



university of
 groningen



UNIVERSITY OF HELSINKI

Computationally modeling Low Saxon variation at different linguistic levels

Raoul Buurke **Janine Siewert**

University of {Groningen, Helsinki}

Overview



1. Low Saxon

2. Variation in Written Low Saxon

3. Pronunciation variation in the Netherlands

4. Conclusion



Low Saxon

Low Saxon Dialects



Low Saxon



- Where is it spoken
- Around 4 million speakers
- No interregional (written) standard
- Differences the Netherlands and Germany
 - School subject in some states of Germany
 - Bachelor's degree program at the University of Oldenburg



Variation in Written Low Saxon

Corpus



dialect	abbr	group	abbr	time	sent	token
Dutch North Saxon	NNS	Dutch	DLS	1800-1939	1.830	44.117
		Low		1980-2022	17.659	259.233
Dutch Westphalian	NWF	Saxon		1800-1939	4.960	103.093
				1980-2022	9.113	134.388
German North Saxon	DNS	German	NLS	1800-1939	23.113	429.521
		North		1980-2022	3.525	61.130
Mecklenburgish	MVP	Low		1800-1939	20.148	579.349
West Pommeranian		Saxon		1980-2022	3.055	42.438
German Westphalian	DWF	German	SLS	1800-1939	17.493	338.730
		South		1980-2022	15.937	272.071
Eastphalian	OFL	Low		1800-1939	1.674	32.888
		Saxon		1980-2022	8.383	144.624
Middle Low Saxon			MLS	1200-1650	171.890	1.406.956

CoNLL-U format



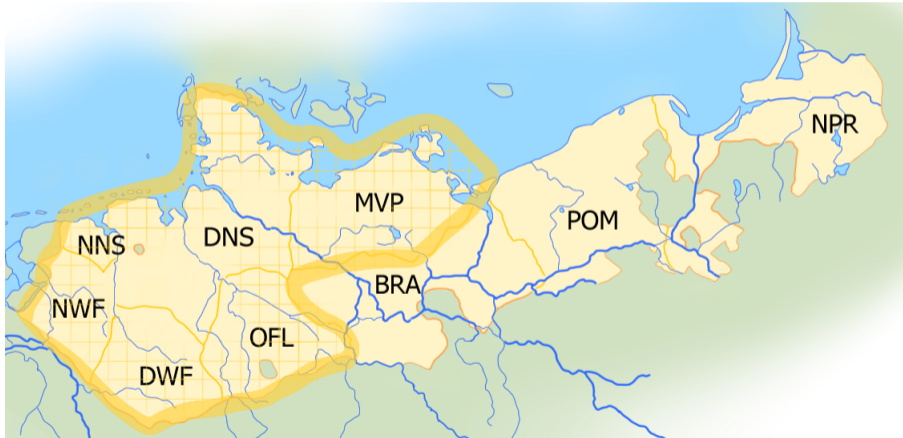
```
# sent_id = LSDC_1980_00142_NNS
# text = Joa, dat wait ik ook wel.
1 Joa jâ INTJ _ _ 4 discourse _ _
2 , , PUNCT _ _ 1 punct _ _
3 dat dê,dê,dâ PRON _ Case=Acc|Gender=Neut|Person=3|PronType=Dem 4 obj _ _
4 wait wêten VERB _ Mood=Ind|Number=Sing|Person=1|Tense=Pres 0 root _ _
5 ik ik PRON _ Case=Nom|Number=Sing|Person=1|PronType=Prs 4 nsubj _ _
6 ook ôk ADV _ _ 4 advmod _ _
7 wel wol ADV _ _ 4 advmod _ _
8 . . PUNCT _ _ 4 punct _ _
```


