Changing usage of Low Saxon auxiliary and modal verbs

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

Martijn Wieling
University of Groningen
m.b.wieling@rug.nl

Yves Scherrer
University of Oslo
University of Helsinki
yves.scherrer@ifi.uio.no

Abstract

We investigate the usage of auxiliary and modal verbs in Low Saxon dialects from both Germany and the Netherlands based on word vectors, and compare developments in the modern language to Middle Low Saxon. Although most of these function words have not been affected by lexical replacement, changes in usage that likely at least partly result from contact with the state languages[1] can still be observed.

1 Introduction

Low Saxon[2] is an unstandardised West Germanic language primarily spoken in the north-eastern Netherlands and northern Germany. As the contact situation with the state languages Dutch and German has led to divergence of Low Saxon dialects at the border, the primary research question we want to investigate is whether the usage of certain auxiliary and modal verbs can also be found to diverge.

This study is part of our broader research into dialectal variation and change in Low Saxon, cf. Siewert et al. (2022), and constitutes a first exploration of the field of lexical variation. Auxiliary and modal verbs are a suitable starting point because they form a relatively closed group for which automatic annotation works more reliably than for many others.

We use word vector representations to compare certain auxiliary and modal verbs and investigate changes in usage from Middle Low Saxon to Modern Dutch Low Saxon and German Low Saxon. These vectors representations were trained on lemmata in concatenation with dependency relations and PoS (Part-of-Speech) information.

2 Background

The divergence of Low Saxon dialects at the border has been investigated in the form of lexical replacement as well as changes at the phonological, morphological and syntactic level, e.g., by Niebaum (1990) and Kremer (1990). A more quantitative study looking at frequencies of local phonological, morphological and syntactic traits in contrast with state language traits is presented by Smits (2009), who examines the stability of dialectal characteristics. All three authors mention the lexical level as an area particularly susceptible to influence from the state languages. Instead of lexical replacement mostly referred to by them, we will however focus on changing usage of the same lexical items.

2.1 Auxiliary and modal verbs

The auxiliary and modal verbs included in the comparison are *dôn* 'to do', *dȫren* 'to dare', *dörven* 'to dare, to be allowed to', *hebben* 'to have', *künnen* 'can', *mȫgen* 'may, like', *môten* 'must', *schȫlen* 'shall, will', *wērden* 'to be (+ past participle), will', *wēsen* 'to be' and *willen* 'want, will'[3]. In particular, we will focus on two groups of auxiliary or modal verbs that exhibit partly overlapping usage: future auxiliaries on the one hand and models of permission, prohibition and obligation on the other hand.

2.2 Future auxiliaries

The first group consists of the verbs *wērden*, *schȫlen* and *willen*. While *wērden*, like its Dutch and German cognates *worden* and *werden*, has traditionally functioned as the auxiliary verb for forming the passive, it has developed the additional

---

[1]'State languages' refers to Standard Dutch and Standard German here, because they are the only languages with state-wide official status in the respective countries. Contact with regional official languages, such as the Frisian languages or Danish, is not taken into account here although this would certainly be an interesting research question as well.

[2]Also called 'Low German'.

[3]The English cognates are *do, dare, tharf*[†]*, have, can, may, must, shall, worth*[†]*, be* and *will*, with the forms marked with a [†] being dialectal or historical. The translations represent some common usages today. Due to the internal diversity and change over time, it is not possible to provide translations covering all varieties and time periods here.

function of the future auxiliary in German. A similar development can be observed in German Low Saxon with first attestations already in Middle Low Saxon (Härd, 2000, 1458), but in older Modern Low Saxon texts, an inchoative reading is often still possible or the more likely interpretation (cf. Lindow et al., 1998, 101–103). In Dutch Low Saxon, on the other hand, we have not encountered usage of *wērden* as an auxiliary for the future tense. Therefore, we expect to see differences in the distance of *wērden* to *schōlen* and *willen*, of which *schōlen* already functioned as a future auxiliary in Middle Low Saxon (Härd, 2000, 1458) and can still do so in both Dutch Low Saxon and German Low Saxon (Lindow et al., 1998, 106). The usage of *willen* as a future auxiliary in German Low Saxon is described at least by Lindow et al. (1998, 104).

