University of Amsterdam

An agent-based model of a historical word order change

Jelke Bloem, Arjen Versloot, Fred Weerman

Agent-based modeling of language

- Multiple language models that communicate
- Models a community of speakers (agents)
- Used in evolutionary linguistics
- Applications in historical linguistics?
  - Same mechanisms, different goals
  - Informed by historical data

Van Trijp (2012), Landsbergen et al. (2010), Pijpops and Beuls (2015)

Starting point

```
<w l="hjirmei" t="nulfoarm">Hier mey</w> <w
l="wolle4" t="[modaal]1-mt-nt">wolle</w> <w
l="wy" t="nulfoarm">wy</w> <w l="dy"
t="nulfoarm">dy</w> / <w l="ús"
t="nulfoarm">ws</w> <w l="Leavenheare"
t="nulfoarm|=it">liæwin Heer</w> <w
l="befelle" t="[bytrans.]nulfoarm|
=nf">byfelle</w>.
```

- Minimal assumptions about language faculty
- Start with data from historical corpora
- Model factors that may be relevant to the change
- Account for current state of language(s)

- Generation of hypotheses on language change

Case study: Verbal cluster word order

- Free order variation in Dutch
  1. ik denk dat ik het begrepen$_2$ heb$_1$

     I think that I it understood$_2$ have$_1$
  2. ik denk dat ik het heb$_1$ begrepen$_2$

     I think that I it have$_1$ understood$_2$
- German, Frisian: Only 2-1 order
- English, Scandinavian: Only 1-2 order *

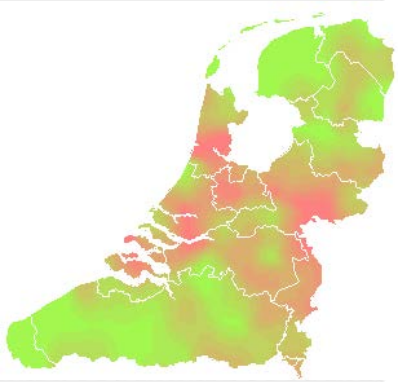
Why did they diverge?

Language variation and change

How do we find factors involved in change?

- Language variation often caused by change

- Start by looking at the language with variation
  1. ik denk dat ik het begrepen$_2$ heb$_1$
  2. ik denk dat ik het heb$_1$ begrepen$_2$
- A language change in progress?



Correlates of variation: Meaning and function

- **Type of clause** — main clause / subordinate clause
- **Type of auxiliary** — ‘have’ / copular / modal

- Separable main verb — … heeft **af**gewassen (has washed **up**)
- Constituent after cluster — … heeft gezien **dat het gebeurde**
- Length of the middle field — … dat [hij naar hun auto] is gelopen
- Syntactic persistence — …afgewassen heeft en …weggelopen is
- Main verb frequency — … naar hun auto is **gelopen**
- Inherence (multi-word units) — … dat hij [rekening zou houden met] …
- …

The model

- An example sentence looks like this:

	Modal	'*to have*'	Copular
main clause (**MC**)	X		
subordinate clause (**SUB**)			

- Create *a* agents each with *n* exemplar sentences
- Each agent has its own language model
  - Features have word order preference based on known exemplars
- Starting situation based on West-Germanic ~500AD

The simulation

- Series of interactions between two random agents
- Speaker agent generates verb cluster based on its language model
  - Features are taken from random stored exemplar
  - Word order is assigned based on features:

    p(asc|sub-mod) = p(asc|sub) + p(asc|mod)
- Recipient agent stores it as exemplar (incl. order)
- Speaker deletes the exemplar from its language

Word order change in the model

p(asc|sub-mod) = p(asc|sub) + p(asc|mod)

- Simulates the fact that people do not perfectly copy a language from each other
- Functional bias -> change
  - i.e. deep structure bias, or efficiency
- Learning bias changes probability distributions in the agents and causes language change

Historical changes relating to model factors

- Constructions with *to have* growing from a very low level:

	Old	Modern
English: *have*	2%	31%
German: *haben*	1%	36%

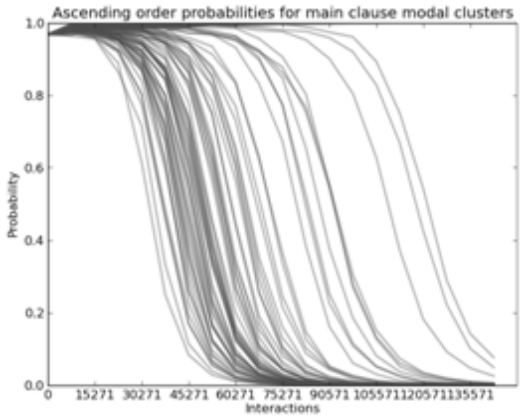
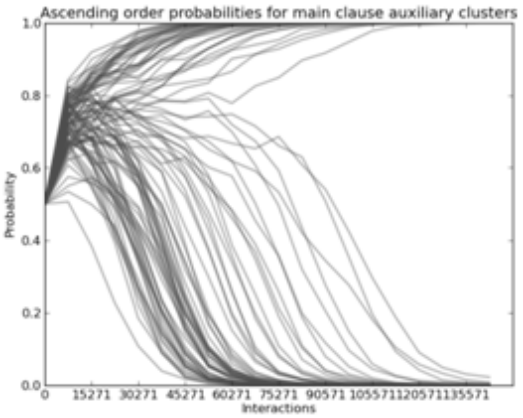
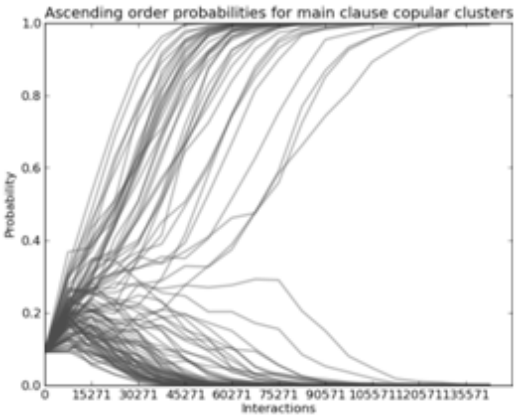
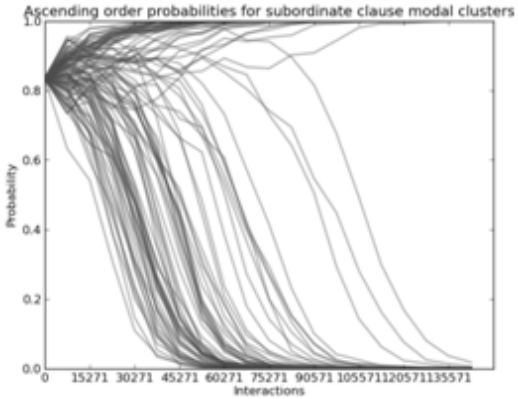
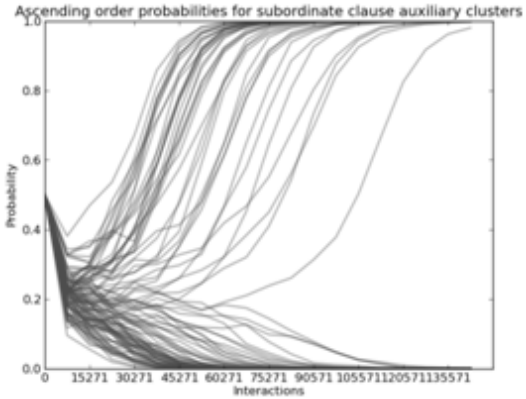
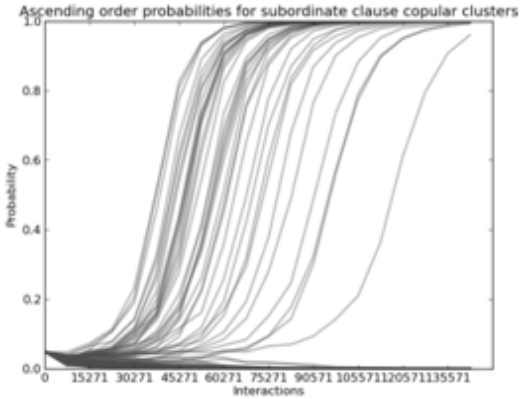
- Emerged later than the first clusters, the modal+inf combination
- Implemented as growth phase in the model

- Increasing number of subordinate clauses

understood have | have understood

Outcome for 30 agents, 5000 interactions

Equal increase of *to have*-constructions and subordinate clauses

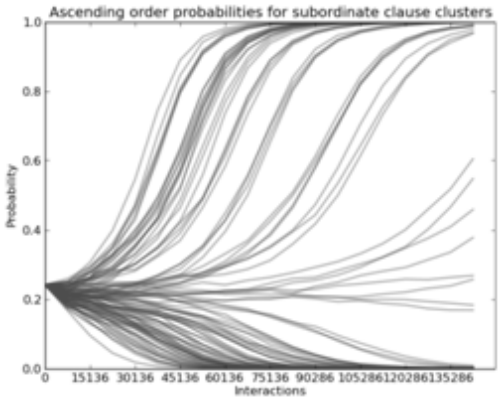
	Modal	‘to have’	Copular
MC			
SUB			

The model correctly predicts both dominant 1-2 (English) and 2-1 (German)

understood have | have understood

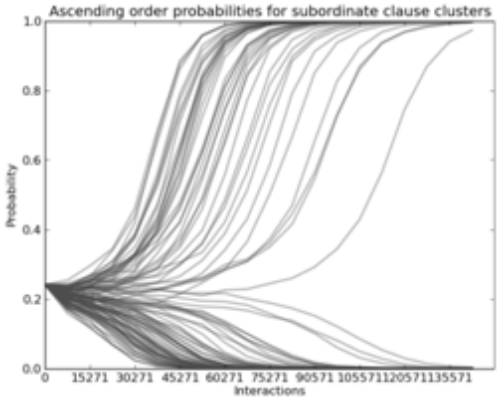
Influence of the relative growth velocity of *to have*-constructions

quick growth (‘English’)