Methods

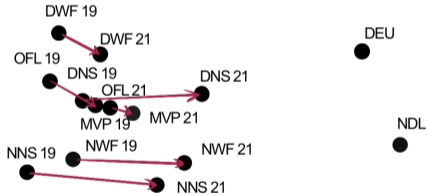


- Aggregate dialect distances
 - Bigram and trigram counts of PoS tags and morphological features
 - Tf-idf normalisation
 - PCA and hierarchical clustering (Python library Scikit-learn)
- Similarity of modal and auxiliary verbs
 - Lemma forms with PoS tag or dependency relation, e.g. *wērdēn_aux*
 - Word vectors trained with fastText's skipgram model with subword units
 - Euclidean distance between word vectors

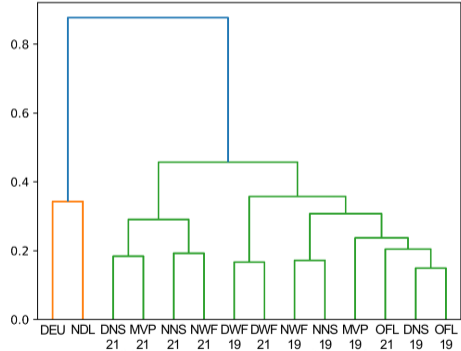
N-Gram-Based Dialect Distances



Results – PoS Distances

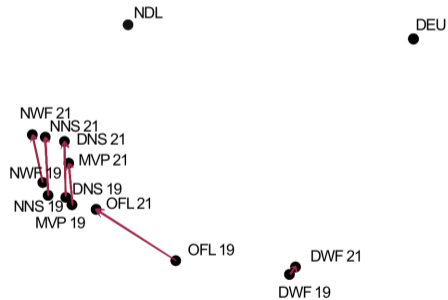


PCA

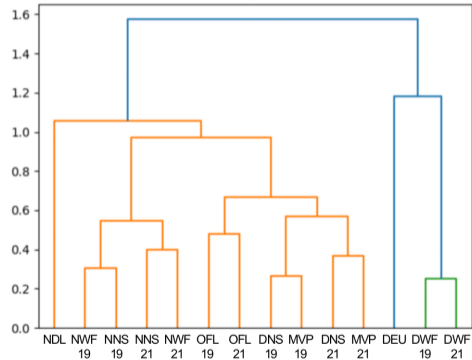


Hierarchical

Results – PoS + Morph. Distances



PCA



Hierarchical

Similarity of Modal and Auxiliary Verbs



Results – wēerden, with dependency relation



MLS	DLS1	DLS2	NLS1	NLS2	SLS1	SLS2
wēsen	wēsen	wēsen	wēsen	wēsen	mōten	wēsen
hebben	dōrven	dōren	schōlen	schōlen	kūnnen	mten
kūnnen	dōren	dōrven	dōrven	dōrven	dōrven	schōlen
<u>willen</u>	mōgen	mten	mten	kūnnen	wēsen	mōgen
mten	mten	mōgen	kūnnen	dōren	dōren	dōrven
dōren	hebben	kūnnen	dōren	hebben	schōlen	hebben
dōn	kūnnen	dōn	hebben	mten	dōn	kūnnen
schōlen	dōn	schōlen	<u>willen</u>	<u>willen</u>	mōgen	dōren
mōgen	schōlen	<u>willen</u>	mōgen	dōn	<u>willen</u>	<u>willen</u>
dōrven	<u>willen</u>	hebben	dōn	mōgen	hebben	dōn



Pronunciation variation in the Netherlands

Data collection



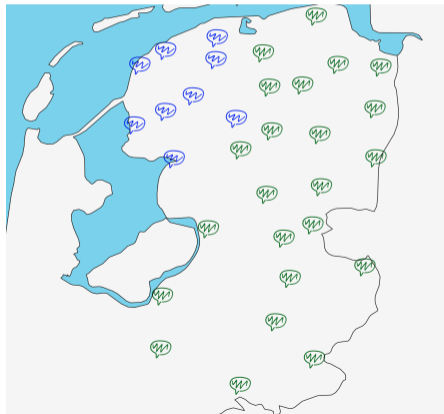
Data collection

- Mobile laboratory
- Word translation task (150 words)
- 32 locations, 23 Low Saxon



