Distributional characterization of constructional meaning

Florent Perek
Albert-Ludwigs-Universität Freiburg / Université Lille III
florent.perek@gmail.com

Corpus Linguistics 2009 – Liverpool – 23th July 2009
• Goal
  – Design and test ways to derive the meaning of grammatical constructions from corpus data
• Outline
  – Theoretical background
  – First corpus study
    • Distributional comparison of constructions
    • Limits of this approach
  – Follow-up study based on LSA
  – Conclusion and prospects
Theoretical background

- Grammar in Cognitive Linguistics
  - Grammar = inventory of form-meaning pairs
  - No principled separation between syntax and lexicon
  - Syntactic patterns = form-meaning pairs
    - Argument Structure Constructions (Goldberg 1995, 2006)
    - Syntactic meaning
      - = schematized experience of an event type: transfer, movement, change of state, ...
      - Most evident when verbs are used creatively
        - e.g. John sneezed the napkin off the table
        - Predicts which verbs are allowed
    - Experimental evidence for constructions
Theoretical background

• Example: the ditransitive construction
  (from Goldberg 1995)

  e.g. Mary gave her sister a penny.
  Sam kicked Peter the ball.

  Syntax: Subject \_\_ Agent V Object1 \_\_ Recipient Object2 \_\_ Theme
  Semantics: Agent CAUSES Recipient TO RECEIVE Theme
Theoretical background

• The origin of constructional meaning
  – Current hypothesis: abstraction of verbal meaning
    • e.g. NP V NP NP (ditransitive)
      occurs with verbs of transfer: give, send, hand, …
      => the syntactic pattern is associated to a transfer meaning
    • Evidenced by the facilitating factor of a biased distribution
      (cf. Goldberg et al. 2004)
  – We investigated this idea with two corpus studies
Study 1

- Hypothesis: distributional distance correlate with semantic distance
  - Supported by:
    - Goldberg's model: constructional meaning constrains the verbs occurring in the construction
    - Corpus studies
      - Stefanowitsch and Gries (2003), Gries et al. (2005)
      - The strongest collocates have the meaning closest to that of the construction
  - Constructions with different meaning should have different distributions, and conversely
  - Distributional differences should reflect semantic differences between constructions
Study 1

• How to compare distributions?
  – Vector space approach to distributional similarity
    • Verbal distribution = vector (of frequencies)
    • The verbs define a multidimensional space
    • Distributional distance = distance between vectors
**Study 1**

- **Simple example with only 2 dimensions**

<table>
<thead>
<tr>
<th>Verb</th>
<th>Frequency in the ditransitive</th>
<th>Frequency in the caused motion</th>
</tr>
</thead>
<tbody>
<tr>
<td>give</td>
<td>153</td>
<td>34</td>
</tr>
<tr>
<td>send</td>
<td>12</td>
<td>18</td>
</tr>
</tbody>
</table>

\[
\text{ditransitive} = \vec{C}_1 = \begin{bmatrix} 153 \\ 12 \end{bmatrix} \\
\text{caused motion} = \vec{C}_2 = \begin{bmatrix} 34 \\ 18 \end{bmatrix}
\]

\[
\text{similarity} = \cos(\vec{C}_1, \vec{C}_2)
\]

\[
\cos(\vec{C}_1, \vec{C}_2) \in [0,1]
\]

1 \(\Rightarrow\) identity

the closer to 0, the more different
Study 1

- Six constructions under study

<table>
<thead>
<tr>
<th>Construction</th>
<th>Semantics</th>
<th>Syntax</th>
<th>Example</th>
<th>Freq.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intransitive Motion</td>
<td>$X_{\text{Theme}}$ MOVE $Y_{\text{Path}}$</td>
<td>Subject-Verb-Oblique$_{\text{Path}}$</td>
<td>The ball rolled down the hill</td>
<td>564</td>
</tr>
<tr>
<td>Change of state</td>
<td>$X_{\text{Theme}}$ BECOME $Y_{\text{State}}$</td>
<td>Subject-Verb-Oblique$_{\text{State}}$</td>
<td>The pond froze solid</td>
<td>471</td>
</tr>
<tr>
<td>Caused motion</td>
<td>$X_{\text{Agent}}$ CAUSE $Y_{\text{Patient}}$ TO MOVE $Z_{\text{Path}}$</td>
<td>Subject-Verb-Object-Oblique$_{\text{Path}}$</td>
<td>Bill broke the hell into the bowl</td>
<td>290</td>
</tr>
<tr>
<td>Resultative</td>
<td>$X_{\text{Agent}}$ CAUSE $Y_{\text{Patient}}$ TO BECOME $Z_{\text{State}}$</td>
<td>Subject-Verb-Object-Oblique$_{\text{State}}$</td>
<td>Bill watered the tulip flat</td>
<td>175</td>
</tr>
<tr>
<td>Ditransitive</td>
<td>$X_{\text{Agent}}$ CAUSE $Y_{\text{Recipient}}$ TO RECEIVE $Z_{\text{Theme}}$</td>
<td>Subject-Verb-Object1-Object2</td>
<td>Joe painted Sally a picture</td>
<td>307</td>
</tr>
<tr>
<td>Conative</td>
<td>$X_{\text{Agent}}$ DIRECT ACTION AT $Y_{\text{Target}}$</td>
<td>Subject-Verb-Oblique$_{\text{at}}$</td>
<td>Bill kicked at the ball</td>
<td>178</td>
</tr>
</tbody>
</table>

