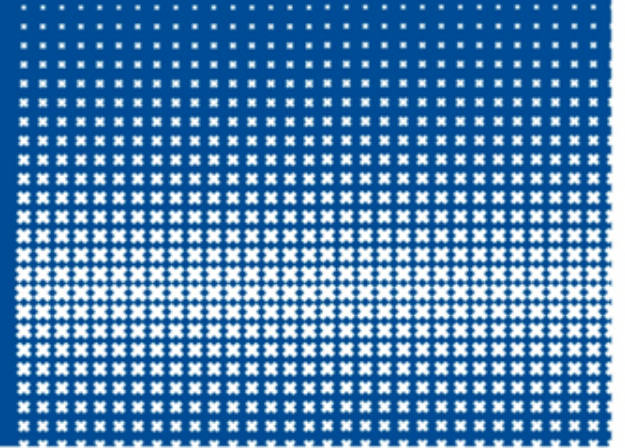




Jelke Bloem, Arjen Versloot, Fred Weerman



Modelling Germanic Syntax

Word order changes and grammaticalization in verbal clusters

Verbal clusters

■ Free order variation in Dutch

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

■ Frisian, German: Only descending order

■ English, Scandinavian: Only ascending order *

Verbal clusters

- Free order variation in Dutch
- Frisian, German: Only **descending order**
- English, Scandinavian: Only **ascending order** *
 1. I think that I **have**₁ **understood**₂ it

Why did they diverge?

Word order changes

- Can we model or simulate historical changes in verbal cluster word order?
- Start with proto-West-Germanic, end at the current state of the West-Germanic languages
- Language agents that produce and perceive verbal clusters

Agent-based modeling of language

- Multiple language models that communicate
- Models a community of speakers
- Simulates the fact that people do not perfectly copy a language from each other
- (innate) Learning bias -> change
 - i.e. deep structure bias
- Learning bias changes probability distributions

Language variation and change

■ Free order variation in Dutch

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

■ Language variation often caused by change

■ A language change in progress?

Dutch order variation

- **Type of clause**
main clause / subordinate clause
- **Type of auxiliary**
copular-*zijn*/passive-*zijn*/time/*worden*
- **Separable main verb**
... heeft afgewassen (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
- **Pre-verbal constituent: Informativity and inheritance**

Model of a verbal cluster

■ Type of cluster

- mod+inf ik denk dat ik het **zien wil**
- have+PP ik denk dat ik het **gezien heb**
- cop+PP ik denk dat hij **gezien is**

■ Clause type: main clause or subordinate clause

1. This can not **be denied**.

■ 2 outcomes: **ascending** or **descending** order

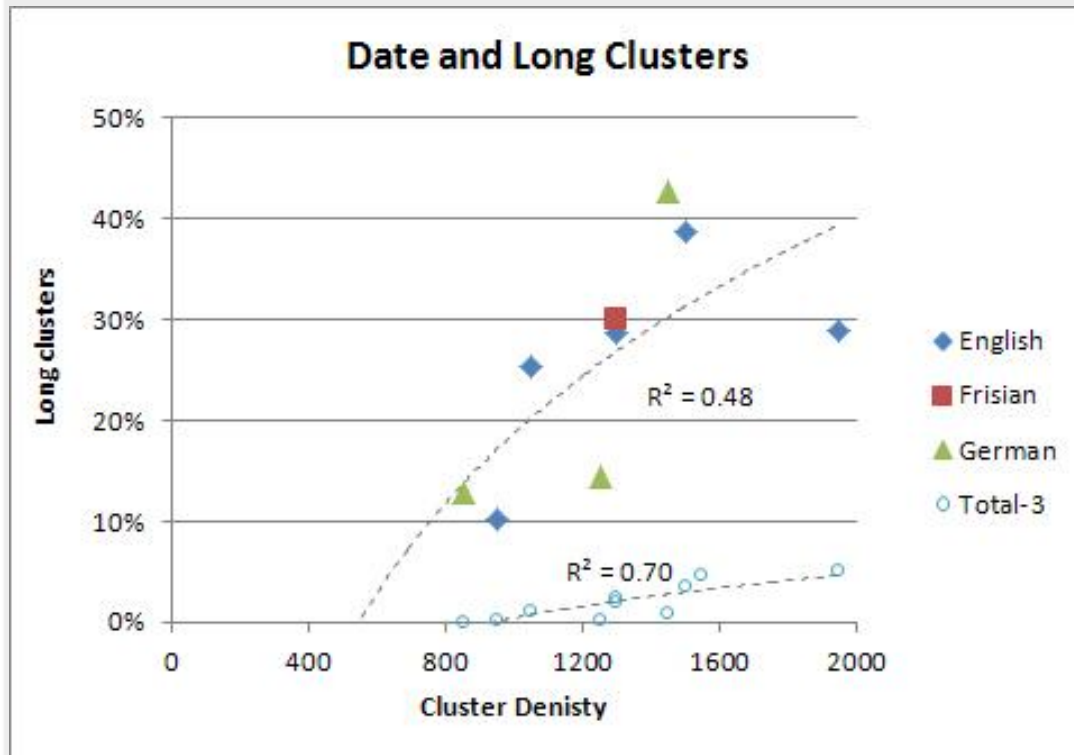
Modeling of agents (speakers)

