Language-specific differences in regularization rates of the Germanic preterite

Isabeau De Smet\textsuperscript{1}
Katrien Beuls\textsuperscript{2}
Dirk Pijpops\textsuperscript{1, 3}
Freek Van de Velde\textsuperscript{1}

\textsuperscript{1} University of Leuven, QLVL
\textsuperscript{2} AI-Lab, Vrije Universiteit Brussel
\textsuperscript{3} FWO Flanders
Strong and weak preterites

- Germanic languages have two morphological strategies for building preterites (not counting analytic perfects, *he has written a book*):

1. **Strong inflection**:
   - English *sing* – *sang*
   - Ablaut, based on Indo-European aspectual system (perfect > preterite)

2. **Weak inflection**
   - English *work* – *worked*
   - Dental suffix, based on analytic formation `[VERB + *d<sub>1</sub>eh-1, *d<sub>1</sub>oh-1 ('did')]`
Changes

• Various changes occur:
  – irregularisation (Eng. buy – bought)
  – one strong ablaut class to another (Du. heffen – hief < hoef (Germ. hob, hub))
  – weak to strong (Du. vragen – vroeg < vraagde (vs. Germ. fragte))
  – strong to weak (Eng. carve – carved < cearf (Du. kerfde < karf))

⇒ Long-term drift, over many centuries
Quantifying the weakification

- Lieberman et al. (2007):
  - tracked all originally strong Old English verbs (that still exist)
  - noted when they weakened (Middle or Modern English)
  - reference grammars
  - binary encoding (strong = 1, weak = 0)
  - 6 log-frequency bins
Quantifying the weakification

- Lieberman et al. (2007):
  - tracked all originally strong Old English verbs (that still exist)
  - noted when they weakened (Middle or Modern English)
  - reference grammars
  - binary encoding (strong = 1, weak = 0)
  - 6 log-frequency bins
- Carroll et al. (2012):
  - German
  - same method
  - Old, Middle, Early New, New High German
Quantifying the weakification

- Dutch data (2017)
  - Old, Middle, Modern (1500-1800) and present-day Dutch (1800-now)
  - controlled for type-token frequency and vowel pattern (ABA, ABB or ABC)
ENGLISH: Lieberman et al. 2007

GERMAN: Carroll et al. 2012
Lieberman et al. 2007: Constant rate of regularisation through time, only dependent on frequency

$$\Rightarrow$$ lines follow the same power law curve (linear on log-log plot) and overlap
Lieberman et al. 2007: Constant rate of regularisation through time, only dependent on frequency

Carroll et al. 2012: Constant rate does not work for German

If the constant regularisation rate were true, the lines would follow the same power law curve and overlap

... neither for Dutch
Lieberman et al. 2007: Constant rate of regularisation through time, only dependent on frequency

⇒ lines follow the same power law curve (linear on log-log plot) and overlap

Lieberman et al. 2007: three measurement points:

Replication with fourth measurement point:
Lieberman et al. 2007: Constant rate of regularisation through time, only dependent on frequency

But the constant rate breaks down when we add an extra measurement point for E. Mod. Eng.:
Socio-demographical factors

• Can we attribute these changes to demography?

"[A] social characteristic with structural consequences is dialect or language contact. Increased exposure to different varieties often – though not always – corresponds to patterns of morphological and other leveling or simplification (...). The ENHG period, when verb regularization picks up dramatically in the history of German, is a period notable for increased geographical mobility, in particular urbanization."

(Carroll et al. 2012: 169)
Historical demographic data

• Problem: no clear data on population size or migration
Historical demographic data

• Problem: no clear data on population size or migration

• We can work with urbanisation:
  – In pre-industrial times, population growth is too high to be explained solely by natural growth (De Vries 1984:199-266, Howell 2006:208)
  – Migration, leading to koineization (Kerswill 2002), due to an influx of L2 speakers
    • Language diversity was higher in Medieval and Early Modern cities
    • Dialects were often mutually unintelligible

• Data:
Historical demographic data

• 3 Areas:
  – ENGLISH: United Kingdom
  – DUTCH: Belgium & Netherlands
  – GERMAN: Germany & Austria

• 4 Periods:
  – Old: mean number of inh. of 3 largest cities in 1200, weighted for their rank
  – Middle: mean number of inh. of 5 largest cities in 1500, weighted for their rank
  – New: mean number of inh. of 7 largest cities in 1700, weighted for their rank
  – Present: mean number of inh. of 9 largest cities in 1900, weighted for their rank

• Cumulative percentage of weak verbs over total number of originally strong verbs
Dutch between English and German (Van Haeringen 1956)
**Pearson corr.**

<table>
<thead>
<tr>
<th></th>
<th>English</th>
<th>Dutch</th>
<th>German</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Demography</strong></td>
<td>0.87</td>
<td>0.72</td>
<td>0.96</td>
</tr>
<tr>
<td>(p=0.31)</td>
<td>(p=0.28)</td>
<td>(p=0.03)</td>
<td></td>
</tr>
<tr>
<td><strong>Dutch</strong></td>
<td>0.72</td>
<td>0.72</td>
<td>0.97</td>
</tr>
<tr>
<td>(p=0.44)</td>
<td>(p=0.28)</td>
<td>(p=0.03)</td>
<td></td>
</tr>
<tr>
<td><strong>German</strong></td>
<td>0.69</td>
<td>0.56</td>
<td>0.99</td>
</tr>
<tr>
<td>(p=0.31)</td>
<td>(p=0.44)</td>
<td></td>
<td>(p=0.01)</td>
</tr>
</tbody>
</table>
Observing & Simulating

- Observed data
- Simulated data
- Theoretical explanation
- Empirical research
- Computer simulations
Observing & Simulating

Empirical research

- Weakification from inferior type & token frequency
- Conserving Effect
- Class Resilience

General Applicability

Computer simulations

Pijpops, Beuls & Van de Velde (2015)
Observing & Simulating

Empirical research

Different rates of weakification

Computer simulations

Demography

Pijpops, Beuls & Van de Velde (2015)
Parameters:
- Number of series: 20
- Number of agents: 100
- Time: 5,000,000 times units (average interactions per agent)
- Replacement rate: 1/5,000, 1/10,000, 1/20,000, 1/100,000
- Replacement number: 1
- Verbal replacement: none

Pijpops, Beuls & Van de Velde (2015)
Conclusions

• No constant rate of weakification

• Different rates can be explained by language/dialect contact
Thanks!

Acknowledgement: thanks to Ryan Carroll, Ragnar Svare and Joseph Salmons for sharing their dataset

More:


References