Westphalian, MVP: Mecklenburgish – West Pomeranian, OFL: Eastphalian. Other dialects: BRA: Brandenburgish, POM: East Pomeranian, NPR: Low Prussian

REFERENCES


Martijn Bartelds and Martijn Wieling. Quantifying language variation acoustically with few resources, 2022.

