Vector-space semantic maps

A data-driven approach to the study of syntactic productivity in diachrony

Florent Perek
Princeton University
fperek@princeton.edu
Syntactic productivity

• Property of a construction to attract new lexical fillers

• In diachrony:
  – The distribution of constructions may vary over time
  – e.g., verb slot in the way-construction (Israel 1996)
    • Verbs of physical actions attested from the 16th century
      \textit{They hacked their way through the jungle.}
    • Abstract means of reaching a goal only appear in the 19th century
      \textit{She typed her way to a promotion.}

• Appears to be partly arbitrary, but actually tied to prior usage
Usage-based determinants of syntactic productivity

Type frequency

Goldberg (1995)
Bybee and Thompson (1997)
Barðdal (2008)
Wonnacott et al. (2012)

Semantic variability

Goldberg (2006)
Barðdal (2008)

Semantic similarity

Bybee and Eddington (2006)
Bybee (2010)
Suttle and Goldberg (2011)
Previous research

• Points to a strong semantic component in syntactic productivity
  – Productivity depends on the structure of the semantic space
  – Novel uses must be semantically consistent with prior usage
  – cf. the notion of coverage (Suttle and Goldberg 2011, Osherson et al. 1990)
    “the degree to which attested instances ‘cover’ the category determined jointly
    by attested instances together with the target coinage” (Suttle and Goldberg
    2011: 1254)

• How to operationalize semantics?
  – In previous studies: introspection, semantic norming study (Bybee and
    Eddington 2006)
  – Proposal: using distributional semantics to measure semantic similarity
Distributional semantics

“You shall know a word by the company it keeps.” (Firth 1957: 11)

• Words that occur in similar contexts tend to have related meanings (Miller and Charles 1991)

• Therefore, a way to characterize the meaning of words is through their distribution in large corpora

• Semantic similarity is quantified by similarity in distribution
Distributional semantics

- Vector-space model
  - Assigns an array of values (i.e., a vector) derived from distributional information to each word
  - Semantic similarity measured by similarity between vectors
  - Here, ‘bag of words’ approach: based on lexical co-occurrences
- Example: shared collocates of *drink*, *sip*, *eat*, and *hear*
  - *drink* and *sip*: names for beverages (*beer*, *coffee*, *tea*), containers for liquids (*glass*, *cup*, *bottle*)
  - *eat* and *drink*/*sip*: words related to dining practices (*bar*, *table*, *dinner*)
  - *hear* share very few collocates with the first three
Case study: The “hell-construction”

- V the hell out of NP, e.g., You scared the hell out of me!
- Intensifying function (broadly defined)
- Scare and beat most typical, but also a wide range of other verbs:
  
  Then I [...] avoided the hell out of his presence
  But you drove the hell out of it!
  I've been listening the hell out of your tape.

I know the hell out of women!

I voice the hell out of ‘b’
(Phillip Hamrick yesterday)
The *hell*-construction in diachrony

- Data from the Corpus of Historical American English (Davies 2010)
- First attestations in the 1930s
- Steady increase in token frequency since then
The *hell*-construction in diachrony

- More and more verb types are used in the construction
The *hell*-construction in diachrony

- The increase in type frequency reflects an increase in productivity
- But what kinds of verbs joined the distribution?
  - Did it become more semantically diverse?
  - Are there particular semantic domains favored by the construction?
- Proposal: track the semantic development of the construction by using distributional semantics
Method

• Vector-space model
  – Data from COCA (Davies 2008)
  – Collocates within a 5-word window, lemmatized and PoS-tagged
  – Nouns, verbs, adjectives, and adverbs from the 5,000 most frequent words

• Output of vector-space model: distance matrix
  – Pairwise semantic distances between verbs
  – Cosine distance (between 0 and 1)
Evaluation of the model

- Hierarchical cluster analysis: groups objects together in a hierarchy by recursively merging the nearest neighbors
- Output: tree diagram (dendrogram)
- Useful to visualize the major semantic distinctions in the distribution
mental verbs (emotions, feelings, cognition, ...)

*amuse, bore, frighten, like, puzzle, surprise, worry, ...*
forceful, violent actions

verbs of hitting, exertion of force, destruction

other concrete actions

concrete
How to visualize the semantic domain of the construction?

