HG8003 Technologically Speaking: The intersection of language and technology.

Representing Meaning

Francis Bond
Division of Linguistics and Multilingual Studies
http://www3.ntu.edu.sg/home/fcbond/
bond@ieee.org

Lecture 3
Location: LT8

HG8003 (2014)
<table>
<thead>
<tr>
<th>Lec.</th>
<th>Date</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>01-16</td>
<td>Introduction, Organization: Overview of NLP; Main Issues</td>
</tr>
<tr>
<td>2</td>
<td>01-23</td>
<td>Representing Language</td>
</tr>
<tr>
<td>3</td>
<td>02-06</td>
<td>Representing Meaning</td>
</tr>
<tr>
<td>4</td>
<td>02-13</td>
<td>Words, Lexicons and Ontologies</td>
</tr>
<tr>
<td>5</td>
<td>02-20</td>
<td>Text Mining and Knowledge Acquisition</td>
</tr>
<tr>
<td>6</td>
<td>02-27</td>
<td>Structured Text and the Semantic Web</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Quiz</td>
</tr>
<tr>
<td>7</td>
<td>03-13</td>
<td>Citation, Reputation and PageRank</td>
</tr>
<tr>
<td>8</td>
<td>03-20</td>
<td>Introduction to MT, Empirical NLP</td>
</tr>
<tr>
<td>9</td>
<td>03-27</td>
<td>Analysis, Tagging, Parsing and Generation</td>
</tr>
<tr>
<td>10</td>
<td>Video</td>
<td>Statistical and Example-based MT</td>
</tr>
<tr>
<td>11</td>
<td>04-03</td>
<td>Transfer and Word Sense Disambiguation</td>
</tr>
<tr>
<td>12</td>
<td>04-10</td>
<td>Review and Conclusions</td>
</tr>
<tr>
<td>Exam</td>
<td>05-06</td>
<td>17:00</td>
</tr>
</tbody>
</table>

➢ Video week 10
Overview

➢ Review of representing text and speech

➢ Word Meaning: Lexical Semantics
  ➢ Why do we want to represent meaning
  ➢ Various approaches (linguistic and computational)
    • Attributional Meaning
    • Relational Meaning
    • Distributional Meaning

➢ Meaning and Usage
Revision
Revision of Representing Language

➢ Writing Systems

➢ Encodings

➢ Speech

➢ Bandwidth
Three Major Writing Systems

➢ Alphabetic (Latin)
  ➢ one symbol for consonant or vowel
  ➢ Typically 20-30 base symbols (1 byte)

➢ Syllabic (Hiragana)
  ➢ one symbol for each syllable (consonant+vowel)
  ➢ Typically 50-100 base symbols (1-2 bytes)

➢ Logographic (Hanzi)
  ➢ pictographs, ideographs, sounds-meaning combinations
  ➢ Typically 10,0000+ symbols (2-3 bytes)
    (2 bytes for currently used, 3 bytes for all variants in all languages)
Encoding

➤ Need to map characters to bits

➤ More characters require more space

➤ Moving towards unicode for everything

➤ If you get the encoding wrong, it is gibberish
Speech

- Speech is an analog signal
  - considerable variation
  - no clear boundaries

- Hard to convert to symbols
  - single speaker trained models work OK
  - noisy speech is still an unsolved problem
Speed is different for different modalities

Speed in words per minute (one word is 6 characters)
(English, computer science students, various studies)

<table>
<thead>
<tr>
<th>Activity</th>
<th>Reading</th>
<th>Writing</th>
<th>Speaking</th>
<th>Hearing</th>
<th>Typing</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>300</td>
<td>31</td>
<td>150</td>
<td>150</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>200 (proof reading)</td>
<td>21 (composing)</td>
<td></td>
<td>210 (speeded up)</td>
<td>19 (composing)</td>
</tr>
</tbody>
</table>

➢ Reading >> Speaking/Hearing >> Typing

⇒ Speech for input
⇒ Text for output
Meaning
Why meaning?