- An example sentence looks like this:

	Modal + inf.	<i>to have</i> + part.	copula + part.
main clause	X		
subordinate clause			

- Initialize n agents with exemplar sentences
- Random agents transfer exemplars:
 - $p(\text{asc}|\text{mod-main}) = p(\text{asc}|\text{mod}) * p(\text{asc}|\text{main})$
 - Learning bias

Growth of multi-verb constructions in Germanic Languages



- The growth of 2-verb clusters in Germanic languages since ca. 500
- The growth of 3-verb clusters in Germanic languages since ca. 800.

Historical patterns underlying the model's starting position

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

	Old	Modern
English: <i>have</i>	2%	31%
German: <i>haben</i>	1%	36%

- Emerged later than the first clusters, the modal+inf combination
- Growth phase in the model
- Increasing number of subordinate clauses

Historical patterns underlying the model's starting position

- Constructions with *participles* biased towards subordinate clauses:

% participles	main clauses	subordinate clauses
Old High German	70%	95%
Old Frisian	15%	37%

Historical patterns underlying the model's starting position

- *Modals + infinitives* have a preference for **ascending** word order:

ascending	copula + part.	modal + infinitive
Old High German	58%	83%
Old Frisian	15%	66%

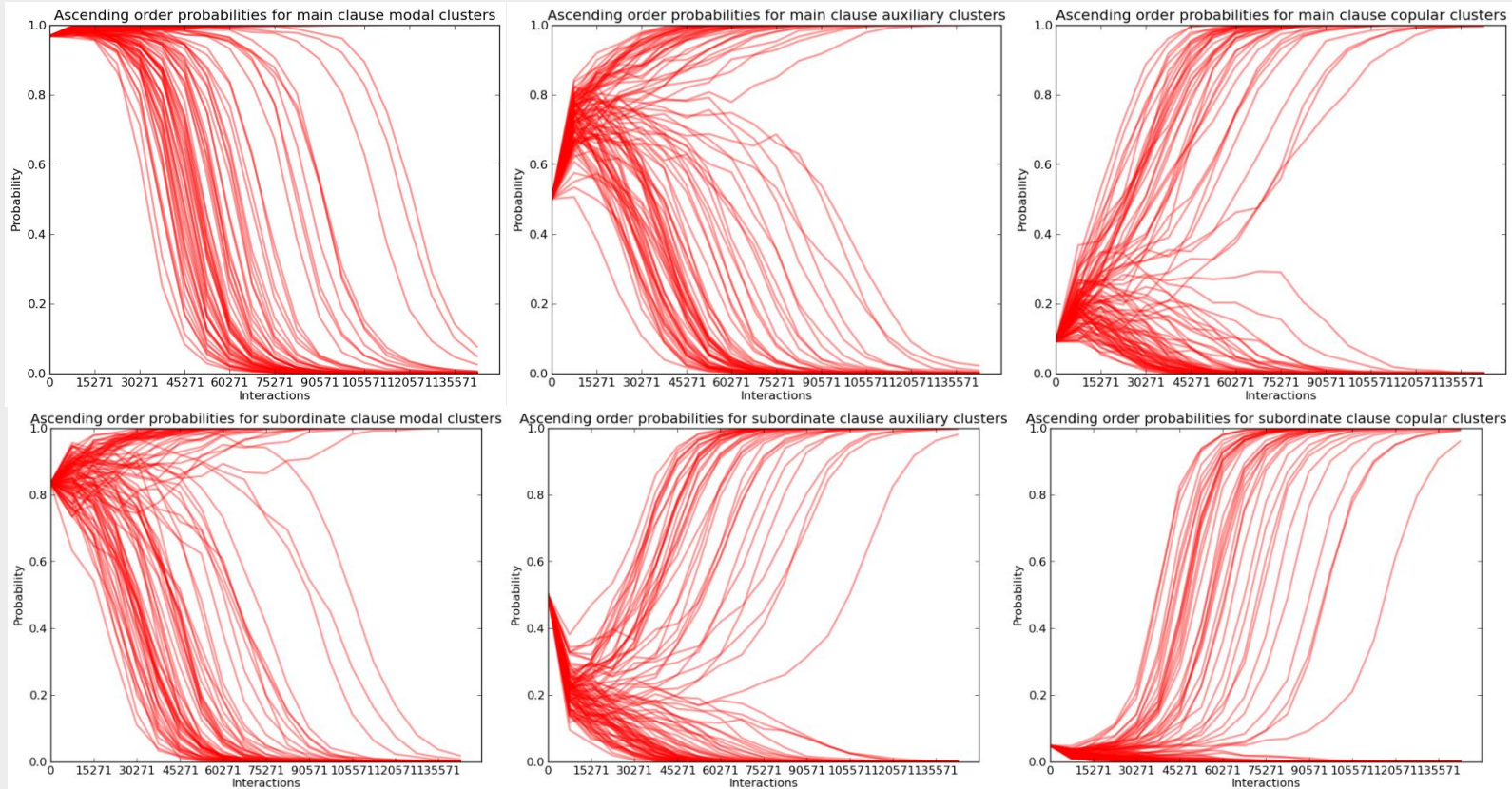
Starting position for the algorithm

An “idealized” situations starts with:

- **ascending** *modal + infinitive* constructions, predominantly appearing in main clauses
- **descending** participium + copula constructions, predominantly appearing in subordinate clauses

	Modal + inf.	<i>to have</i> + part.	copula + part.
main clause	30	1	10
subordinate clause	5	1	20

Outcome for 30 agents, 5000 interactions even increase of *to have*-constructions and subordinate clauses



The model correctly predicts both dominant
ascending (English) and **descending** (German)

