The rise of the 19th century English progressive: variation between individual verbs

Alina Ladygina and Igor Yanovich

DFG Center for Advanced Study “Words, Bones, Genes and Tools”, Universität Tübingen

NANRiHS

July 23, 2017
Outline

- Introduction
  - Overview: development of English progressive
  - Motivation for current research
- Our expectations and hypotheses for the 19th century
- Data and methods
- Data analysis
  - Exploratory data analysis
  - Regression modeling
- Concluding remarks

Two “big-picture” ideas of this talk

(i) there is enough data out there to track individual words’ grammatical-change trajectories
(ii) such tracking may lead to surprises!
Development of English progressive

- Old English:
  - Stylistic, non-obligatory usage [Nehl, 1988]
- Middle English:
  - Fluctuations in frequency [Kranich, 2010]
- Early Modern English:
  - Integrating in verbal aspectual system, more consistent use [Elsness, 1994, Núñez-Pertejo, 2004]
- Late Modern English: dramatic rise in frequency [Arnaud, 1998, Smitterberg, 2005]
- Present-Day English: obligatory in "semantically progressive" contexts [Kranich, 2010]
19th century: rapid development of progressive

Late Modern English:
- Crucial period for the process of the progressive integration

19th century specifically:
- Increase in frequency across all genres [Smitterberg, 2005]
- Fast development in colloquial speech (e.g. personal letters, dramas)
The existing works investigate the distribution of progressives across genres or genders [Smitterberg, 2005] → little attention to concrete verbs

Previous studies were conducted on relatively small corpora (e.g., mostly colloquial texts [Arnaud, 1998]) → extrapolation from small amounts of data, rather than examining detailed trajectories of changes

Limited repertoire of statistical methods (chi-squared, factor analysis) → demonstrates the general trend, but not finer-grained variation
Our expectations based on the previous literature

Our initial expectations:
- The frequency of progressives will increase considerably by the end of the period
- Verbs of particular semantic classes (e.g. verbs of movement) will demonstrate faster development

Question:
- How exactly would that expected rise look in individual verbs?
Data: (British) Hansard Archive

- Collection of the (British) House of Commons parliamentary debates (from 1803 to 2004, of which we considered 1803-1889).
  http://www.hansard-archive.parliament.uk/

  **Example**: Mr. Corry said, that for obvious reasons, and out of respect to the House, he should certainly confine himself to explanation, and be very short: what he had to say arose out of the observations of a right hon. gent, whom he never heard without great respect. He should, however, state how the matter stood.
Publicly available data
A very large corpus
Speech of a roughly homogeneous community:
- Speakers are in regular linguistic contact with each other.
- Equal social status.
- The same register of the texts (formal speech).

Caveat: we do not know whether we are dealing with the language of the MPs or the transcribers, but either way the texts have been produced by a focussed group. (Cf. [Mollin, 2008] for recent-times Hansard material.)

Intuition about the progressive: significant increase in frequency in official texts would indicate fossilization of the progressive in the English grammar
The retrieved debates cover the period from 1803 to 1889.

The texts were divided into 9 samples by calendar decades.

Each debate was lemmatized and POS-annotated with TreeTagger [Schmid, 1994]

The 50 most frequent V-ing forms were selected for further analysis
  ...so that we have enough forms to look at the trajectories

The metric we are looking at:

$$\frac{\text{# of progressive forms}}{\text{# of all forms for the same lexical verb}}$$

The idea is that if (i) the progressive starts to become obligatory for some contexts, and (ii) the relative frequency of different types of context stays roughly the same, we will see an increase in our metric.
Extraction of progressives