- Can be plotted by means of multidimensional scaling (MDS)
  - Places objects in a 2-dimensional space such that the between-object distances are preserved as well as possible
  - Distance matrix converted to a set of coordinates

- Four vector-space semantic maps (1 per 20-year period)
  - 1930-1949
  - 1950-1969
  - 1970-1989
  - 1990-2009
Summary

- Densely populated regions are more likely to attract new members
- New verbs appear either close to or inside a cluster
- In line with previous accounts
- Can we derive quantitative evidence for these observations?
Statistical analysis

• Assumptions
  – Semantic contribution of the construction is constant
  – Hence, all verbs that ever occurred in it are equally plausible
  – Why do some verbs first occur later than others?

• Hypothesis
  The probability of occurrence of a new item is related to the density of the semantic space around this item in prior usage
Measure of density

- Considers a subset of the nearest neighbors of a verb
- Density = 1 – mean distance to the $N$ nearest neighbors
- $N = 3$ to $8$

with $N = 3$

$$\text{Density} = 1 - \frac{0.1 + 0.1 + 0.1}{3} = 0.9$$

with $N = 4$

$$\text{Density} = 1 - \frac{0.3 + 0.1 + 0.1 + 0.1}{4} = 0.850$$
Mixed effects logistic regression

- Dependent variable: Occurrence (0/1)
- Data: Verb x Period x Occurrence triplets
- Density around the verb calculated for the preceding period
- For each verb, Occurrence = 1 on its first attestation, 0 in earlier periods
- Model:
  \[ \text{Occurrence} \sim \text{Density} + (1 + \text{Density} \mid \text{Verb}) \]
Results

- Positive effect of Density, significant for N = 7 and 8
  - Evidence for a relation between density and the likelihood of a new coinage
- Effect strength increases and $p$-value decreases with N
  - Considering a higher number of neighbors increases the strength and reliability of the measure of density as a predictor of novel uses

<table>
<thead>
<tr>
<th>Number of neighbors (N)</th>
<th>Effect of Density</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0.7211</td>
<td>0.195</td>
</tr>
<tr>
<td>4</td>
<td>0.8836</td>
<td>0.135</td>
</tr>
<tr>
<td>5</td>
<td>1.0487</td>
<td>0.091</td>
</tr>
<tr>
<td>6</td>
<td>1.2367</td>
<td>0.056</td>
</tr>
<tr>
<td>7</td>
<td>1.4219</td>
<td><strong>0.034</strong> *</td>
</tr>
<tr>
<td>8</td>
<td>1.6625</td>
<td><strong>0.017</strong> *</td>
</tr>
</tbody>
</table>
The role of $N$: example with *adore*

<table>
<thead>
<tr>
<th>$N = 3$</th>
<th>$1930s - 1940s$</th>
<th>$1950s - 1960s$</th>
</tr>
</thead>
<tbody>
<tr>
<td>adore?</td>
<td></td>
<td>love</td>
</tr>
<tr>
<td>love</td>
<td>surprise</td>
<td>adore?</td>
</tr>
<tr>
<td>surprise</td>
<td>whip lick bore</td>
<td>impress</td>
</tr>
<tr>
<td>whip</td>
<td>lick</td>
<td>love</td>
</tr>
<tr>
<td>lick</td>
<td>bore</td>
<td>adore?</td>
</tr>
<tr>
<td>chase</td>
<td></td>
<td>impress</td>
</tr>
</tbody>
</table>

for $N = 3$: little variation in density

<table>
<thead>
<tr>
<th>$N = 7$</th>
<th>$1930s - 1940s$</th>
<th>$1950s - 1960s$</th>
</tr>
</thead>
<tbody>
<tr>
<td>adore?</td>
<td></td>
<td>love</td>
</tr>
<tr>
<td>love</td>
<td>surprise</td>
<td>adore?</td>
</tr>
<tr>
<td>surprise</td>
<td>whip lick bore</td>
<td>impress</td>
</tr>
<tr>
<td>whip</td>
<td>lick</td>
<td>love</td>
</tr>
<tr>
<td>lick</td>
<td>bore</td>
<td>adore?</td>
</tr>
<tr>
<td>chase</td>
<td></td>
<td>impress</td>
</tr>
<tr>
<td>bore</td>
<td></td>
<td>love</td>
</tr>
<tr>
<td>please</td>
<td></td>
<td>adore?</td>
</tr>
</tbody>
</table>

for $N = 7$: density increases
Conclusion

- Distributional semantics is appropriate for the study of syntactic productivity in diachrony; benefits:
  - Turns the informal notion of meaning into a quantified representation
  - Fully automatic and data-driven
  - Virtually no limit on the number of items to be considered
  - Enables the use of visualization techniques and statistical analysis
- Distribution-based account consistent with current views
- Promising approach to the study of syntactic productivity
I thank the hell out of you!