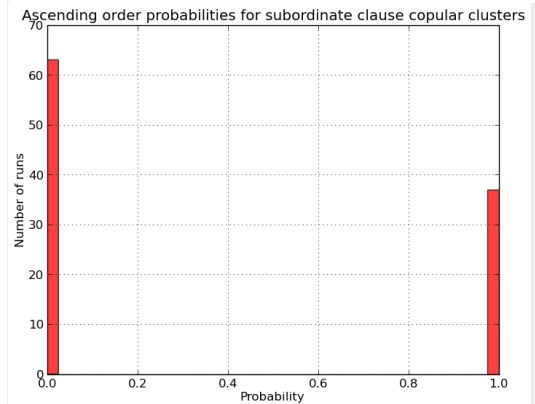
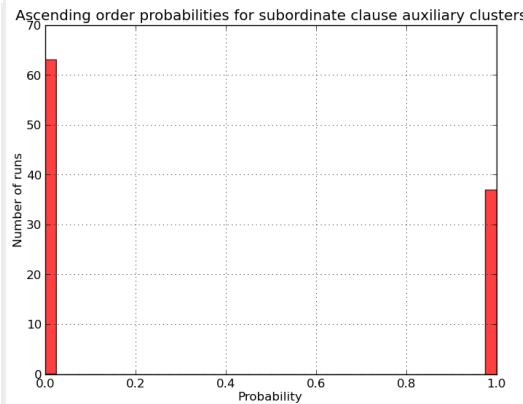
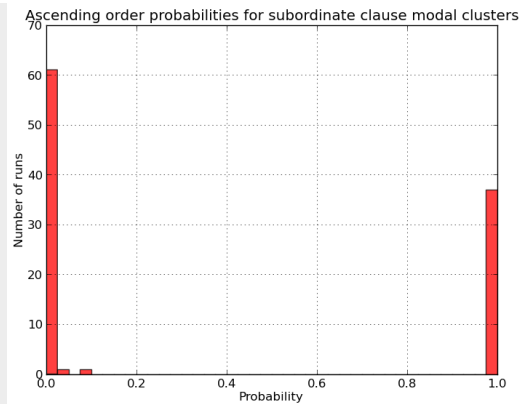
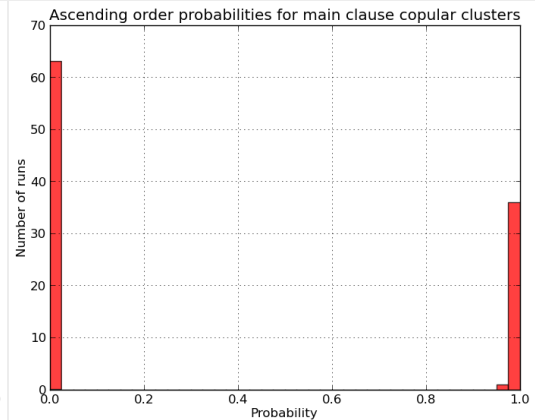
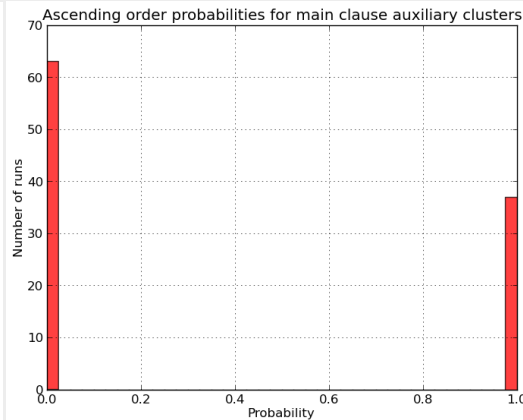
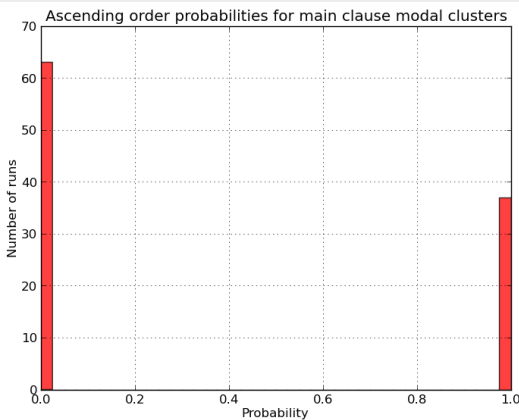
begrepen heb

heb begrepen

Modelling Germanic syntax

15

Outcome for 30 agents, 5000 interactions even increase of *to have*-constructions and subordinate clauses