- Manually identified in the ICE-GB (spoken)
- Extraction of the verbal distribution
Study 1

- Comparison of all six constructions
Study 1

• Possible improvement
  – Collexeme analysis (Stefanowitsch and Gries 2003)
    • Some verbs are more important collocates for some constructions than for others
    • Not rendered by raw frequencies
    • Collostruction strength instead of frequencies
• With collexeme analysis
Interim conclusion

- Do reflect relevant meaning similarities
  - Change of state / change of location (cf. Goldberg & Jackendoff 2004)
  - Caused motion / transfer
- But purely distributional
  - Does not take the meaning of verbs into account
  - Would the result be different if we do?
- Two issues:
  - Representation of verbal meaning
  - Representation of constructional meaning derived from the former
Interim conclusion

• How to represent word meaning in a corpus?
  – Semantic annotations (e.g. WordNet ids)
  – Distributional characterization

• Latent Semantic Analysis
  – Used in data mining
  – Based on co-occurrences of words in documents
  – Correlates with human judgements on semantic similarity, cf. Landauer, Foltz and Laham (1998)
  – Often used as an objective measure of semantic similarity, e.g. in Bencini & Goldberg (2000)
Word meaning in LSA = a vector
- Semantic distance = distance between vectors
- Construction vectors can be derived from the vectors of the verbs in its distribution

\[ \vec{C} = \sum (freq(V) \times \vec{V}) \]

E.g. the ditransitive

<table>
<thead>
<tr>
<th>Verb</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>allow</td>
<td>2</td>
</tr>
<tr>
<td>ask</td>
<td>6</td>
</tr>
<tr>
<td>write</td>
<td>1</td>
</tr>
</tbody>
</table>

\[ \text{ditransitive} = 2 \times \text{allow} + 6 \times \text{ask} + \ldots + 1 \times \text{write} \]

- Sum of vectors = vector in the same space
- Same distance metrics than for words
Study 2

- Corpus submitted to LSA
- Two representations of constructional meaning:
  - Vectors calculated from word vectors + distribution
  - Vector of the construction symbol (=word) artificially inserted in the corpus
- Questions:
  - To what extent do symbolic meaning and distributionally derived meaning correlate in the LSA space?
  - How semantic differences between constructions are reflected if verbal meaning is taken into account?
Comparison of symbolically derived vs. distributionally derived constructional meaning

- They are strongly similar

<table>
<thead>
<tr>
<th>Construction</th>
<th>Similarity between construction symbol and distribution vector</th>
</tr>
</thead>
<tbody>
<tr>
<td>Change of state</td>
<td>0.7981</td>
</tr>
<tr>
<td>Resultative</td>
<td>0.7391</td>
</tr>
<tr>
<td>Caused motion</td>
<td>0.7378</td>
</tr>
<tr>
<td>Ditransitive</td>
<td>0.7157</td>
</tr>
<tr>
<td>Conative</td>
<td>0.7649</td>
</tr>
<tr>
<td>Intransitive Motion</td>
<td>0.8322</td>
</tr>
</tbody>
</table>

Seems to validate the view that constructional meaning originates with verbal meaning
Study 2

- Comparison of distributionally derived constructional meaning

- No sharp differences between constructions
- The most frequent words are not distinctive enough
- They appear in too wide a range of contexts
- LSA does not capture sharp semantic differences

A possible answer
• Solution: use “stopwords”
  – Frequent verbs are simply ignored in the analysis
  – Semantic differences between constructions are captured by the less frequent verbs
A slight improvement
- Conative evidently different from other constructions
- But still no clear differences
• Conclusion
  – Distribution captures semantic differences between constructions
  – Less so clear with an account of verbal meaning
    • Seems to capture the meaning of constructions
    • But not semantic differences between constructions

• Prospects
  – Bigger corpus to derive word vectors from
  – Use dictionary-based semantic distance
    • more reliable, especially for highly frequent verbs
References


