Analyzing historical changes in Irish English adjective amplification

Slides available at
www.martinschweinberger.de
m.schweinberger@uq.edu.au
R code upon request

Dr. Martin Schweinberger
Phenomenon: Adjective Amplification

(1) Adjective amplification in predicative contexts
   a. they were very kind to me on bord (HCIE:Alderd01:1880)
   b. John was prety sick all spring (HCIE:CarroA36:1887)
   c. I am so glad (HCIE:Burke:03:1884)
   d. I was really glad to hear that (HCIE:Spratt01:1869)

(2) Adjective amplification in attributive contexts
   a. I have very good health (HCIE:GambJ:06:1876)
   b. I have pretty good wages at present
      (HCIE:ButleJ02:1850)
   c. we have so large grounds to pley (HCIE:Sloane04:1892)
   d. I got a real nice Passage (HCIE:Markey01:1889)
Phenomenon: Adjective Amplification

- Subtype of intensification
- Related to the semantic category of *degree* (degree adverbs) and ranges from low (downtoners) to high (amplifiers) (Quirk et al. 1985: 589–590)
  - Amplifiers
    - Boosters, e.g. *very*
    - Maximizers, e.g. *completely*
  - Downtoners
    - Approximators, e.g. *almost*
    - Compromisers, e.g. *more or less*
    - Diminishers, e.g. *partly*
    - Minimizers, e.g. *hardly*
Previous Research

Amplification

- Extensive history of research and a substantial amount of corpus-based research on intensification (e.g. Aijmer 2011, 2018; Bauer and Bauer 2002; Borst 1902; Bolinger 1972; D'ArCY 2015; Fuchs 2016, 2017; Ito and Tagliamonte 2003; Núñez Pertejo and Palacios 2014; Palacios and Núñez Pertejo 2012; Tagliamonte and Roberts 2005; Macaulay 2006; Tagliamonte 2006, 2008)

- Intensification is considered a major area of grammatical change in English which makes it an ideal test case for processes of language change (cf. Brinton and Arnovick 2006: 441)

- Crucial for “social and emotional expression of speakers” (Ito and Tagliamonte 2003: 258)
Previous Research

- Amplifying really replaces very (lexical replacement)
  
  (see D’Arcy (2015) for NZE; see Ito and Tagliamonte (2003) and Barnfield and Buchstaller (2010) for North East British English, Tagliamonte (2008) and Tagliamonte and Denis (2014) for Toronto English; see Tagliamonte and Denis (2014) for South Eastern Ontario English)

Figure 1: Adapted from D’Arcy (2015: 468)
Previous Research

- Amplifying *really* replaces *very* (lexical replacement)

(see Schweinberger (subm) for Irish English)

Figure 2: Adapted from Schweinberger (subm)
Motivation

Adjective amplification has been described as a

- “major area of grammatical change” (cf. Brinton and Arnowick 2006: 441)

- “[…] waxing and waning of forms, alongside invention and renewal” (D’Arcy 2015: 450)

- “Competition, change and recycling among intensifiers have been going on in English since the Old English period” (Tagliamonte 2008: 361)

- “in this sea of change, processes of expansion and contraction are occurring all the time” (Partington 1993: 180)
Research Question

Q

How did the adjective amplifier system in Irish English evolve during the last 300+ years?

→ Is adjective amplification really a site of perpetual waxing and waning of forms?
INTRODUCTION

DATA AND METHODOLOGY

THE HAMBURG CORPUS OF IRISH ENGLISH

DATA PROCESSING

STATISTICAL PROCEDURES

RESULTS

DISCUSSION & OUTLOOK
Data and Methodology
Analyzing historical changes in Irish English adjective amplification  

Data and Methodology

The Hamburg Corpus of Irish English (HCIE)
Hamburg Corpus of Irish English (HCIE)

- Letters written by Irish emigrants between 1675 and 1935
- App. 1,000 letters amounting to app. 500,000 words
- Accompanied by a limited range of metadata (e.g. year of composition) and information about the “speakers” (gender)