- Using POS-annotation and lemmas, we extracted finite progressives in present and past tense forms.
- Balance between accuracy and time investment: catch by regular expressions adjacent BE+ing as well as BE+X+ing 3-word sequences
  - After the first pass, examine the Xs in BE+X+ing
    - Retain negation, 1-verb adverbials compatible with the progressive
    - Exclude articles, prepositions, incompatible adverbials
  - After the second pass, exclude verbs with too much noise
- The contexts with non-progressive meanings that were discarded, following [Smitterberg, 2005]'s observations:
  - Construction BE+going+to → futurate meaning
  - Inverted forms (e.g. in questions)
  - Progressives with modal verbs
  - Appositively used gerunds
going
wanting
making
proceeding
doing
acting
speaking
suffering
passing
taking
sitting
endeavouring
preparing
deserving
coming
fitting
giving
pursuing
living
carrying
surprising
labouring
increasing
exceeding
binding
contending
calling
saying
looking
becoming
paying
fighting
alluding
putting
approaching
asking
beginning
stating
lying
serving
treating
advancing
seeking
falling
writing
discussing
pressing
arguing
addressing
returning
Manual processing of retrieval output

- To make sure that the extracted examples represent the true progressive, we investigated randomly selected examples of each verb
  - 20 examples per period per verb (180 sentences in total for each verb)
- Verbs occurring mostly in non-progressive contexts were filtered out

1. That correction is **wanting**

2. The articles were equally **binding** on both countries, and all the House had to do was to see that they were strictly complied with.

- Total # of examples: N=89553
  ([Smitterberg, 2005] employs N=53 for the comparable period in the debates genre, and slightly more than 1000 for the overall dataset.)
Filtered list: 42 forms in total

go
going
wanting
making
proceeding
doing
acting
speaking
suffering
passing
taking
sitting
endeavouring
preparing
deserving
coming
fitting
giving
pursuing
living
carrying
surprising
labouring
increasing
exceeding
binding
contending
calling
saying
looking
becoming
paying
fighting
alluding
putting
approaching
asking
beginning
stating
lying
serving
treating
advancing
seeking
falling
writing
discussing
pressing
arguing
addressing
returning
General trend and trajectories of individual verbs throughout the period

- Metric: proportion of progressive forms to total number of occurrences
- General trend: slight increase in frequency
- Individual tendencies: heterogeneous patterns for each verb
Grouping of verbs sharing similar trajectories

For more detailed investigation via exploratory data analysis, the verbs were divided into three classes:
- Increasing frequency (by at least 1% of the verb’s overall instances)
- Stable frequency
- Decreasing frequency (by at least 1%)
First class: increasing development

Most frequent forms: *suffering, beginning, labouring, endeavouring, acting, pursuing, speaking* ⇒ no clear aspectual or semantic similarities

Progressives increased in frequency (by at least 1%)

![Graph showing the increase in frequency of progressives over time.](image)
Exploratory data analysis

Second class: plateau

Relatively stable progressives

![Graph showing relative frequency of different verb forms over time periods from 1800 to 1880. Verb forms include making, doing, passing, sitting, coming, giving, carrying, surprising, saying, fighting, putting, asking, stating, serving, treating, and returning.]

Alina Ladygina, Igor Yanovich (Tübingen)
Third class: decreasing frequency

Progressives decreased in frequency (by at least 1%)
Is there quantitative support for saying the trajectories differ?

Different regression models can be compared to find out whether adding more parameters leads to a better fit between data and model.
An information-oriented regression investigation

M1: random intercepts (r. i.)

M2: r. i. + individual random slopes

M3: r. i. + random slopes by change category

M4: r. i. + random slopes by sem. category

- Random intercepts: each verb is allowed to have its own starting freq.
- "Random" slopes: trajectories sharing the same slope are all compared to that slope.
  ⇒ if verbs in each category are really similar, the model will have more explanatory power as measured by e.g. AIC.
Regression results

Models:
M1: rel.freq $\sim$ year.trans + (1|verb)
M4: rel.freq $\sim$ year.trans + (1|verb) + (0+year.trans|semgroup)
M3: rel.freq $\sim$ year.trans + (1|verb) + (0+year.trans|traj)
M2: rel.freq $\sim$ year.trans + (1|verb) + (0+year.trans|verb)
M3and2: rel.freq $\sim$ year.trans + (1|verb) + (0+year.trans|traj) + (0+year.trans|verb)