2.3 Modals of permission, prohibition and obligation

In the second group, we look at the distance of *dörven* to *dōren*, *môten* and *mōgen*. The verb *dōren* is especially interesting, because in Modern Low Saxon it has generally been either replaced by or merged with *dörven*. According to Lindow et al. (1998, 110), *dörven* originally carried the meaning 'to be allowed to', while *dōren* meant 'to dare', and these meanings are to varying degree found in *dörven* the modern language.

While negated *müssen* in German carries the meaning 'does not need to', negated *môten* in German Low Saxon can be used like the English equivalent *must not* (Lindow et al., 1998, 110). This usage is similar to negated *dörven*.

The main usages of *mōgen* in German Low Saxon according to Lindow et al. (1998, 112) are the expression of possibility, of an assumption and of a wish. These meanings can be found in German and Dutch as well, but they differ in which meanings dominate.

Since we have not found comparable descriptions for the Dutch Low Saxon verbs, our expectations are mostly based on the corresponding usage in Dutch and our own exposure to Dutch Low Saxon.

3 Data

The Modern Low Saxon data shown in Table 1 comes from the LSDC dataset (Siewert et al., 2020) and is split into two time periods: 1800–1939 and 1980–2022. Furthermore, we split the dataset into

Abbr.	Variety	Time span	Tokens
MLS	Middle Low Saxon	1200–1650	1 406 979
DLS1	Dutch Low Saxon	1800–1939	147 212
DLS2	Dutch Low Saxon	1980–2022	393 619
NLS1	German North Low Saxon	1800–1939	1 008 851
NLS2	German North Low Saxon	1980–2022	103 568
SLS1	German South Low Saxon	1800–1939	371 611
SLS2	German South Low Saxon	1980–2022	416 686

Table 1: Low Saxon varieties and their token counts.

Figure 1: The three major Low Saxon dialect groups included.

three large geographical groups: Dutch Low Saxon (DLS), German North Low Saxon (NLS) and German South Low Saxon (SLS) as shown in Figure 1. All subcorpora contain a variety of genres, among others short stories, fairy tales, theatre plays, historical accounts, speeches, and letters.

The Middle Low Saxon (MLS) data is taken from the Reference Corpus Middle Low German / Low Rhenish (ReN-Team, 2021), and converted to CoNLL-U format including a conversion of the tags to the UD tagset[4] that is used in the LSDC dataset. The genres in the Reference Corpus are specified as prose, document, or verse.

3.1 Annotation

For this research, three layers of annotation are relevant: Lemmatisation, PoS tagging and dependency parsing.

The LSDC dataset comes with PoS tags, but does not include lemmata or dependency relations. The PoS tags are primarily annotated automatically, except for the around 300 sentences per dialect group that were manually corrected for finetuning annotation models.

The lemmata and PoS tags in the original version of the Reference Corpus have been annotated

---

[4] https://universaldependencies.org/u/pos

by human curators, but we needed to make some adaptations and add dependency parsing.

3.1.1 Lemmatisation

For comparison with the Reference Corpus, we needed to lemmatise Modern Low Saxon to Middle Low Saxon. Our lemmata follow the *Mittelniederdeutsches Handwörterbuch* (Lasch et al., 1928 ff.), but we removed superscript numbers and simplified a few graphemes, such as <êⁱ> to <êi>, to speed up the manual lemmatisation of the training, development and test data for the lemmatiser. We furthermore slightly manually adapted the lemmata in the Reference Corpus in the same way as for the modern corpus.

We manually lemmatised these same around 900 sentences which contained gold standard PoS tags in order to train a lemmatiser. Of these, 700 were part of the train set and 100 each formed the development and the test set.

We trained a Stanza (Qi et al., 2020) lemmatiser on a train set that contained the whole Reference Corpus in addition to our small manually annotated Modern Low Saxon training data, whereas we only used Modern Low Saxon data for the development and test set. We reached an accuracy of 83% and lemmatised the remainder of the LSDC data with this model.

3.1.2 Dependency parsing

Due to time constraints, we only managed to annotate dependency relations for around 300 sentences of which 100 became part of the train set.

