

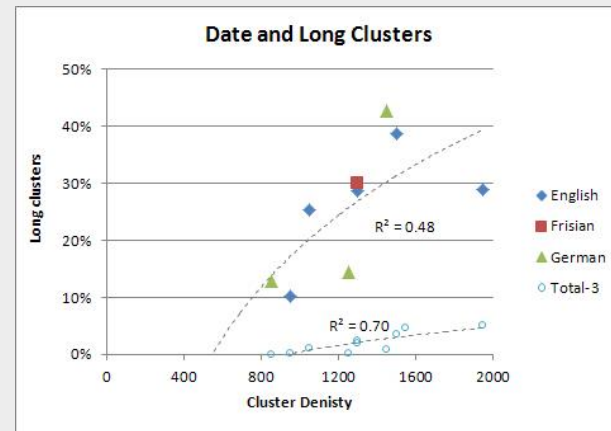
Modeling a historical variety of a low-resource language

Language contact effects in the verbal cluster of Early-Modern Frisian

Jelke Bloem, Arjen Versloot, Fred Weerman
j.bloem@uva.nl

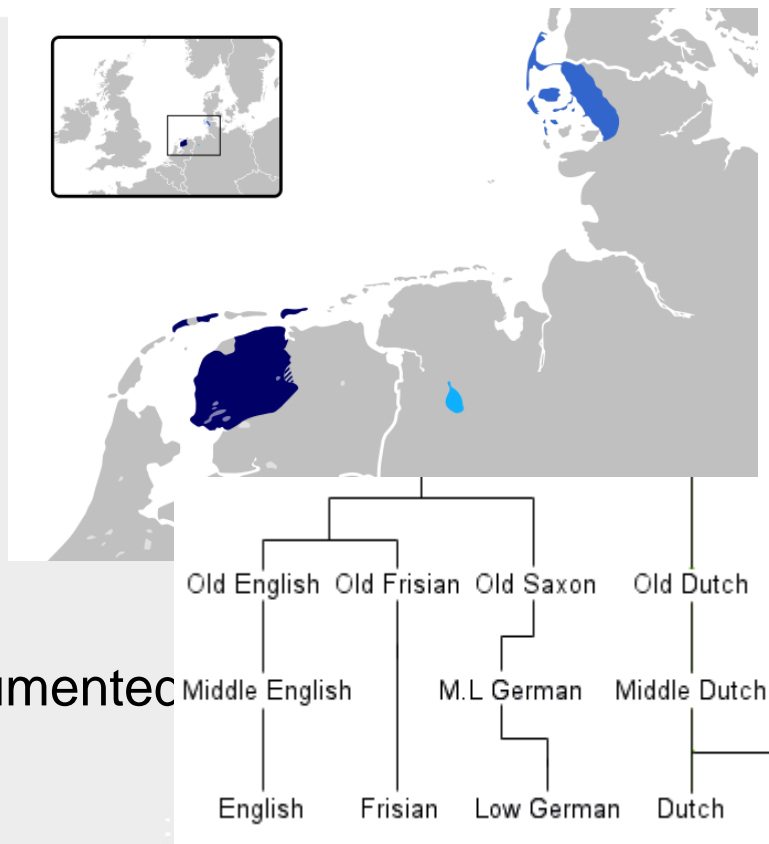
Why look at this?

- Can't study contact phenomena with the BNC
 - More likely to occur in lesser-used language (Weinreich, 1953)
- Can't study language change with the BNC
 - Easier to study a change after it happened
- Most languages are 'low-resource'



Frisian and its contacts

- West-Germanic language, spoken along the North Sea coast
- 400.000-500.000 native speakers
- Substrate language to
 - Middle Low German (14th-16th century)
 - Middle and Modern Dutch
- Limited written tradition, but well-documented
- Low-resource language



Case study: Frisian verb clusters

Anne sei dat er my sjoen₂ hie₁

Anne said that he me seen₂ had₁

‘Anne said that he had seen me’