The model correctly predicts both dominant
ascending (English) and **descending** (German)

begrepen heb

heb begrepen

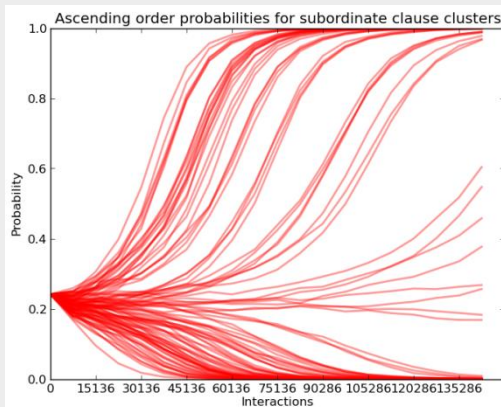
Modelling Germanic syntax

Influence of the relative growth velocity of *to have*-constructions (typical for English)

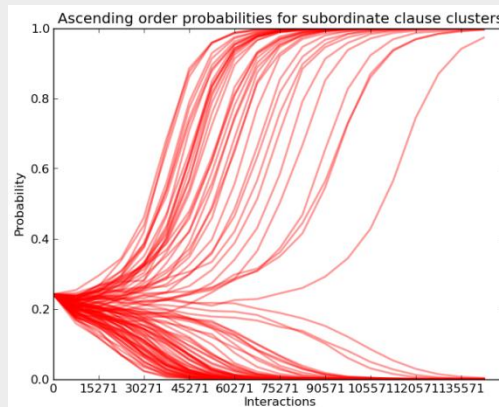
quick growth ('English')