➢ When we read (or listen) what we understand is more than is actually said or written

➢ Understanding is useful for

➢ Translation
➢ Search
➢ Inference
➢ Paraphrasing
➢ …
The Core Problem of MT (& NLU)

(1) 頭を掻いた
atama wo kaita
head ACC scratched
“I scratched my head.”

➢ The Japanese text doesn’t say

1. That 掫く should be scratch, not shovel, row, . . .
2. Who scratched; Whose head it is
3. That 頭 should be head, not boss, top, . . .
4. That head needs a possessive pronoun

➢ A native speaker of Japanese would know (2), could deduce (1,3)
➢ A native speaker of English knows (4)
➢ How do we teach a computer?
Different Strokes for Different Folks

➤ Most languages care about possession

➤ English: pronouns
  my head

➤ Japanese: politeness, evidentiality
  your honorable head vs my head
  I itch vs you seem to itch

➤ Russian: reflexives
  I scratch self head

➤ Swedish: definiteness
  I scratch the head (head-et)

➤ Shared level somewhere beyond syntax
  This is the level that we call meaning or semantics
(2) 頭を掻いた
atama wo kaita
head ACC scratched
“I scratched my head.”

Utterance

Syntax

Lexical & Structural Semantics
Introduction to Lexical Semantics

➢ Attributional Meaning
define meaning through attributes (definitions, semantic primitives)

➢ Relational Meaning
define meaning through relations (semantic graphs)

➢ Distributional Meaning
define meaning as points in semantic space

➢ the Syntax-Semantics Interface
  ➢ Verb Diathesis
  ➢ Countability
What is Lexical Semantics?

➤ Working definition:

the study of what individual lexical items mean, why they mean what they do, how we can represent all of this, and where the combined interpretation for an utterance comes from
Predict the morphosyntax (esp. countability) of:

coagulopathy:

muntjac:

Countability (syntactic property of English)

Countable has singular and plural, takes “a”: a dog, dogs
typically things

Uncountable singular only, no “a”: gold, *golds
typically stuff
Example of Lexical Semantics in Action (1)

➢ Predict the morphosyntax (esp. countability) of:

➢ **coagulopathy**: group of conditions of the blood clotting (coagulation) system in which bleeding is prolonged and excessive; a bleeding disorder

➢ **muntjac**: small Asian deer with small antlers and a cry like a bark

➢ What part of speech are these? (noun, verb, adjective)

➢ Are they countable or uncountable?
Interpret the following compound nominalisations:

- risk recognition
- doctor involvement

Compare:

- police failure $\equiv$ (the) police$_{\text{SUBJ}}$ fail
- player selection $\equiv$ [someone] selects (the) player$_{\text{OBJ}}$
Lexical semantics is concerned with the identification and representation of the semantics of lexical items.

If we are to identify the semantics of lexical items, we have to be prepared for the eventuality of a given word having multiple interpretations.

- **polysemy**: having multiple meanings
- **mono-semy**: having only one meaning
Distinguishing Polysemes

➢ The polysemy of a word can be tested by a variety of means, including:

➢ antagonism: can the word be used in a sentence with multiple competing interpretations?

* Kim can't bear children
  * Cannot have children
  * Doesn’t like children

➢ zeugma: when the word is used in a context where multiple competing interpretations are simultaneously evoked, does it become a pun?

* Kim and her visa expired
  * died
  * ran out

You are free to execute your laws, and your citizens, as you see fit.
(From the television program Star Trek: The Next Generation)
➢ independent truth conditions: can the word be used in a given sentence with different truth conditions according for different interpretations?

* Kim is wearing a light jacket
* not heavy
* not dark

➢ definitional distinctness: it is impossible to come up with a unified definition which encompasses the different sub-usages of the word.