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>AIC</th>
<th>logLik</th>
<th>deviance</th>
<th>Chisq</th>
<th>Chi</th>
<th>Df</th>
<th>Pr(&gt;Chisq)</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>4</td>
<td>-2053.6</td>
<td>1030.8</td>
<td>-2061.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M4</td>
<td>5</td>
<td>-2051.6</td>
<td>1030.8</td>
<td>-2061.6</td>
<td>0.000</td>
<td>1</td>
<td></td>
<td>1.00000000</td>
</tr>
<tr>
<td>M3</td>
<td>5</td>
<td>-2106.4</td>
<td>1058.2</td>
<td>-2116.4</td>
<td>54.795</td>
<td>0</td>
<td></td>
<td>&lt; 2.2e-16</td>
</tr>
<tr>
<td>M2</td>
<td>5</td>
<td>-2132.0</td>
<td>1071.0</td>
<td>-2142.0</td>
<td>25.588</td>
<td>0</td>
<td></td>
<td>&lt; 2.2e-16</td>
</tr>
<tr>
<td>M3and2</td>
<td>6</td>
<td>-2142.1</td>
<td>1077.0</td>
<td>-2154.1</td>
<td>12.080</td>
<td>1</td>
<td></td>
<td>0.0005096</td>
</tr>
</tbody>
</table>
Discussion: is this real?

- Are these differences, ahem, really real?
  - It is notoriously difficult to establish the reality of group differences in change beyond doubt — not just in linguistics, but also in biology
  - But it looks to us like we see a real phenomenon:
    - Frequency changes for each verb remain in the same “region”, so they don’t look like random jumps (artifacts)
    - Exploratory data analysis suggests differences in individual verbs
    - Regression modeling confirms the eyeballing analysis’s results

- We know that something important happened to the progressive.
- Our data suggest that this process was more nuanced one could perhaps think. (Not really a big surprise for people who looked carefully into specific language changes.)
- This has not been shown for the rise of the progressive on this scale
Discussion: what does it mean?

- The rise of the obligatory progressive is a sum of quite different individual-verb trajectories.

- Open question: maybe fine-grained semantics is to blame?
  - Possible story: the progressive is gradually established in novel semantic micro-niches; not all verbs occur in those micro-niches in the first place; as the result, we see frequency rising in some, but not all verbs

- On the other hand, it could also be that individual verbs are truly idiosyncratic...
  - Changes outside phonology are rarely studied on the word-by-word level. Such lexical effects may be very common.

- On the empirical side, we need more studies
- On the theoretical side, perhaps models for explaining the grammar-level and word-level behavior jointly?
Thank you!
Acknowledgements

We would like to thank the anonymous reviewers for their comments, and to acknowledge gratefully the support by DFG under FOR 2237 (DFG Center for Advanced Study “Words, Bones, Genes and Tools”).
Verbs of movement

Trajectories for the verbs of literal and metaphorical movement

- going
- proceeding
- passing
- coming
- pursuing
- approaching
- advancing
- returning

Time periods

Relative frequency

Alina Ladygina, Igor Yanovich (Tübingen)
Appendix

Verbs of speech

Trajectories for the speech verbs

- speaking
- saying
- asking
- stating
- discussing
- arguing

Time periods
Relative frequency
1800 1820 1840 1860 1880
0.00 0.05 0.10 0.15 0.20 0.25

Alina Ladygina, Igor Yanovich (Tübingen)

The rise of the 19th cent. progressive
The development of the progressive in 19th century English: A quantitative survey.

On the progression of the progressive in early Modern English.

The progressive in Modern English. A Corpus-Based Study of Grammaticalization and Related Changes.

The Hansard hazard: gauging the accuracy of British parliamentary transcripts.

On the Development of the Grammatical Category of Verbal Aspect in English.

Lincom, München.

In *Proceedings of International Conference on New Methods in Language Processing*, Manchester, UK.

Language and computers : studies in practical linguistics. Rodopi.