A Dialectometric Study of
Low Saxon Syntactic Variation through Time

Janine Siewert
University of Helsinki, Finland
janine.siewert@helsinki.fi

Paper Abstract

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).

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

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.
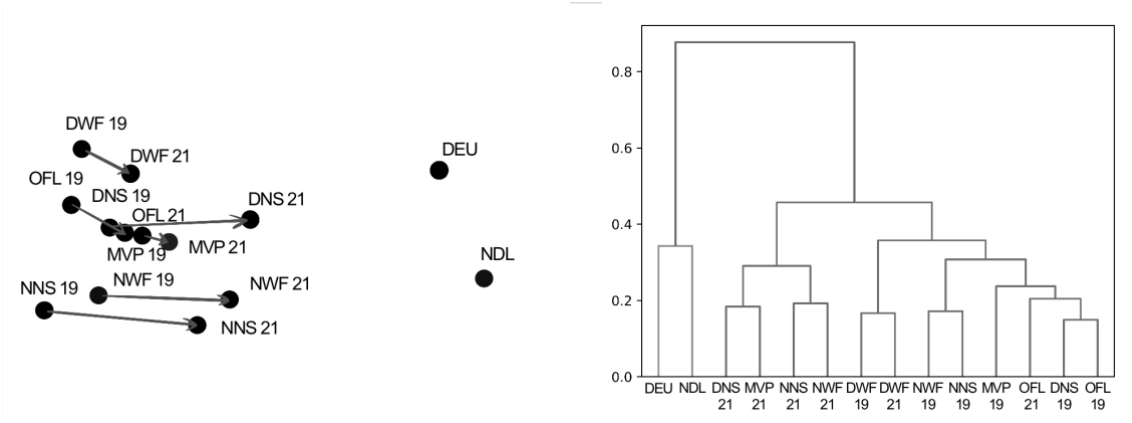
Figure 2. PoS level PCA and hierarchical clustering. NNS: Dutch North Saxon, NWF: Dutch Westphalian, DNS: German North Saxon, DWF: German Westphalian, MVP: Mecklenburgish – West Pomeranian, OFL: Eastphalian. 19: 1800–1939, 21: 1980–2022

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 and morphological features (explained here: https://universaldependencies.org/guidelines.html) 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).
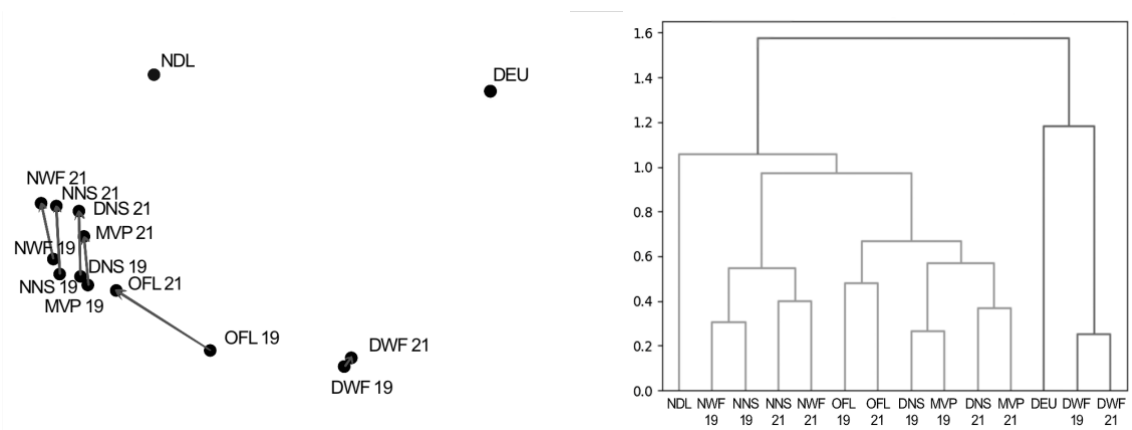

Figure 3. PoS and morphological features, PCA and hierarchical clustering. NNS: Dutch North Saxon, NWF: Dutch Westphalian, DNS: German North Saxon, DWF: German Westphalian, MVP: Mecklenburgish – West Pomeranian, OFL: Eastphalian. 19: 1800–1939, 21: 1980–2022

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.

REFERENCES

Bartelds, Martijn & Wieling, Martijn (2022). Quantifying language variation acoustically with few resources. In Marine Carpuat, Marie-Catherine de Marneffe & Ivan Vladimir Meza Ruiz (Eds.), *Proceedings of the 2022 Conference of the North American Chapter of the As-sociation for Computational Linguistics: Human Language Technologies* (3735-3741). Association for Computational Linguistics. doi: 10.18653/v1/2022.naacl-main.273.

Buurke, Raoul Sergio Samuel Jan, Sekeres, Hedwig G, Heeringa, Wilbert, Knooihuizen Remco & Wieling, Martijn (2022). Estimating the level and direction of aggregated sound change of dialects in the northern Netherlands. *Taal & Tongval, 74*(2), 183-214.

Elmentaler, Michael & Borchert, Felix (2012). Niederdeutsche Syntax im Spannungsfeld von Kodex und Sprachpraxis. In Robert Langhanke, Kristian Berg, Michael Elmentaler & Jörg Peters (Eds.), *Germanistische Linguistik 220 – Niederdeutsche Syntax* (101-135). Georg Olms Verlag.

Lameli, Alfred (2016). Raumstrukturen im Niederdeutschen. Eine Re-Analyse der Wenkerdaten. *Niederdeutsches Jahrbuch .Jahrbuch des Vereins für niederdeutsche Sprachforschung, 139*, 131-152.

ReN-Team (2019). *Referenzkorpus Mittelniederdeutsch/Niederrheinisch (1200-1650), 2019*. URL http://hdl.handle.net/11022/0000-0007-D829-8. Archived in Hamburger Zentrum für Sprachkorpora. Version 1.0. Publication date 2019-08-14.

Schröder, Ingrid (2004). Niederdeutsch in der Gegenwart: Sprachgebiet – Grammatisches – Binnendifferenzierung. In Dieter Stellmacher (Ed.), *Niederdeutsche Sprache und Literatur der Gegenwart* (35-97). Georg Olms Verlag.

Siewert, Janine, Scherrer, Yves, Wieling, Martijn & Tiedemann, Jörg (2020). LSDC - a comprehensive dataset for Low Saxon Dialect Classification. In *Proceedings of the 7th Workshop on NLP for Similar Languages, Varieties and Dialects* (25-35). International Committee on Computational Linguistics (ICCL).

Acknowledgements/Disclaimers