- Verbs considered to have fixed word order

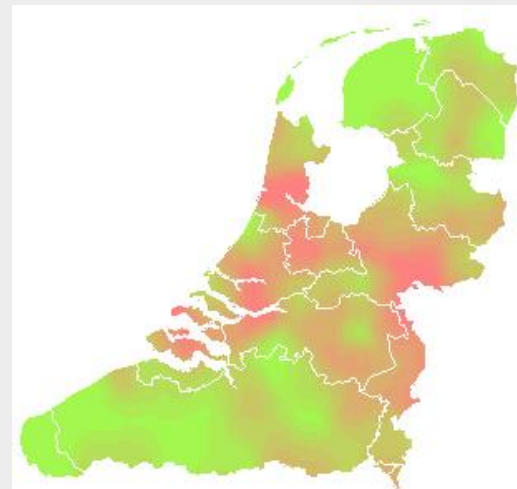
Het Nederlands kent in enkele gevallen een zekere vrijheid van plaatsing.
... dat hij zou komen / komen zou
Deze vrijheid is in het Fries niet aanwezig.

(Prisma grammar of Frisian)

- However, speakers have started using both orders (de Haan, 1996)
-> Recent language contact!

Dutch verb clusters

- Dutch two-verb clusters: Free order variation
 1. ik denk dat ik het begrepen₂ heb₁
 I think that I it understood₂ have₁
 2. ik denk dat ik het heb₁ begrepen₂
 I think that I it have₁ understood₂
- What's the difference?
 - No syntactic restrictions
 - Not meaning (?)



Not just the kids these days?

- However... 'Dutch' order also found in older Frisian texts

Oudfriesse Oorkonden (Middle Frisian charters, 14th and 15th century): 40% 1-2 (Larooij, 1991)

Elder Skeltenariucht (legal text, ~1300): 10% 1-2 (Bor, 1971)

Essay by *Gysbert Japicx* (17th century): 10% 1-2 (Hoekstra, 2010)

- Existence of 'mixed' texts of Middle Low German and Middle Frisian

Contact hypotheses

■ Widespread bilingualism

- Similar to present situation in province of Frisia
- Change from below (Labov 1965)
- Characterized by early acquisition of prestige language (Weerman 2011)

■ Learned contact

- Comparable to e.g. Latin influence on Old High German
- Change from above
- Characterized by late acquisition of prestige language

Early-modern Frisian

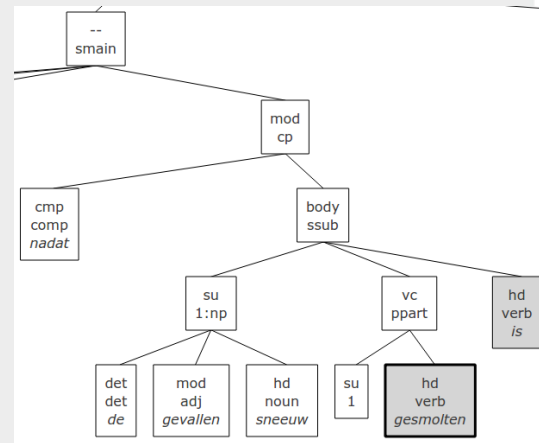
- Study of Early-Modern Frisian texts (1550-1800)
 - Integrated Language Database (Fryske Akademy)
 - Part-of-speech tagged corpus of historical Frisian

- Were Dutch-style 1-2 orders used? In what contexts?
 - Dutch-like contexts indicate bilingualism
 - Non-Dutch-like contexts indicate learned borrowing

```
<w l="dizze" s="pronomen" t="nulfoarm">dizze</w>
<w l="grouweldied" s="de-subst." t="nulfoarm|=it">
gruwle-diede</w> <w l="net" s="adverbium" t=
"nulfoarm">naet</w> <w p="oon~schoagjen">oon</w>
<w l="te" s="preposysje" t="nulfoarm">to</w> <w l=
"oanskôgje" p="oon~schoagjen" s="haadtw." t=
"[monotr.+NP]df">schoagjen</w>, <w l="selm2" s=
"adverbium" t="nulfoarm">salm</w> <w l="de" s=
"artikel" t="nulfoarm">de</w> <w l="Melezer" s=
```


