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

- 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?

<table>
<thead>
<tr>
<th>understood have</th>
<th>have understood</th>
</tr>
</thead>
</table>
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 \textit{begrepen}_2 heb_1
  2. ik denk dat ik het heb_1 \textit{begrepen}_2

- A language change in progress?

understood have \hspace{1cm} have understood

An agent-based model of a historical word order change
Correlates of variation: Meaning and function

- **Type of clause**
  main clause / subordinate clause

- **Type of auxiliary**
  ‘have’ / copular / modal

- 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

- Inherence (multi-word units)
  … dat hij [rekening zou houden met] …

understood have | have understood
The model

- An example sentence looks like this:

<table>
<thead>
<tr>
<th></th>
<th>Modal</th>
<th>‘to have’</th>
<th>Copular</th>
</tr>
</thead>
<tbody>
<tr>
<td>main clause (MC)</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>subordinate clause (SUB)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- 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 \(~500\text{AD}\)
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:
    \[ p(\text{asc}|\text{sub-mod}) = p(\text{asc}|\text{sub}) + p(\text{asc}|\text{mod}) \]
- Recipient agent stores it as exemplar (incl. order)
- Speaker deletes the exemplar from its language

understood have | have understood

An agent-based model of a historical word order change
Word order change in the model

\[ p(\text{asc}|\text{sub-mod}) = p(\text{asc}|\text{sub}) + p(\text{asc}|\text{mod}) \]

- 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:

<table>
<thead>
<tr>
<th></th>
<th>Old</th>
<th>Modern</th>
</tr>
</thead>
<tbody>
<tr>
<td>English: <em>have</em></td>
<td>2%</td>
<td>31%</td>
</tr>
<tr>
<td>German: <em>haben</em></td>
<td>1%</td>
<td>36%</td>
</tr>
</tbody>
</table>

- Emerged later than the first clusters, the modal+inf combination
- Implemented as growth phase in the model
- Increasing number of subordinate clauses
Outcome for 30 agents, 5000 interactions
Equal increase of *to have*-constructions and subordinate clauses

The model correctly predicts both dominant 1-2 (English) and 2-1 (German)
Influence of the relative growth velocity of to have-constructions

Quick growth (‘English’)  moderate growth  slow growth (‘German’)

<table>
<thead>
<tr>
<th>Quick growth ('English')</th>
<th>Moderate growth</th>
<th>Slow growth ('German')</th>
</tr>
</thead>
<tbody>
<tr>
<td>56%/35%</td>
<td>63%/36%</td>
<td>92%/7%</td>
</tr>
</tbody>
</table>

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

understood have  |  have understood  | An agent-based model of a historical word order change
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

understood have | have understood

An agent-based model of a historical word order 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