<table>
<thead>
<tr>
<th>Period</th>
<th>Letters (N)</th>
<th>Words (N)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1675-1750</td>
<td>16</td>
<td>9,818</td>
</tr>
<tr>
<td>1751-1850</td>
<td>355</td>
<td>225,032</td>
</tr>
<tr>
<td>1901-1935</td>
<td>608</td>
<td>322,107</td>
</tr>
<tr>
<td>Unknown</td>
<td>1</td>
<td>223</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>980</strong></td>
<td><strong>557,180</strong></td>
</tr>
</tbody>
</table>
Hamburg Corpus of Irish English (HCIE)

- Semi-literate authors (dictated or self-written) who used a profoundly speech-like style (the most speech-like data we will get. . . )
- Lower social strata (not upper-class, highly literate and multilingual writers as in the INNSBRUC letter corpus)
- Letters predominantly from the US, but also from Australia and New Zealand (but substantially fewer)
Data Processing
Data Processing

- Cleaned and automatically corrected/spell-checked the letters
- Part–of-speech tagged (OpenNLP vis R) the cleaned letters
- Retrieved adjectives (PoS–tag JJ)
- Determined whether adjective were preceded by an amplifier (member of a predefined set of amplifiers)
- Manual cross–evaluation of automated classification
- Metadata and speaker information
Data Processing

Removed...
- negated adjectives
- comparative and superlative forms
- adjectives that were not amplified by at least two different amplifier types
- adjectives that were preceded by downtoners
- strange forms (e.g. *much*)
Statistical procedures
Lexical diversity scores

Dividing the number of adjective types a given amplifier co-occurs with by the number of tokens of that amplifier type (see 3).

\[ (3) \quad LD = \frac{N_{Adj.\ Types}}{N_{Amp.\ Tokens}} \]

LD can reach a maximum value of 1 (high lexical diversity). The lower the LD value, the lower the degree of lexical diversity.

<table>
<thead>
<tr>
<th>Amplifier</th>
<th>Amp. Tokens (N)</th>
<th>Adj. Types (N)</th>
<th>Calculation</th>
<th>LD</th>
</tr>
</thead>
<tbody>
<tr>
<td>variantA</td>
<td>10</td>
<td>1</td>
<td>1/10</td>
<td>.1</td>
</tr>
<tr>
<td>variantB</td>
<td>10</td>
<td>5</td>
<td>5/10</td>
<td>.5</td>
</tr>
<tr>
<td>variantC</td>
<td>10</td>
<td>10</td>
<td>10/10</td>
<td>1</td>
</tr>
<tr>
<td>very1675−1750</td>
<td>11</td>
<td>17</td>
<td>11/18</td>
<td>.647</td>
</tr>
<tr>
<td>very1751−1850</td>
<td>86</td>
<td>368</td>
<td>86/368</td>
<td>.234</td>
</tr>
<tr>
<td>very1851−1930</td>
<td>102</td>
<td>845</td>
<td>102/845</td>
<td>.121</td>
</tr>
</tbody>
</table>
Covarying collexeme analysis

- Member of the collostructional family of analyses (Hilpert (2006); Stefanowitsch and Gries (2005))

- Quantifies and evaluates of attraction between elements that occur in two distinct slots within a specified construction (here: 1\textsuperscript{st} slot = amplifier slot; 2\textsuperscript{nd} slot = adjective slot)

- How much more likely is it that nice occurs in the 2\textsuperscript{nd} slot given that really occurs in the 1\textsuperscript{st} slot compared with another amplifier taking the 1\textsuperscript{st} slot?