We used Stanza for dependency parsing[5] as well and complemented the small manually annotated Low Saxon train set with UD datasets in Afrikaans[6], Danish (Johannsen et al., 2015), Dutch (Bouma and van Noord, 2017), English (Zeldes, 2017), German (McDonald et al., 2013), Norwegian (Øvrelid and Hohle, 2016) and Swedish (Nivre and Megyesi, 2007). We included the mainland Scandinavian languages in addition to the West Germanic ones, because they were in close contact with and strongly influenced by Middle Low Saxon during the time of the Hanseatic League. Since Stanza does not allow for finetuning, the train set included all eight languages while the development and test set contained exclusively Low Saxon data. This parser reached an accuracy of 81% LAS for

Modern Low Saxon and was used to parse both the Modern and the Middle Low Saxon corpus, but since it has only encountered Modern Low Saxon data during training, the parsing accuracy on Middle Low Saxon is likely lower.

The lemmatised and dependency-parsed Modern Low Saxon data is publicly available under a CC BY-NC license[7].

4 Methods

The word vectors were trained on the whole dataset – both the manually and the automatically annotated part – using fastText's (Bojanowski et al., 2016) skipgram model with a vector length of 100 and subwords[8] following these two set-ups: lemma + dependency relation (e.g., *dörven_aux*), and lemma + PoS tags (e.g. *wērden_AUX*). Our reason for using subwords during training is that, otherwise, the PoS or dependency information, that is part of the same string, could not be accessed.

Levy and Goldberg (2014) found dependency information to be beneficial for identifying words that behave in a similar way and not only occur in similar contexts. For comparison, we used PoS tags, because the PoS tagging in our dataset is more accurate than the dependency relations.

We initially also tested vectors based on lemmata only, but eventually excluded these, since they showed great fluctuations even within the same variety, when the vectors were trained with different mininum word counts. Furthermore, when working with fastText's function `.get_nearest_neighbors()`, we had observed that the suggested nearest neighbours tended to be more meaningful when dependency or PoS information was added, as otherwise the importance of uninformative subword units such as *nnen* or *llen* seemed to be overestimated.

We first trained common vectors with a mininum word count of 50 for both Middle and Modern Low Saxon to ensure a common initialisation for all variants. Subsequently we fine-tuned this model on Middle Low Saxon and Modern Low Saxon data separately, and finally, with a mininum word count of 25, retrained the general Modern Low Saxon model with data in the subgroups listed in table 1.

Due to the small size of the two subcorpora DLS 1 and NLS 2 (cf. Table 1), we also trained vectors

---

[5] https://universaldependencies.org/u/dep

[6] https://github.com/UniversalDependencies/UD_Afrikaans-AfriBooms/tree/master

[7] https://github.com/Helsinki-NLP/LSDC-morph/tree/main/lchange2023

[8] See training options here: https://fasttext.cc/docs/en/unsupervised-tutorial.html

with mininum counts of 5, 10 and 15 to check the stability of the results.

Subsequently, we used the Python libraries NumPy (Harris et al., 2020) and SciPy (Virtanen et al., 2020) to measure the Euclidean and cosine distances between the resulting word vectors. We will, however, only present the results based on Euclidean distance here, since the other approach produced comparable results.

Despite the common initialisation, the absolute distance values did not compare well across varieties. The reason for this might be found in the different sizes of the subcorpora. Therefore, we only discuss the relative closeness compared with the other modal or auxiliary verbs here.

5 Results

5.1 Future auxiliaries

We use *wērden* as a target verb and list the other auxiliary and modal verbs in order of closeness to *wērden* in Tables 2 and 3.

In both tables, we see that in German Low Saxon, NLS and SLS, *schōlen* is closer to *wērden* than in Middle Low Saxon and Dutch Low Saxon. Curiously, these verbs seem to grow closer in Dutch Low Saxon, but due to the small size of the DLS 1 corpus, one should not draw strong conclusions from this. The increase we see in German South Low Saxon in both tables is likely more reliable. For German North Low Saxon we find contradicting tendencies: While the dependency-based table shows continuity, we find a decrease in closeness in the PoS-based data.

Strikingly, while *willen* was still clearly closer to *wērden* than *schōlen* in Middle Low Saxon, the order has shifted in the modern language and *willen* has become less similar almost without exception.