moderate growth

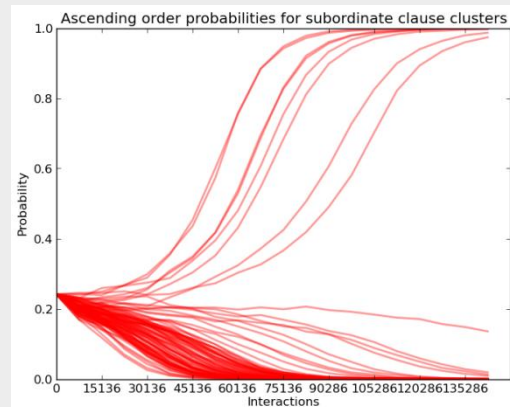
slow growth ('German')



56%/35%



63%/36%



92%/7%

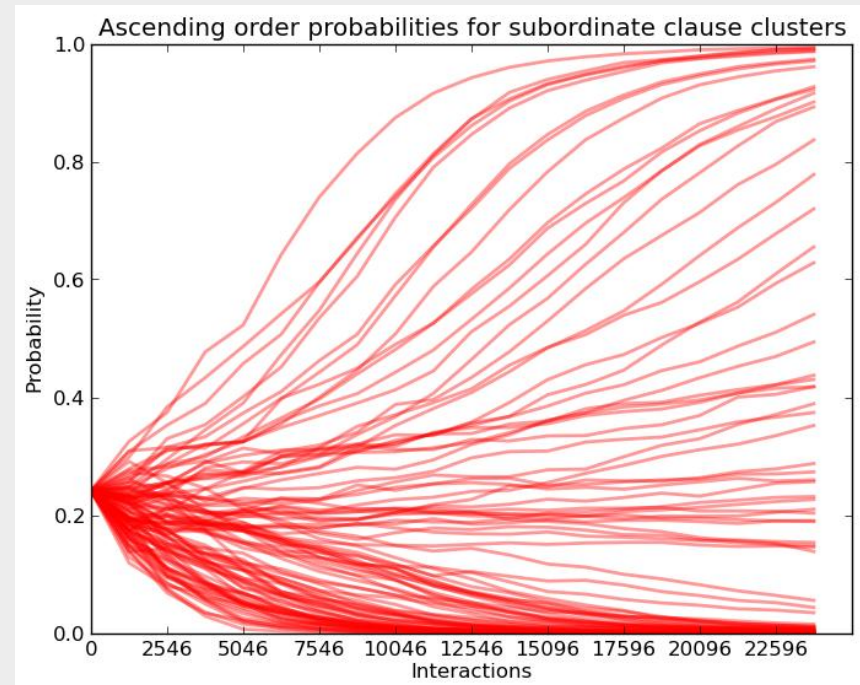
Quicker 'have' growth increases the chances of an **ascending** word order