➢ Note the importance of actual examples in deciding about polysemy
Approaches to Defining Word Meaning

➢ Attributional semantic categorisation

➢ Relational semantic categorisation

➢ Distributional semantic categorisation
For each lexical item, come up with a semantic description of each of its distinct usages, in isolation of the categorisation of other lexical items, e.g.:

**enrichment** *(n)* the act of making fuller or more meaningful or rewarding

Methodologies:

- definitional semantics
- decompositional semantics
Definitional Semantics

➤ Standard lexicographic approach to lexical semantics:

*semantics* = the study of language meaning
*tailor* = a person whose occupation is making and altering garments

➤ Definitions are conventionally made up of;

➤ *genus*: what class the lexical item belongs to
➤ *differentiae*: what attributes distinguish it from other members of that class

➤ Often hard to understand if you don’t already know the meaning!
Definitional Semantics: pros and cons

➢ Pros:

➢ familiarity (look-up and annotation)

➢ Cons:

➢ subjectivity in sense granularity (splitters vs. lumpers) and definition specificity
➢ circularity in definitions
  * lynx: a bobcat; bobcat: a kind of lynx
  * Monday: the day after Sunday; . . .
➢ consistency, reproducibility, . . .
➢ often focus on diachronic (historical) rather than synchronic (current) semantics
Bear (v) in WordNet

1. bear – (have; "bear a resemblance"; "bear a signature")
2. give birth, deliver, bear, birth, have – (cause to be born; "My wife had twins yesterday!")
3. digest, endure, stick out, stomach, bear, stand, tolerate, support, brook, abide, suffer, put up – (put up with something or somebody unpleasant; "I cannot bear his constant criticism")
4. bear – (move while holding up or supporting; "Bear gifts"; "bear a heavy load"; "bear news"; "bearing orders")
5. bear, turn out – (bring forth, "The apple tree bore delicious apples this year"; "The unidentified plant bore gorgeous flowers")
6. bear, take over, accept, assume – (take on as one’s own the expenses or debts of another person; "I’ll accept the charges"; "She agreed to bear the responsibility")
7. hold, bear, carry, contain – (contain or hold; have within; "The jar carries wine"; "The canteen holds fresh water"; "This can contains water")
8. yield, pay, bear – (bring in; "interest-bearing accounts"; "How much does this savings certificate pay annually?")
9–13
The Corpus Revolution in Definitional Semantics

- Moves towards corpus-based lexicography in an attempt to reduce subjectivity in sense granularity and definition specificity
  - move from type- to token-based sense discrimination/annotation

- Look at many examples of a word in context

- Started with COBUILD in the 1970s

- Now fairly standard
normally organise everything and BEAR the costs of running and advertising o that we may take it thence and BEAR it to the chapel. HAMLET: Do not believe her with right sides facing, and BEAR in mind that one curtain as laid out their religious convictions and BEAR witness to the power of faith to solve at the age of two and a half and BEAR the first young when they are three. rued interests on such awards as BEAR interest, certified pursuant to section crued interest on such awards as BEAR interest. _(D)_ The Secretary of the unpaid principal of such awards BEAR to the total amount in the fund avail the overall curtain length, but BEAR in mind individual window shapes. Val ). However, other strategies can BEAR fruit and are described under three sit one more day is more than I can BEAR -- Love is already turning into hate. picture; but consider, if you can BEAR it, what might have happened if MacAr than the physical and mental, can BEAR overstraining. And, in the last case, materials for the shell will cost. BEAR in mind that this does not include in quate services, that these costs BEAR disproportionately on the rural poor. beehive voices, for no one could BEAR silence, drowned out the sound of Mrs probably hated more than he could BEAR? And possessed himself- how?- of a ri
Bear separated into senses