5.2 Modals of permission, prohibition and obligation

Tables 4 and 5 present the verbs ordered by closeness to the target verb *dörven*. As mentioned in Section 2, *dōren* has mostly fallen out of use in the Modern Low Saxon period and is only represented by a handful of examples. As a result, the word vectors are largely inherited from the common pretrained vectors. Furthermore, the verb *dörven* has very few occurrences in the Dutch Low Saxon data. Therefore, vectors of this verb likely represent mostly the common Low Saxon pretrain-model.

The verb *dōren* is very close to *dörven* in Middle Low Saxon and Dutch Low Saxon[9], whereas the picture is less consistent in German Low Saxon: While closeness is high in the NLS 2 data, it is only the fourth or fifth most similar verb in the NLS 1 data. The number of occurrences of *dörven*, however, is small (only 10) in the newer data and, therefore, less reliable. Similarly, we observe a decrease in South Low Saxon, particularly in the dependency-based data.

The other verb that shows a contrasting development in Dutch Low Saxon and German Low Saxon is *mōgen*. Curiously, while the similarity compared to Middle Low Saxon seems to increase in Dutch Low Saxon in the dependency data, a decrease appears to occur in the PoS-based data. Nevertheless, in both cases the relative closeness is greater than in German Low Saxon. The only exception to this seems to be newer North Low Saxon (NLS 2) in table 5, but, in fact, the vectors trained with a smaller mininum word count showed a greater distance.

In case of *môten*, we find a contrast between Dutch Low Saxon and German North Low Saxon on the one hand, and German South Low Saxon on the other hand: Whereas in German South Low Saxon, the closeness to *dörven* remains comparable to Middle Low Saxon over both time periods, the other two modern varieties show a decrease in both tables.

6 Discussion and future research

For *wērden*, we found partly expected and partly surprising results. The increased closeness of *schōlen* in German Low Saxon is in line with the development of *wērden* into a future tense auxiliary. The slight increase we see in German South Low Saxon when going from the older to the modern period might tell that this additional usage of *wērden* was not as widespread yet in the 19th and early 20th century.

On the other hand, we do not have an explanation for the decreased closeness of *willen*. However, at least for modern German Low Saxon, the greater distance might show that the usage of *willen* as a future auxiliary is in fact not very widespread.

While the similarity between *dörven* and *dōren*

---

[9]Due to the small number of occurrences, the Dutch Low Saxon vectors might represent mostly a copy of Middle Low Saxon.

MLS	DLS1	DLS2	NLS1	NLS2	SLS1	SLS2
wēsen	wēsen	wēsen	wēsen	wēsen	môten	wēsen
hebben	dörven	dŏren	**schŏlen**	**schŏlen**	künnen	môten
künnen	dŏren	dörven	dörven	dörven	dörven	**schŏlen**
willen	mögen	môten	môten	künnen	wēsen	mögen
môten	môten	mögen	künnen	dŏren	dŏren	dörven
dŏren	hebben	künnen	dŏren	hebben	**schŏlen**	hebben
dôn	künnen	dôn	hebben	môten	dôn	künnen
schŏlen	dôn	**schŏlen**	willen	willen	mögen	dŏren
mögen	**schŏlen**	willen	mögen	dôn	willen	willen
dörven	willen	hebben	dôn	mögen	hebben	dôn

Table 2: Auxiliar and modal verbs most similar to *wērden*, with dependency relation.

MLS	DLS1	DLS2	NLS1	NLS2	SLS1	SLS2
wēsen	hebben	wēsen	wēsen	wēsen	dŏren	wēsen
dôn	dŏren	dŏren	**schŏlen**	künnen	wēsen	dörven
willen	mögen	dörven	hebben	dŏren	môten	hebben
hebben	wēsen	dôn	môten	hebben	künnen	mögen
mögen	dörven	hebben	dŏren	**schŏlen**	dörven	**schŏlen**
dŏren	môten	môten	künnen	dörven	**schŏlen**	willen
schŏlen	künnen	**schŏlen**	dörven	dôn	dôn	dôn
môten	dôn	künnen	willen	mögen	hebben	môten
künnen	willen	mögen	dôn	willen	mögen	künnen
dörven	**schŏlen**	willen	mögen	môten	willen	dŏren

Table 3: Auxiliar and modal verbs most similar to *wērden*, with PoS information.