- High values = strong attraction: the amplifier and adjective co-occur more often together than would be expected by chance;

  Low values = the amplifier and adjective co-occur less often together than would be expected by chance.
Semantic vector space models (Levshina 2015)

Semantic Vector Space Models (SVSM) are based on co-occurrence profiles of adjectives and amplifiers

- Idea: words that have similar collocational profiles are similar in meaning; e.g. amplifiers that co-occur with the same adjectives are semantically similar

- Advantages of such models are that SVSM . . .
  - provide a data-driven measure of semantic similarity (do not rely on any theoretical framework)
  - take variety specific differences in use/meaning into account (fine-grained classifications)
  - inform about statistical significance
Semantic vector space models

Example

<table>
<thead>
<tr>
<th></th>
<th>get</th>
<th>see</th>
<th>use</th>
<th>hear</th>
<th>eat</th>
<th>kill</th>
</tr>
</thead>
<tbody>
<tr>
<td>knife</td>
<td>31</td>
<td>16</td>
<td>69</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>cat</td>
<td>36</td>
<td>38</td>
<td>4</td>
<td>4</td>
<td>6</td>
<td>20</td>
</tr>
<tr>
<td>???</td>
<td>66</td>
<td>58</td>
<td>9</td>
<td>34</td>
<td>28</td>
<td>12</td>
</tr>
<tr>
<td>boat</td>
<td>46</td>
<td>21</td>
<td>17</td>
<td>4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>cup</td>
<td>59</td>
<td>6</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>pig</td>
<td>4</td>
<td>15</td>
<td>3</td>
<td>1</td>
<td>7</td>
<td>21</td>
</tr>
<tr>
<td>banana</td>
<td>7</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>12</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 1: Co-occurrences of selected nouns and verbs in the British National Corpus.
## Semantic vector space models

<table>
<thead>
<tr>
<th></th>
<th>knife</th>
<th>cat</th>
<th>???</th>
<th>boat</th>
<th>cup</th>
<th>pig</th>
</tr>
</thead>
<tbody>
<tr>
<td>cat</td>
<td>.62</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>???</td>
<td>.60</td>
<td>.21</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>boat</td>
<td>.48</td>
<td>.33</td>
<td>.32</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cup</td>
<td>.76</td>
<td>.59</td>
<td>.58</td>
<td>.36</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pig</td>
<td>.72</td>
<td>.36</td>
<td>.45</td>
<td>.64</td>
<td>.88</td>
<td></td>
</tr>
<tr>
<td>banana</td>
<td>.71</td>
<td>.57</td>
<td>.47</td>
<td>.60</td>
<td>.72</td>
<td>.64</td>
</tr>
</tbody>
</table>

Table 2: Distance matrix based on co-occurrences of selected nouns and verbs in the British National Corpus.
Cluster Dendrogram

Height

banana
pig
cat
??
knife
boat
cup

mx
hclus (*, "ward.D")

Dr. Martin Schweinberger
Results
Variants across real-time

Figure 3: Percent of amplifier variants in IrE (HCIE) across real-time.
Interpretation

Variants across real-time

- Attributive contexts
  - Remarkable(!) stability and stasis
- Predicative contexts
  - Little change in until 1900
  - 1900 and later: so replaces very

Q

Is the replacement of very by so accompanied by changes in co-occurrence patterns?
Scenario 1 (Broadening)

so increases because it associates with many (but infrequent) adj. types

(Mair 2004: “delayed increase of discourse frequency” hypothesis)

Prediction

Co-occurrence with many different adjective types

→ high lexical diversity

→ weak coll. attraction with specific adj. types
Scenario 2 (Specialization)

so increases because it associates with few but frequent adj. types (HFAs)

(Lorenz 2002: 144; Méndez-Naya 2003: 375; Tagliamonte and Roberts 2005: 285)

Prediction

Co-occurrence with few high frequency adjectives

→ low lexical diversity

→ strong coll. attraction with high-freq. adj. types
Scenario 3 (Randomness)

so increases because it associates with random adj. types
→ We cannot predict which variants become successful based on their coll. profile.
Lexical diversity scores of amplifier variants across real-time

Figure 4: Lexical diversity scores in IrE (HCIE) across real-time.
Interpretation

The increase in *so* is not accompanied by any discernible increase or decrease in lexical diversity...