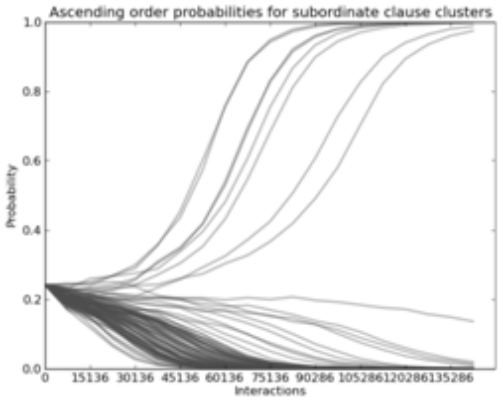
56%/35%

moderate growth



63%/36%

slow growth (‘German’)



92%/7%

Quicker ‘have’ growth increases the chances of a **1-2** word order

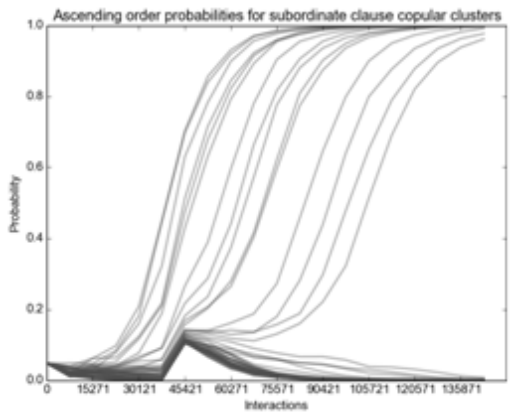
Results

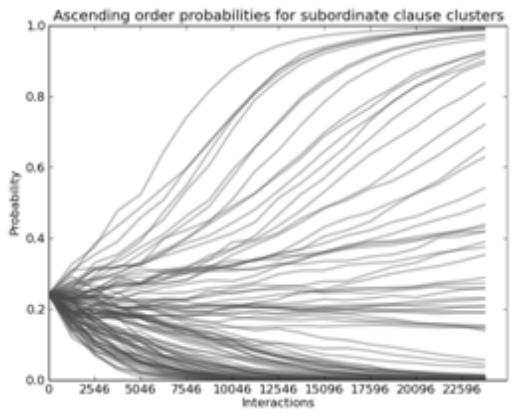
- Growth of ‘have’ supports **1-2** order
  - Prediction: more ‘have’ in English
- Growth of subclauses supports **2-1** order
  - Prediction: more sub clauses in German

-> The dominant word order may depend on different preference for specific constructions

Dutch: Another process of change?

- Dutch followed pattern of German until ~1500
- Now changing to 100% **1-2** order?
  - **1-2** order is acquired first (Meyer & Weerman, 2014)
  - **1-2** order is catching on quickly in Frisian
- Some other factor triggered a second change





Discussion

- Agent-based model as a tool for historical linguistics
- Two hypotheses about historical word order change:
  - "Have" clusters grammaticalizing faster

    supported the **1-2** order (i.e. English)
  - Increased use of subordinate clauses

    supported the **2-1** order (i.e. German)
- Predictions can be tested using historical corpora

Discussion

- Can test what change is possible in a language, given the assumptions and starting conditions
- Our model has few assumptions:
  - Single learning bias
  - Does not depend on framework
- Only needs features, frequencies and a change over time caused by these features
- Applicable to other cases of language change

References

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.

Frank Landsbergen, Robert Lachlan, Carel ten Cate, and Arie Verhagen. A cultural evolutionary model of patterns in semantic change. *Linguistics*, 48(2):363–390, 2010.

Caitlin Meyer and Fred Weerman. Cracking the cluster: The acquisition of verb raising in Dutch. *Manuscript in preparation*, submitted

Dirk Pijpops and Katrien Beuls. Strong ”island of resilience” in the weak flood. Dutch strategies for past tense formation implemented in an agent-based model. Presented at Computational Linguistics in the Netherlands (CLIN) 25, Antwerp, 2015.

Remi Van Trijp. Self-assessing agents for explaining language change: A case study in German. In ECAI, pages 798–803, 2012.