Features of Dutch 1-2 orders (Coussé et al, 2008, Bloem et al. 2017)

- Semantic properties and associations
 - Type of auxiliary verb (modal, passive etc.)
 - 'to'-infinitives
- Discourse factors and priming
- Processing difficulty of the cluster and context
 - Length of clause
 - Use of separable verbs
 - Function words near clusters
 - Frequency of main verb
- Stylistic factors
 - Prose or rhyming text



Processing complexity: Morphological structure of the participle

- Some Germanic complex verbs are separable

1. dat ik het **heb** **volgehouden**

that I it **have** endured

2. dat ik het **vol** **heb** gehouden (*volbracht)

that I it **PT** have endured

- Separable verbs more difficult to process

- Separable verbs more in **1-2** order in Dutch (Bloem et al. 2017)

Verb morphology	Odds ratio
Non-separable	1.00
Separable	7.06

Task

- Find Early-Modern Frisian **1-2** clusters
 - Not trivial: Annotation limitations
- Model using features transferred from Dutch
 - Are features of **1-2** orders in modern Dutch also informative for Frisian?
 - Not every feature can be operationalized for Frisian
 - Rhyme and prose difference (stylistic factor)
 - Logistic regression model for identifying contribution of each feature

Why modern Dutch? Not enough data on Middle Dutch or Modern Frisian

Data sources

■ Fryske Akademy Integrated Language Database

- Reyner Bogerman's Frisian Sayings (1542) - Rhyming
- Burmania-spreekwoorden (1614) - Prose
- Wedding poetry J. van Hichtum (1639) - Rhyming
- Collected prose texts of *Gysbert Japicx* (1639-1666) - Prose
- Collected rhyming texts of *Gysbert Japicx* (1639-1666) – Rhyming
- Waatze Gribberts Bruyloft (1701) – Prose
- Prose of Eelke Meinderts (1777-1783) - Prose
- Poetry of Eelke Meinderts (1777-1783) - Rhyming

■ Manually tokenized, lemmatized, annotated with part-of-speech tags

■ Automatic extraction of candidate verb clusters

■ 1546 automatically identified two-verb clusters

1169 prose, 377 rhyming

understood have | have understood

Verb cluster order in Early-Modern Frisian

12

GYSBERT JAPIX
FRIESCHE
RYMLERE;

In trye dielen forfchaet:

d'Eerste binne

LJEAFD in BORTLYCKE
MINGEL-DEUNTJES:

's Oorde fette

GEMIENE af HUZMANNE
PETEAR, in oere KATERYE:

's Efteske iz:

HYMMELSCH HARPLUWD.

Dat iz to fassen, yfelycke fen

DAVIDS PSALMEN.

Dy oorde druwck, mei ien twaede Diel forgreat.