MLS	DLS1	DLS2	NLS1	NLS2	SLS1	SLS2
môten	môten	**dŏren**	künnen	schŏlen	schŏlen	môten
dŏren	**dŏren**	*mögen*	môten	**dŏren**	môten	willen
willen	*mögen*	willen	schŏlen	künnen	willen	künnen
mögen	künnen	môten	willen	willen	**dŏren**	dôn
schŏlen	dôn	künnen	**dŏren**	môten	künnen	**dŏren**
künnen	wērden	schŏlen	hebben	hebben	wēsen	wēsen
hebben	schŏlen	dôn	wēsen	*mögen*	dôn	hebben
dôn	hebben	wēsen	*mögen*	dôn	hebben	schŏlen
wēsen	wēsen	hebben	dôn	wēsen	*mögen*	wērden
wērden	willen	wērden	wērden	wērden	wērden	*mögen*

Table 4: Auxiliary and modal verbs closest to *dörven* based on lemmata with dependency relations.

MLS	DLS1	DLS2	NLS1	NLS2	SLS1	SLS2
dŏren	**dŏren**	**dŏren**	willen	**dŏren**	môten	môten
môten	wēsen	willen	hebben	schŏlen	**dŏren**	**dŏren**
mögen	wērden	schŏlen	**môten**	hebben	künnen	künnen
willen	*mögen*	dôn	**dŏren**	*mögen*	dôn	dôn
künnen	**môten**	*mögen*	schŏlen	willen	wērden	willen
schŏlen	künnen	hebben	künnen	künnen	schŏlen	wēsen
dôn	hebben	künnen	dôn	wēsen	wēsen	hebben
hebben	dôn	wēsen	wēsen	dôn	*mögen*	schŏlen
wērden	willen	môten	wērden	**môten**	willen	*mögen*
wēsen	schŏlen	wērden	*mögen*	wērden	hebben	wērden

Table 5: Auxiliary and modal verbs closest to *dörven* based on lemmata with PoS information.

in Dutch Low Saxon cannot be judged reliably due to data sparsity, we see an interesting decrease in German Low Saxon. This might be related to the usage of German *dürfen*, which generally does not carry the meaning 'to dare'.

A similar phenomenon can apparently be observed in case of *mögen*. While the meaning of 'to be allowed to' is still dominant in Dutch *mogen*, it has become less common for German *mögen*, and the distances we find in Table 4 and 5 suggest that the Low Saxon varieties might again follow the state languages.

Moreover, a shift in the usage of negated *môten* from 'must not / to not be allowed to' to 'do not need to' as in German might explain the decreased similarity in NLS.

In conclusion, we find that lexical change and divergence at the border is not only visible in the form of lexical replacement, but also at the level of word usage. For some of the developments described above, such as the increased closeness of *wērden* and *schŏlen* and the decreased closeness of *dörven* and *mögen* in German Low Saxon, state language influence likely plays a role.

6.1 Future Research

In order to increase the reliability of our results, we want to further improve dependency parsing accuracy. In particular, separate train, development and test data for Middle Low Saxon dependency parsing would be desirable.

We plan to also use these dependency relations for the detection of syntactic structures and a comparison of dialect similarity and change at the syntactic level, since this is often considered more stable than the lexical level.

Moreover, the reference corpus contains metadata information on place and time, so one might take a look at the internal variation of Middle Low Saxon as well.

Computing overall differences of the modal verbs to their Dutch and German cognates would be an interesting research direction as well, but on the one hand this might require subcorpora of more equal size, as discussed in Section 4 and on the other hand – and even more importantly – comparable Dutch and German corpora from the same time periods and with the same annotation.

Limitations

The Middle Low Saxon reference corpus does not cover the Netherlands, so our dataset does not include the predecessor to Modern Dutch Low Saxon. Unfortunately, to our knowledge, there is no reference corpus for the Middle Low Saxon varieties from today's Dutch side of the border.

Unlike in the Middle Low Saxon data, the lemmatisation and PoS tags of the modern data are

not gold-standard, and the dependency parsing was done fully automatically for both. This needs to be kept in mind when judging the reliability of the results.

Acknowledgements

References

Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2016. Enriching word vectors with subword information. *arXiv preprint arXiv:1607.04606*.