Buurke & Siewert

Recording locations



Method

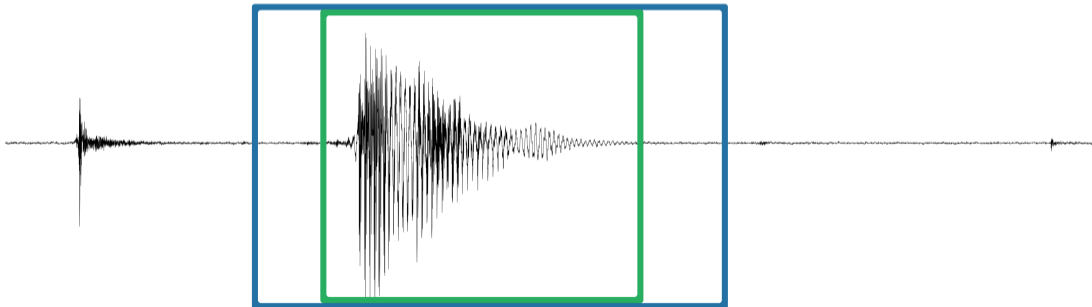


Voice activity detection

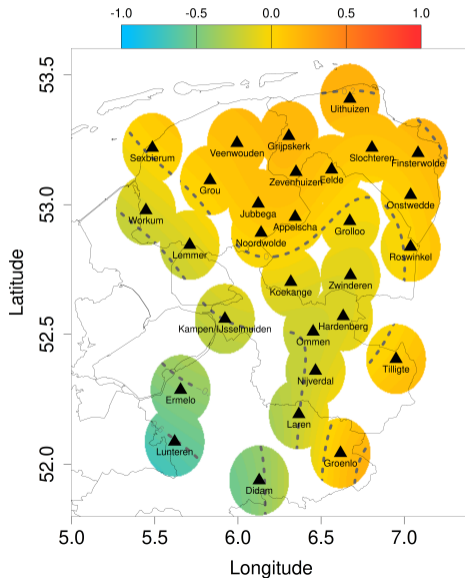
- Pyannote
- SoX

Neural acoustic distance

- Neural feature representations
- Dynamic time warping



Results





Conclusion

Conclusion



- North-south split in Low Saxon
 - Dutch Low Saxon as part of the northern group
- At the sound level, there are further splits in the Netherlands
 - Influence of Standard Dutch
- Attempts to maximize 'dialectalness'
 - Gronings accent
 - German Low Saxon speakers approximating Dutch forms and vice versa?
 - Language activism?



university of
 groningen



UNIVERSITY OF HELSINKI

Dank veur t luustern!

Raoul Buurke Janine Siewert

University of {Groningen, Helsinki}

Writing traditions



NWF: Ziene olders hadden altied hard ewarkt en wazzen gezene leu in den naoberschop.

NNS: Daor, kiek man ijs goud, 't kan best wezen, dat 't nog familie van die is.

DNS: Arfest neem twe Kaarten to de eerst Klaß, un as ik daröver grote Ogen maak, lach he un meen, dat kunn darop staan, ik schull man instigen.

BRA: Unn so wo de Doot dat den Fischer vertellt hett, isset ook ekâmen; dat ganze Dörp is uutstorven, man de Fischer is aarbliiwen unn issen riiken riiken Mann wâren, [...]

OFL: Ik kann nich sä güt wiet lupen un dorumme schölle mik miene Fründin hier ne Parkbuchte friehulen.

DWF: Eunige Dage später frogere de Magister, biu de veuer Johrestyien herren: Hiärmen sprank op, un de Magister mennte all, hai härr' et wieten.

German Low Saxon dialects according to Lameli



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