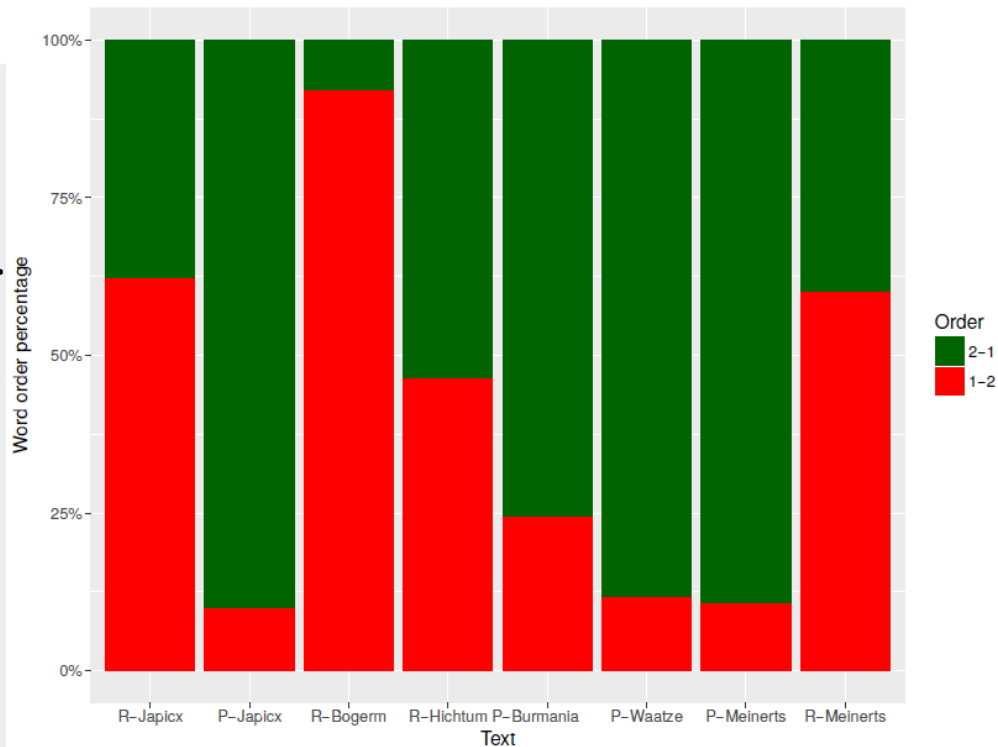
TO LEUWARD.

By KERST TJALLINGS in GERIT
HUGSLAG, Boekfckpers. 1681.

Results by text

More 1-2 orders in rhyming text ->
Stylistic option

In prose texts, some of the
1-2 clusters rhyme
(but not all)



Results (multifactorial logistic regression, R glm)

18.69 means: 1-2 order
18.69 times more likely

*** means: $p < 0.001$

Feature	Contrast	Odds
Text type	+Rhyme / -Prose	*** 18.69
Text year	Per additional year	0.998
Type of auxiliary verb	+Modal / -Cop Asp Passive Past Future +Copula / - Asp Passive Past Future +Aspectual / -Passive Past Future +Passive / -Past Future +Past / -Future	1.19 3.03 *** 9.67 1.11 *** 5.04
To-infinitive	+Yes / -No	*** 8.33
Priming	+1-2 / -2-1	0.95
Separable verb	+Separable / -Inseparable	0.64
Information value	+High / -Low Medium +Medium / -Low	0.91 *** 0.24
Frequency of main verb	Per additional order of magnitude	0.959

Results for Dutch (Bloem et al. 2017)

Verb clusters extracted from Wikipedia data

Issues in model comparison

- Different features
- Different annotation
- Different corpus size

Comparison of conclusions, not models

Feature	Contrast	Odds
Text type		
Text year		
Type of auxiliary verb	Modal Aspectual <i>hebben (to have)</i> <i>worden (to be)</i> <i>zijn (to be)</i>	*** 159.27 ***1135.14 ***2.00 *** 1.47 (baseline)
To-infinitive	Yes	*** 0.29
Priming	1-2 -2-1	*** 1.24 *** 0.59
Separable verb	Separable	*** 7.06
Information value	Low High	(baseline) ***0.79
Frequency of main verb	Per additional order of magnitude	*** 1.06

Discussion

- Constructional factors:

To-infinitive

+Yes / -No

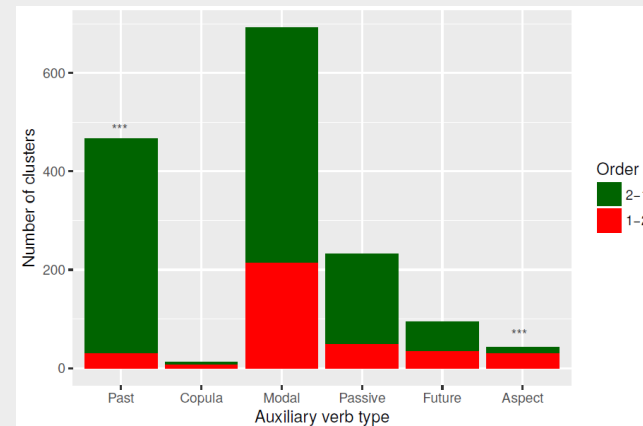
*** 8.33

- High proportion of 'te'-infinitive constructions in 1-2

- Hoekstra (2010) took these as a sign of contact

- Auxiliary verb types not similar to Dutch:

In Dutch, modals predict 1-2 order most strongly



Discussion

- Two-verb clusters in the 1-2 order were used in Early-Modern Frisian texts
- Usage patterns similar to Dutch:
 - `to'-infinitives in 1-2 order
- Dissimilar usage patterns:
 - No evidence of a processing effect (priming, separable verbs)
 - Different preferences for auxiliary verbs
- Strong 1-2 order preference in rhyming text, formal text

Conclusion

- The 1-2 order appears to be a stylistically marked option rather than an integral part of EMF grammar
- Points towards 'learned borrowing' and late acquisition
- Change through early acquisition happened later
- Restrict hypothesis space based on findings from related language to model probabilistic phenomena even in a low-resource language
- Some types of language change can only be studied in this way

References

- J. Bloem, A.P. Versloot & F.P. Weerman (2017). Verbal cluster order and processing complexity. In: Enoch Aboh, editor, *Complexity in human languages: A multifaceted approach*, volume 60, pages 94–119. Elsevier.
- J. Bloem, A.P. Versloot & F.P. Weerman (2014). [Applying automatically parsed corpora to the study of language variation](#). In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers* (pp. 1974-1984). Dublin, Ireland: Dublin City University and Association for Computational Linguistics.
- Bor , Arie (1971). *Word-groups in the language of the Skeltana Riucht: a syntactic analysis with occasional lexicological observations; followed by an inquiry into its punctuation and the possibility of the influence on the text of spoken language*. H. Veenman & Zonen
- Buma, W. J. (1957). *Aldfryske houlikstaspraken*. van Gorcum, 1957.
- Coussé, E., Arfs, M., & De Sutter, G. (2008). Variabele werkwoordsvolgorde in de Nederlandse werkwoordelijke eindgroep. Een taalgebruiksgebaseerd perspectief op de synchronie en diachronie van de zgn. rode en groene woordvolgorde. In G. Rawoens (Ed.), *Taal aan den lijve. Het gebruik van corpora in taalkundig onderzoek en taalonderwijs* (pp. 29–47). Gent: Academia Press.
- De Haan, G. J. (1996). Recent changes in the verbal complex of Frisian. In: H. F. Nielsen, & A. Petersen (Eds.), *Frisian and Germanic Miscellany*. Published in Honour of Nils Århammar on his Sixty-Fifth Birthday, 7 August 1996. Odense: Odense University Press.

References

- Hoekstra, E. (2010). Reade wurdfolchoarders en dêrmei gearhingjende aspekten yn 17e-ieusk Frysk. In ûndersyk nei de tiidwurdkloft yn Gysbert Japicx syn "Yen suwnerlinge forhânlinge Fen it Libben In fenne Deade". *It Beaken. Tydskrift fan de Fryske Akademy*, 72.
- Labov, W. (1965). On the mechanism of linguistic change. Georgetown University Monographs on Language and Linguistics, 18:91–114.
- Larooij, Pieter (1991). Undersyk nei de posysje fan it helptiidwurd in de bysin yn *Oudfriesse Oorkonden, Bewerkt door O. Vries*, Vierde Deel. Master's thesis, University of Groningen
- Weerman, F. (2011) Diachronic change: Early versus late acquisition. *Bilingualism: Language and Cognition*, 14(2):149151.
- Weinreich, U. (1953). *Languages in contact, findings and problems*. New York: Linguistic Circle of New York.