normally organise everything and BEAR the costs of running and advertising so that we may take it thence and BEAR it to the chapel. HAMLET: Do not believe rued interests on such awards as BEAR interest, certified pursuant to section unpaid principal of such awards BEAR to the total amount in the fund available her with right sides facing, and BEAR in mind that one curtain as laid out the overall curtain length, but BEAR in mind individual window shapes. Val their religious convictions and BEAR witness to the power of faith to solve at the age of two and a half and BEAR the first young when they are three. ). However, other strategies can BEAR fruit and are described under three s

one more day is more than I can BEAR -- Love is already turning into hate. icture; but consider, if you can BEAR it, what might have happened if MacAr
han the physical and mental, can BEAR overstraining. And, in the last case, bEEhive voices, for no one could BEAR silence, drowned out the sound of Mrs
Decompositional Semantics

Define in terms of primitives:

Bachelor: MARRIED −, MALE +

Hard to define the primitives

Define words by way of a constrained representation language, in an attempt to avoid circularity and enforce consistency of annotation, e.g. Lexical Conceptual Semantics (LCS):

`(DEF_WORD "give"
 LCS (cause (* thing 1)
 (go poss (* thing 2)
  ((* to 5) poss (thing 2) (at poss (thing 2) (thing 6))))
 (give+ingly 26)))

Jackendoff (1990)
Decompositional Semantics: Pros and Cons

➢ Pros:
   ➢ systematic representation/in-built definition of well-formedness
   ➢ language independent, consistent descriptions

➢ Cons:
   ➢ obscurity of representation
   ➢ disagreement about primitives/semantic language
     * number grows over time
     * 14 → 40 → 60
   ➢ subtle semantic distinctions can be impossible to make due to restrictions in the representation language
   ➢ it is hard to go from the definition back to the word
Relational
Relational Semantic Categorisation

➤ Capture correspondences between lexical items by way of a finite set of pre-defined semantic relations

➤ Methodologies:
  ➤ lexical relations
  ➤ constructional relations
Propositional synonymy: X is a propositional synonym of Y if

(i) X and Y are syntactically identical,
(ii) substitution of Y for X in a declarative sentence doesn’t change its truth conditions

e.g., violin and fiddle

Why propositional synonymy is over-restrictive:

- syntactic identity (cf. eat and devour)
- collocations (cf. cemetery and graveyard)
- gradability (cf. sofa/settee vs. boundary/frontier)
Near Synonymy

➤ Synonyms are substitutable in some/most rather than all contexts.

➤ Synonymy via semantics: synonyms share “common traits” or attributional overlap, walking the fine line between “necessary resemblances” and “permissible differences”:

\[ \text{grain vs. granule; green vs. purple; alsatian vs. spaniel} \]

➤ Permissible differentiation via clarification:

\[ \text{Here is a grain, or granule, of the substance.} \]
\[ * \text{The cover is green, or, that is to say} \, \text{purple.} \]

and contrast:

\[ \text{Here is a grain or, more exactly, granule} \]
\[ * \text{He likes alsations, or more exactly, spaniels} \]
Properties of synonymy

- Symmetric
- applies only to lexical items of the same word class
- applied at the sense or lexical item-level?
- \( \approx \) converse of polysemy
Hypernymy and Hyponymy

- **Hyponymy**: $X$ is a hyponym of $Y$ iff $f(X)$ entails $f(Y)$ but $f(Y)$ does not entail $f(X)$:

  - *Kim has a pet dog* $\rightarrow$ *Kim has a pet animal*
  - *Kim has a pet animal* $\not\rightarrow$ *Kim has a pet dog*

  N.B. complications with universal quantifiers and negation:

  - *Kim likes all animals* $\rightarrow$ *Kim likes all dogs*
  - *Kim likes all dogs* $\not\rightarrow$ *Kim likes all animals*

- **Hypernymy**: $Y$ is a hypernym of $X$ iff $X$ is a hyponym of $Y$
Properties of hypernymy/hyponymy