Q

Could the change in the amplifier system be triggered by changes in the frequency of adjectives?
Change in adjective types across real-time

Figure 5: Adjective types in IrE (HCIE) across real-time.
Interpretation

The increase in *so* is not triggered by any discernible change in the frequencies of adjectives. . .

Q

Could the change in the amplifier system be based on changes in the collocation profile of amplifier-adjective bigrams?
Changes in collocation strength across real-time (covarying collexeme analysis)

Figure 6: Collocation strength in IrE (HCIE) across real-time.
Interpretation

In the newest data, so is increasingly co-occurring with good (the most frequent adjective) and glad (another HFA). These changes in collocation strength are, however, relatively small and not significant...

Could the change in the amplifier system be dependent upon semantic similarity?
Figure 7: Semantic similarity of amplifier types in IrE (HCIE).
Figure 8: Semantic similarity of amplifier types in IrE (HCIE).
Interpretation

so and very are, indeed, most similar with respect to their collocational profiles. This suggest that replacement is dependent upon semantic similarity (Schweinberger (fcb))! However, this does not explain why so replaces very in the HCIE data in the first place. . .

Q

Is this pattern unique to Irish English?

→ Is this an isolated phenomenon?
Figure 9: Percent of amplifier variants in AmE (COHA) across real-time. Schweinberger (fca).
Figure 10: Semantic similarity of amplifier types in AmE (COHA) Schweinberger (fca).
Figure 11: Semantic similarity of amplifier types in AmE (COHA) Schweinberger (fca).
Discussion & Outlook
Summary

The analysis shows that . . .

- adjective amplification can show prolonged phases of stasis (no or little changes in relative frequency)
  → amplification is, at least not unconditionally, a domain of “[. . .] waxing and waning of forms” (D’Arcy 2015: 450)

- observed change is restricted to predicative contexts (thus making it unlikely that so replaces very long-term or across moralities (as is the case with really in informal spoken discourse)

- prolonged phases of stasis can be interrupted by sudden shifts which contradicts the notion that change is (always) gradual and step-wise
  → punctuated equilibrium!?!
Discussion

- so can successfully replace the dominant form (very) because both variants are semantically similar → replacement requires semantic similarity! (see also Schweinberger (fcb))
- No signs that so broadens before taking over the system.
- Moderate signs that so specialized during its increase (increase in collocation strength with good and glad)
Problems and Outlook

Problems

- Uneven distribution of letters across time periods (maybe use CORICOR for future research)
- Speaker-variability is not yet included into the models (run a mixed-effects model with speakers as random effect)
- No answer to the question as to what triggered the increase in *so* in both IrE and AmE...

Outlook

- Similar findings have been described for data from the fiction(!) section of the *Corpus of Historical American English* (COHA)
  → universal trend in “speech-like” writing or is it a contact phenomenon or something else?
Thank you so, really, very much!

Acknowledgements

I would like to thank...

Peter Siemund and his team (Lukas Pietsch, Kalynda Beal)
for providing me with a preliminary version of the HCIE
(without them the current study would not have been possible)

my colleagues at UQ
for comments and their feedback on earlier versions of this talk

This study will appear as:
Schweinberger, Martin. accepted. Analyzing Historical Changes in the
Irish English Amplifier System. Anglistik – International Journal
of English Studies (Special Issue: Focus on English Linguistics -
Varieties meet Histories, Eds. Daniela Kolbe-Hannah & Ilse Wischer)


Palacios, I. and P. Núñez Pertejo (2012). He’s absolutely massive. it’s a super day. madonna, she is a wicked singer. youth language and intensification: A corpus-based study. Text and Talk 32(6), 773–796.


Analyzing historical changes in Irish English adjective amplification

Slides available at www.martinschweinberger.de
m.schweinberger@uq.edu.au
R code upon request

Dr. Martin Schweinberger