Jean Mark Gawron
San Diego State University, Department of Linguistics
January 28, 2020
1. Introduction

2. Difficulties with senses
Outline

1. Introduction

2. Difficulties with senses
What is a lexicographer?

“. . . a lexicologist with a deadline” (Fillmore)
Dictionary entries

Three tasks

1. Identify senses by looking at data.
2. Compose definitions. (Many of you would get fired your first day on the job.)
3. Provide *illustrative* examples. (Not so easy!)
Outline

1. Introduction

2. Difficulties with senses
Identifying (and **individuating**) senses

1. How many senses are there? Look at the dictionary entries for *line* that follow, and compare with other dictionaries. Atkins and Levin (1991) find numerous cognitive and theoretical difficulties in merging *whistle, whistler* entries from two dictionaries.


3. Modern Lexicographers have moved away from relying on that classical notion of what a dictionary is (American Heritage, Collins, New Oxford) to entries heavily fortified with examples, proliferating senses whenever necessary to accommodate a variety of examples.
Verbs are worse (Atkins 1993, Palmer et al. 2007)

Hector Project Dictionary (1992-93), used in Senseval-1, seed corpus for BNC

<table>
<thead>
<tr>
<th>Verb</th>
<th>Number of senses</th>
</tr>
</thead>
<tbody>
<tr>
<td>bury</td>
<td>15</td>
</tr>
<tr>
<td>float</td>
<td>18</td>
</tr>
<tr>
<td>seize</td>
<td>11</td>
</tr>
</tbody>
</table>
### Example problem verbs (Palmer et al. 2007)

<table>
<thead>
<tr>
<th>Hector</th>
<th>WN</th>
</tr>
</thead>
<tbody>
<tr>
<td>shake</td>
<td></td>
</tr>
<tr>
<td>shake hands with so.</td>
<td>WN1</td>
</tr>
<tr>
<td>shake o.’s fist at so.</td>
<td>WN1</td>
</tr>
<tr>
<td>shake o’s head</td>
<td>WN1</td>
</tr>
<tr>
<td>My hands shook from the cold.</td>
<td>WN1</td>
</tr>
<tr>
<td>He shook the bag violently.</td>
<td>WN1</td>
</tr>
<tr>
<td>shaken by his father’s death</td>
<td>WN5</td>
</tr>
<tr>
<td>gentle tremors (WN2)</td>
<td></td>
</tr>
<tr>
<td>rapid vibrations (WN3)</td>
<td></td>
</tr>
<tr>
<td>swaying (WN4)</td>
<td></td>
</tr>
</tbody>
</table>

Sense grouping methodologies proposed (syntactic and semantic criteria); psychological priming experiments support the idea of core/peripheral senses. (Williams 1992)
Kilgariff (1997:19)

The primary implication is that a task-independent set of word senses for a language is not a coherent concept. Word senses are simply undefined unless there is some underlying rationale for clustering, some context which classifies some distinctions as worth making. For people, homonyms like *pike* are a limiting case; in almost every situation where a person considers it worth their while attending to a sentence containing *pike*, it is also worth their while making the fish/weapon distinction.
Coactivation: two senses at once

1. For better or worse, this would bring competition to the licensed trade. (Kilgarriff 1992)
   - competition = more competitors
   - competition = more competitiveness

2. On streets where life can be as hard, gray, and dirty as the sidewalks, people seem to want art. (Anastasia Hernandez)
   - impervious to pressure or impact (said of a surface)
   - fraught with difficulty, perilous
Word senses are simply undefined unless there is some *underlying rationale* for clustering, some context which classifies some distinctions as worth making. (emphasis mine) Kilgarriff (1997:19)

The importance of the concept of activation in disambiguation was demonstrated in experiments by Tanenhaus et al. (1979) and Swinney (1979) . . . word senses are not selected instantaneously in sentence comprehension. Rather both senses of a word with a two-way ambiguity are activated initially. The sense that receives contextual support remains active, whereas the the contextually inappropriate sense is gradually deactivated . . . Schuetze (2000:206)
(Statistical) word vectors for all words . . .

For each token of word \( w \) (say, \textit{cup}), average together the word vectors of the words nearby (words like \textit{drink}, \textit{win}, \textit{suction}, \textit{of coffee}) into a \textbf{context vector}.

You now have, say, 1500 context vectors. Cluster these into \( k \) “senses.”

Each “sense” is a group of similar context vectors.
How to do semantics if the foundations of meaning are so creaky?

We **do** understand each other (a lot of the time, anyways . . . )

Particular aspects of word meaning get activated/primed by context, not necessarily by logical associations (The astronomer married a star.) So we think of a word’s senses as groups of context appropriate for it.

We focus in this course on the semantics/pragmatics connection, grammatical meaning, and how word meanings **combine**
Bibliography

Atkins, Beryl T, and Beth Levin. 1991.
Admitting impediments.
*Lexical acquisition: Exploiting on-line resources to build a lexicon* 233.

Atkins, Sue. 1993.
Tools for computer-aided corpus lexicography: the hector project.

Cambridge, MA: MIT Press USA.

Dictionary and word sense distinctions: An inquiry into their nature.
   I don't believe in word senses.

   Introduction to the special issue on senseval.

Palmer, Martha, Hoa Trang Dang, and Christiane Fellbaum. 2007.
   Making fine-grained and coarse-grained sense distinctions, both
   manually and automatically.
   *Natural Language Engineering* 13(02):137–163.

   Disambiguation and connectionism.
   University Press Oxford.

   Lexical access during sentence comprehension: (re) consideration of
   context effects.