➢ Asymmetric

➢ applies only to lexical items of the same word class

➢ applies at the sense level

➢ Transitive
  
  \( \text{dog} \subseteq \text{mammal} \subseteq \text{animal} \)
Antonymy (opposites)

➢ **Complementarity:** X and Y are complementaries if X and Y define mutually-exclusive sets which encompass all of a conceptual domain, cf.:

   *The door is neither open nor shut*
   
   *I am feeling neither good nor bad*

➢ **Antonyms:**

  ➢ are fully gradable
  ➢ when intensified move in opposite directions along their scale of domain (cf. *heavy* vs. *light*)
  ➢ do not bisect their domain of operation

➢ Similarity with synonymy, in terms of attributional overlap
Antonymy is generally considered to operate at the lexical item-level (cf. *rise/fall* vs. *ascend/descend*)

Morphological influences (cf. *long/short* vs. *lengthen/shorten*)

Other properties of antonymy:

- symmetric
- applies only to lexical items of the same word class (esp. adjectives and verbs)
There are many, many more lexical relations advocated by various theories including:

- meronymy/holonymy (part-whole)
- troponymy/hypernymy (cf. *walk* vs. *lollop*)
- entailment (cf. *snore* vs. *sleep*)
- Element/Group (cf. *bee* vs. *swarm*)
- Operator (cf. *question* vs. *ask*)
- Magnifier (cf. *wound* vs. *badly*)
Word Meaning as a Graph

➢ You need a very big graph to capture all meanings
Distributional
Firth (1957) famously made the observation:

*You shall know a word by the company it keeps*

which is commonly known as the **distributional hypothesis**

Look at the **contexts** in which words appear
A Case in Point

Acyclovir is a specifically anti-viral drug ...
Acyclovir has been developed and marketed by ...
Acyclovir given intravenously, ...

Coagulopathy is a well recognised complication ...
... could stimulate a coagulopathy ...
... is also probably responsible for a coagulopathy ...
... a patient with a coagulopathy.
Lexico-syntactic context is commonly used by corpus linguists to analyse lexical semantics, through a combination of:

- concordancing
- analysis of common verb–argument collocations
- analysis of passives and other constructions
- analysis of co-occurrence with certain adverbs/auxiliaries
Do you know what a blag is?

➢ *The blag bit the postman.*

➢ *The big hairy blag . . .*

➢ *He was walking his blag.*

➢ *The blag barked.*

Now do you know what a blag is?
We can learn word meaning from context

➢ all we needed to learn *blag* was the context
  ➢ there was no grounding or definition
  ➢ no real world example, photograph or other representation

➢ all of this was learnt from seeing it in context
Distributional Hypothesis

- Similar terms appear in similar contexts

- Distributional Similarity $\neq$ Co-occurrence
  - Distributional similarity requires shared context
  - The terms themselves don’t have to appear together
  - i.e. distributionally similar terms need not co-occur
  - this is important since synonyms don’t always co-occur
What are terms?

- Similar terms appear in similar contexts
- single words
- multi-word expressions
  - noun compounds: *machine translation*
  - verb particle: *give up*
What are contexts?

- Some way of defining a semantic space
- E.g., Define each word as a vector of attributes (1, 2, 0, 3, 5, 1, ...)
- Similarity is defined as being close in this space
- Also known as: Latent Semantic Indexing (LSI)
Consider these two sets

c1: Human machine interface for Lab ABC computer applications

c2: A survey of user opinion of computer system response time

c3: The EPS user interface management system

c4: System and human system engineering testing of EPS

m1: The generation of random, binary, unordered trees

m2: The intersection graph of paths in trees

m3: Graph minors IV: Widths of trees and well-quasi-ordering

m4: Graph minors: A survey

<table>
<thead>
<tr>
<th>Term</th>
<th>c1</th>
<th>c2</th>
<th>c3</th>
<th>c4</th>
<th>m1</th>
<th>m2</th>
<th>m3</th>
<th>m4</th>
</tr>
</thead>
<tbody>
<tr>
<td>human</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>computer</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>user</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>system</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>survey</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>trees</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>graph</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>minors</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
Contexts: Locality