Gosse Bouma and Gertjan van Noord. 2017. Increasing return on annotation investment: The automatic construction of a Universal Dependency treebank for Dutch. In *Proceedings of the NoDaLiDa 2017 Workshop on Universal Dependencies (UDW 2017)*, pages 19–26, Gothenburg, Sweden. Association for Computational Linguistics.

Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. 2020. Array programming with NumPy. *Nature*, 585(7825):357–362.

John Evert Härd. 2000. Syntax des Mittel-niederdeutschen. In Werner Besch, Anne Betten, Oskar Reichmann, and Stefan Sonderegger, editors, *Sprachgeschichte – Ein Handbuch zur Geschichte der deutschen Sprache und ihrer Erforschung*. Walter de Gruyter, Berlin and New York.

Anders Johannsen, Héctor Martínez Alonso, and Barbara Plank. 2015. Universal dependencies for danish. In *International Workshop on Treebanks and Linguistic Theories (TLT14)*, page 157.

Ludger Kremer. 1990. Kontinuum oder Bruchstelle? Zur Entwicklung der Grenzdialekte zwischen Niederrhein und Vechtegebiet. In Ludger Kremer and Hermann Niebaum, editors, *Germanistische Linguistisk 101-103*, pages 85–123. Olms.

Agathe Lasch, Conrad Borchling, Gerhard Cordes, Dieter Möhn, Ingrid Schröder, Jürgen Meier, and Sabina Tsapaeva. 1928 ff. *Mittelniederdeutsches Handwörterbuch*. Wachholtz Verlag, Neumünster.

Omer Levy and Yoav Goldberg. 2014. Dependency-based word embeddings. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 302–308, Baltimore, Maryland. Association for Computational Linguistics.

Wolfgang Lindow, Dieter Möhn, Hermann Niebaum, Dieter Stellmacher, Hans Taubken, and Jan Wirrer. 1998. *Niederdeutsche Grammatik*. Verlag Schuster Leer.

Ryan McDonald, Joakim Nivre, Yvonne Quirmbach-Brundage, Yoav Goldberg, Dipanjan Das, Kuzman Ganchev, Keith Hall, Slav Petrov, Hao Zhang, Oscar Täckström, Claudia Bedini, Núria Bertomeu Castelló, and Jungmee Lee. 2013. Universal Dependency annotation for multilingual parsing. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 92–97, Sofia, Bulgaria. Association for Computational Linguistics.

Hermann Niebaum. 1990. Staatsgrenze als Bruchstelle? Die Grenzdialekte zwischen Dollart und Vechtegebiet. In Ludger Kremer and Hermann Niebaum, editors, *Germanistische Linguistisk 101-103*, pages 49–83. Olms.

Joakim Nivre and Beata Megyesi. 2007. Bootstrapping a swedish treebank using cross-corpus harmonization and annotation projection. In *Proceedings of the 6th international workshop on treebanks and linguistic theories*, pages 97–102. Association for Computational Linguistics Pennsylvania, PA.

Peng Qi, Yuhao Zhang, Yuhui Zhang, Jason Bolton, and Christopher D. Manning. 2020. Stanza: A Python natural language processing toolkit for many human languages. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*.

ReN-Team. 2021. Reference Corpus Middle Low German/Low Rhenish (1200–1650); Referenzkorpus Mittelniederdeutsch/Niederrheinisch (1200–1650).

Janine Siewert, Yves Scherrer, and Martijn Wieling. 2022. Low Saxon dialect distances at the orthographic and syntactic level. In *Proceedings of the 3rd Workshop on Computational Approaches to Historical Language Change*, pages 119–124, Dublin, Ireland. Association for Computational Linguistics.

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

Tom F.H. Smits. 2009. Prinzipien der Dialektresistenz – Zur Bestimmung einer dialektalen Abbauhierarchie. In Alexandra N. Lenz, Charlotte Gooskens, and Siemon Reker, editors, *Low Saxon Dialects across Borders – Niedersächsische Dialekte über Grenzen hinweg*, pages 317–338. Franz Steiner Verlag.

Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. 2020. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272.

Amir Zeldes. 2017. The GUM corpus: Creating multilayer resources in the classroom. *Language Resources and Evaluation*, 51(3):581–612.

Lilja Øvrelid and Petter Hohle. 2016. Universal dependencies for norwegian. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, Paris, France. European Language Resources Association (ELRA).