Results

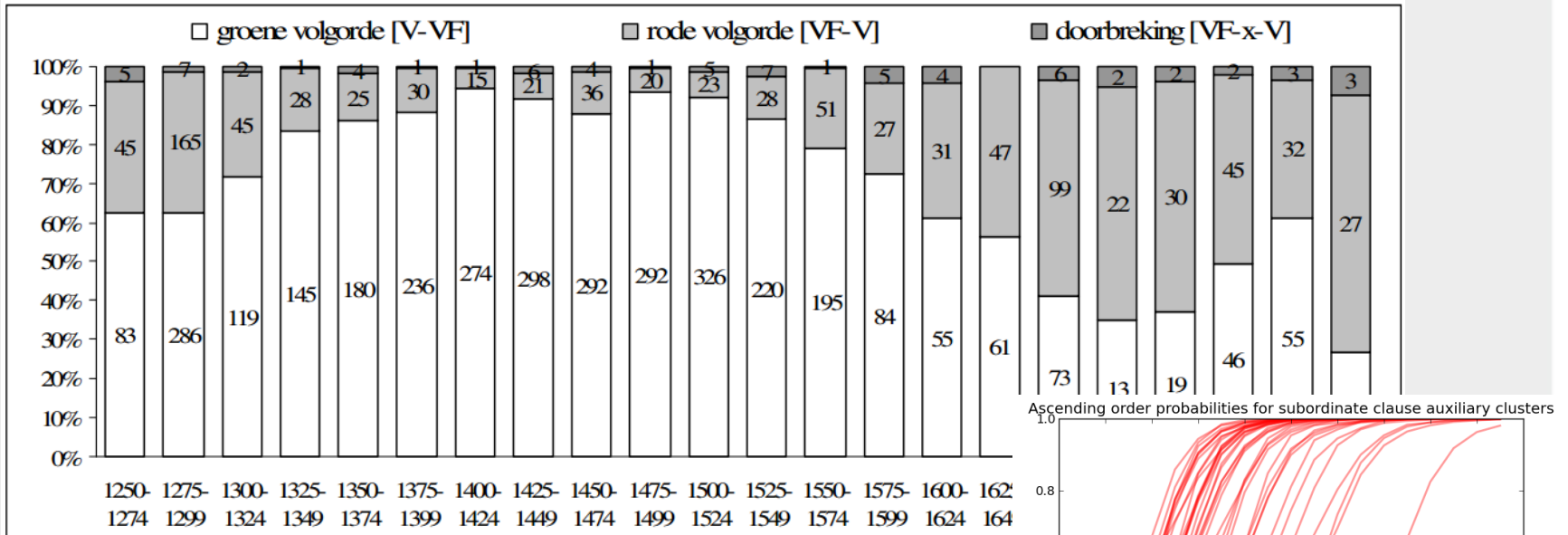
- Growth of 'have' supports **ascending** order
 - Prediction: more 'have' in English
 - Growth of subclauses supports **descending** order
 - Prediction: more sub clauses in German
- > The dominant word order may depend on different preference for specific constructions

Dutch variation: a change in progress?

- Model may remain in unstable state for a while
 - Optionality
- Dutch is changing to 100% ascending order?

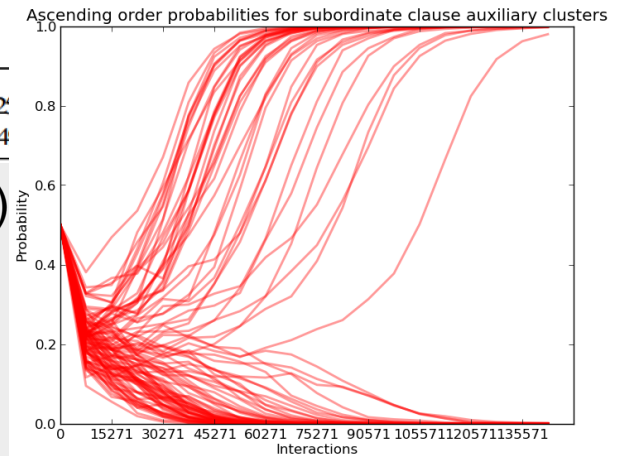


Dutch historical change



Verb order in official texts (n = 4327) (Coussé 2008)

Or: 2 different processes of change



Diachronic change in cluster order

Model predictions:

%green	mod+inf	habba+PP	cop+pp
main	100%	92%	70%
sub	98%	33%	9%

Probabilities from early Modern Frisian text (c. 1550):

%green	mod+inf	habba+PP	cop+pp
main	100%	100%	100%
sub	100%	33%	20%

But: the 100% **ascending** main clause is the V2-effect, which our model does not yet account for

Discussion

- Auxiliary type and clause type may be used as a diachronic explanation
- Grammaticalization or embedding?
 - This can not **be denied**. (main clause)
 - ... that it not **denied** can **be**. (Contrasting)
- Increased use of subordinate clauses may have changed base order to **descending**
- “Have” clusters support the opposite **ascending** order (English examples)
- Unstable phenomenon can be modelled well

Outcome probabilities (over ideal distributions)

■ Starting values:

mod-mc	30	cop-mc	10	heb-mc	1
mod-sub	5	cop-sub	20	heb-sub	1

$$p(\text{red}|\text{mod}) = 30 + 5 / (30+5+1+1) = 0.95$$

$$p(\text{red}|\text{mc}) = 30 + 1 / (30+1+10+1) = 0.74$$

$$p(\text{red}|\text{mod-mc}) = p(\text{red}|\text{mod}) * p(\text{red}|\text{mc}) = 0.7 / 70\%$$

$$P(\text{green}|\text{mod-mc}) = 1\%$$