- $n$-words window based co-occurrence
  
  He$_7$ said$_6$ that$_5$ he$_4$ had$_3$ a$_2$ good$_1$ idea about$_1$ that$_2$

- terms within the window are the attributes
- window sizes vary ($\pm 1 - \pm 250$)
- phrase, sentence and document boundaries

- Sentence and document level co-occurrence
  
  Sentence/Documents themselves are attributes

  - this is typical in Information Retrieval
  - IDs are the attributes
  - called a term-document matrix
Contexts: Linguistic Structure

- $n$-word window contexts may include relative position
- filtering on stop words or part of speech (pos) tags
- but these constrain syntactic class of synonyms
- Grammatical relations
  - Verb-Subject, Verb-Object, Verb-Indirect Object: (OBJ have idea)
  - Modifier-Head: (MOD good idea)
- Extracting linguistic structure affects precision:
  - grammatical relations are more correlated (higher precision)
  - parsing errors introduce noise (lower precision)
Stop Words

Words which are filtered out prior to, or after, processing of natural language data (text).

There is no definite list of stop words which all NLP tools incorporate.

Typical examples are function words:

- a, the, this, that
- of, in, on, at
- you, he, who

Stop words can cause problems when using a search engine to search for phrases that include them, particularly in names such as 'The Who', 'The The', or 'Take That'.
**Evaluation Metrics**

**Precision**  Ratio of correctly labeled/Labeled (P: Accuracy)

**Recall**  Ratio of correctly labeled/Should have been labeled (R)

Normally we can raise precision at the cost of lower recall and vice-versa. So we try to optimize a combined score: **F-measure**

**F-measure**  A measure of overall goodness \( \frac{2PR}{P+R} \) (F)

More generally F-measure is \( \frac{(1+\beta^2)PR}{\beta^2P+R} \).

Most often we set \( \beta = 1 \). If Precision is more important, increase \( \beta \).
Another way of looking at it

<table>
<thead>
<tr>
<th>System</th>
<th>Actual</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>target</td>
</tr>
<tr>
<td>selected</td>
<td>tp</td>
</tr>
<tr>
<td>not selected</td>
<td>fn</td>
</tr>
</tbody>
</table>

Precision = \( \frac{tp}{tp + fp} \); Recall = \( \frac{tp}{tp + fn} \); \( F_1 = \frac{2PR}{P + R} \)

**tp** True positives: system says Yes, target was Yes

**fp** False positives: system says Yes, target was No

**tn** True negatives: system says No, target was No

**fn** False negatives: system says No, target was Yes
Example: Similarity

➢ System says *eggplant* is similar to *brinjal*
True positive

➢ System says *eggplant* is similar to *egg*
depends on the application (both food), but generally not so good
False positive

➢ System says *eggplant* is not similar to *aubergine*
False negative

➢ System says *eggplant* is not similar to *laptop*
True negative
Context — Size Trade-offs

- There are precision/recall trade-offs with locality:
  - Larger contexts cause more collisions (higher recall)
  - Larger contexts are less correlated (lower precision)

- Large contexts require lots of storage

- Speed is an important factor
  - Window methods are extremely fast (minutes)
  - Linguistic methods can be much much slower (hours to days)
    but they produce much better quality context information

- More data can trump better quality
  - Given (near) unlimited raw text, speed is very important
Similarity Measures

➤ Once you have context vectors

➤ You need to compare them

➤ Many, many possible measures
  ➤ Based on distance between points
  ➤ Based on importance of attributes

