Sparse Usage Graphs as Model for Word Meaning in Context

November 27, 2020

Dominik Schlechtweg
Supervisor: apl. Prof. Dr. Sabine Schulte im Walde
Institute for Natural Language Processing, University of Stuttgart, Germany
Word Meaning

*a word’s meaning is the knowledge a word can trigger in a human*

(p. 54 Blank, 1997)
Word Meaning

Figure 1: Blank (1997)'s levels of word meaning (p. 95).
Word Meaning

- has traditionally been equated to a set of senses ("polysemy", Bréal 1897)
- abstractions over patterns of use (Kilgarriff, 2007)
- each usage of a word in context disambiguates it / “determines” one of its senses (Navigli, 2009)
History in CL

- before second half of 20 century mainly historical (cf. Blank, 1997)
- WSD (Weaver, 1949/1955)
- Cognitive Semantics, Prototype Theory (Rosch & Mervis, 1975)
- discrete to graded (Erk, McCarthy, & Gaylord, 2009, 2013)
- “I don’t believe in the word senses” (Kilgarriff, 1997)
Word senses are a lexicographer’s attempt to impose order on a word’s role in the language. [...] [The process of a] speaker’s understanding of a word [...] is not one which naturally gives rise to a set of distinct senses.

(Kilgarriff, 2007)

[Polysemy is the] consciously experienceable, intersubjectively comprehensible existence of a semantic relationship between two distinct meanings of a word.

(p. 424 Blank, 1997)
The lexicographic process

1. gather a corpus of citations for a word;
2. divide the citations up into clusters, so that, as far as possible, all the members of each cluster have more in common with any other member of that cluster, than with any member of any other cluster;
3. for each cluster, (post-hoc) work out what it is that makes its members belong together; and
4. take these conclusions and code into a dictionary definition.

(Kilgarriff, 2007)
The lexicographic process

- lexicographer’s criteria in step 2 are not explicit and subjective (Kilgarriff, 1997, 2007)
- we make criteria explicit and inter-subjective
- operationalize Blank’s criteria
- gives rise to a graded scale of word meaning in context (Erk et al., 2013)
Advancing the graded view

- graded word sense annotation (Erk et al., 2013)
- Word Usage Graphs (WUGs)
- idea first proposed in McCarthy, Apidianaki, and Erk (2016)
- first operationalized in Schlechtweg, McGillivray, Hengchen, Dubossarsky, and Tahmasebi (2020)
- large-scale, multi-lingual, diachronic, resource of WUGs based on 100,000 judgments
Word Usage Graphs

Figure 2: Usage graph of Swedish *ledning*. 
Annotation

- 100–200 changing words selected from etymological dictionaries (OED, 2009; Paul, 2002; Svenska Akademien, 2009)
- pre-annotation (rough filtering by one annotator)
- adding of control words with similar frequency properties
- sample 100 uses (30 for Latin) of each word per time period
- graded word sense annotation (Erk et al., 2013)
Data

Table 1: Annotation Table.
Table 2: Blank (1997)'s continuum of semantic proximity (left) and the DURel relatedness scale derived from it (right).
Identity

My arm hurts.

vs.

She grabbed his arm after he had told her the news.
Context variance

\[ \text{My arm hurts.} \]

\[ \text{vs.} \]

\[ \text{Look at the arm of the statue.} \]
Polysemy

My arm hurts.

vs.

An arm of the sea.
My arm hurts.

vs.

The number of men under arms is no longer the decisive factor in warfare.
(1) 1830 but I am bound and thrown into a dark cell.
(2) 1851 ...be fit to burn in a jail; no, not in a condemned cell.

... 

(3) 1990 But I never would return To my cold prison cell. What’s life without liberty?
(4) 2006 She searched the bag for her cell as we headed toward the door.
### Diachronic Data

#### Table 3: Annotation Table.

<table>
<thead>
<tr>
<th>sentence 1</th>
<th>judgment</th>
<th>comment</th>
<th>sentence 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>but I am bound and thrown into a dark cell.</td>
<td></td>
<td>be fit to burn in a jail; no, not in a condemned cell.</td>
<td>But I never would return to my cold prison cell; what’s life without liberty?</td>
</tr>
<tr>
<td>but I am bound and thrown into a dark cell.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>but I am bound and thrown into a dark cell.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>be fit to burn in a jail; no, not in a condemned cell.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>be fit to burn in a jail; no, not in a condemned cell.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>She searched the bag for her cell as we headed toward the door.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>But I never would return to my cold prison cell; what’s life without liberty?</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Diachronic Data

Table 4: Annotation Table.
Word Usage Graphs (WUGs)

Figure 3: Graph visualization four uses of *cell*. 
Word Usage Graphs (WUGs)

Figure 4: Graph visualization four uses of cell.
Clustering

- correlation clustering (Bansal, Blum, & Chawla, 2004)
- optimization criterion: **reduce (weighted) number of cluster-edge conflicts**

\[ L(C) = \sum_{e \in \phi_{E,C}} W(e) + \sum_{e \in \psi_{E,C}} |W(e)| \]

(i) finds the optimal number of clusters on its own  
(ii) handles missing information (non-observed edges)  
(iii) robust to errors by using the global information  
(iv) respects the gradedness of word meaning  
(v) dominated in simulation study
Figure 5: Graph visualization for uses of *cell*.
Clustering

Figure 6: Graph visualization for uses of cell.
Figure 7: Usage graph of Swedish *ledning*. 
Figure 8: Usage graph of Swedish *ledning*.
Figure 9: Usage graph of Swedish *ledning*.
Figure 10: Usage graph of German *Eintagsfliege*.
Figure 11: Usage graph of German *Eintagsfliege*. 
Figure 12: Usage graph of German *Eintagsfliege*.
Senses in WUGs

Figure 13: Usage graph of German *Festspiel*.
Senses in WUGs

Figure 14: Usage graph of German *zersetzen*.
Senses in WUGs

Figure 15: Usage graph of German *Abgesang*. 
The theoretical perspective

- which is the best model for WUGs?
- what is the psychological status of WUGs?
Stochastic Block Model (SBM)

- generative probabilistic model for random graphs with planted clusters
- canonical model for community detection
- simplest model of a graph with communities (Abbe, 2017)
Stochastic Block Model (SBM)

Figure 16: SBM cycle
Senses in WUGs

Figure 17: Usage graph of German *Abgesang*. 
The practical perspective

- how can WUGs be exploited for lexicography?
- how can WUGs be annotated efficiently?
- how can we further automate the lexicographic process?
- DURel annotation interface:
  https://www.ims.uni-stuttgart.de/data/durel-tool
Example of Use Pair

Screenshot of visualized usage graph from DURel system’s beta version.
Figure 18: Screenshot of annotation interface from DUREl system’s beta version.
Sampling procedure

Figure 19: Step 0: No information
Sampling procedure

Figure 20: Step 1: Initial clustering
Sampling procedure

Figure 21: Step 2: Cluster comparison
Sampling procedure

Figure 22: Step 3: Compare non-assignable uses
Sampling procedure

Figure 23: Step 4: Cluster comparison
Sampling procedure

Figure 24: Step 5: Cluster comparison


Svenska Akademien. (2009). *Contemporary dictionary of the Swedish Academy.* The changed words are extracted from a database managed by the research group that develops the Contemporary dictionary.