➤ Typical ones: Cosine, Jacard, Mutual Information

➤ Basically — closer is more similar
## Similarity

<table>
<thead>
<tr>
<th>Term</th>
<th>c1</th>
<th>c2</th>
<th>c3</th>
<th>c4</th>
<th>m1</th>
<th>m2</th>
<th>m3</th>
<th>m4</th>
</tr>
</thead>
<tbody>
<tr>
<td>human</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>computer</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>user</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>system</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>survey</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>trees</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>graph</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>minors</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

\[
Jacard(A, B) = \frac{|A \cap B|}{|A \cup B|}
\]

Jacard(c2,c3) = \(2/3 = 0.67\)
Jacard(c2,m2) = \(0/5 = 0\)
Jacard(m2,m4) = \(1/4 = 0.25\)
Distributional Similarity often Surprises

➢ Other relations are often identified

➢ other related concepts with similar contextual distributions
➢ this is called the tennis problem:
  \( \text{ball, racquet, net, . . .} \)

➢ the worst case is antonyms
  (\text{hot} and \text{cold} share similar contexts)

➢ less problematic is hypernyms/hyponyms
  (particularly for symmetric measures)
Summary: three ways to define meanings

➢ Attributional Meaning
  meaning given by attributes

➢ Relational Meaning
  meaning through relations (semantic graphs)

➢ Distributional Meaning
  points in semantic space
How does syntax effect meaning?

How does semantics effect the way we use words?
Example: Diathesis Alternations

➢ Causative/inchoative alternation:

\[ \text{Kim broke the window} \leftrightarrow \text{The window broke} \]

➢ Middle construction alternation:

\[ \text{Kim cut the bread} \leftrightarrow \text{The bread cut easily} \]\n
Levin (1993)
Conative alternation:

\[ Kim \textit{hit} \text{ the door} \leftrightarrow Kim \textit{hit} \text{ at the door} \]

Body-part possessor ascension alternation:

\[ Kim \textit{cut} \text{ Sandy’s arm} \leftrightarrow Kim \textit{cut} \text{ Sandy on the arm} \]
A verb’s (in)compatibility with different alternations is a strong predictor of its lexical semantics:

<table>
<thead>
<tr>
<th></th>
<th>break</th>
<th>cut</th>
<th>hit</th>
<th>touch</th>
</tr>
</thead>
<tbody>
<tr>
<td>Causative</td>
<td>YES</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
</tr>
<tr>
<td>Middle</td>
<td>YES</td>
<td>YES</td>
<td>NO</td>
<td>NO</td>
</tr>
<tr>
<td>Conative</td>
<td>NO</td>
<td>YES</td>
<td>YES</td>
<td>NO</td>
</tr>
<tr>
<td>Body-part</td>
<td>NO</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
</tbody>
</table>

\[
\text{break} = \{ \text{break, chip, crack, crash, crush, ...} \}
\]
\[
\text{cut} = \{ \text{chip, clip, cut, hack, hew, saw, ...} \}
\]
\[
\text{hit} = \{ \text{bang, bash, batter, beat, bump, ...} \}
\]
\[
\text{touch} = \{ \text{caress, graze, kiss, lick, nudge, ...} \}
\]

Levin (1993)
Corollary: we can predict the syntax of novel words we are given the semantic class for (cf. countability examples earlier)

The principal weakness of syntax-based verb classification is that there are often subtle divergences in semantics between synonyms (cf. *eat* vs. *devour* vs. *gobble*)
Countability and the Syntax-Semantics Interface

Countability:

- A syntactico-semantic property of the noun phrase
- Bounded, indivisible *individuals*
  prototypically COUNTABLE: *a dog, two dogs*
- Unbounded, divisible *substances*
  prototypically UNCOUNTABLE: *gold*
Divisibility and Countability

VS.
Countability Classes

➤ countable: book, button, person (one book, two books)

➤ uncountable: equipment, gold, wood (*one equipment, much equipment, *two equipments)

➤ plural only: clothes, manners, outskirts (*one clothes, clothes horse)

➤ bipartite: glasses, scissors, trousers (*one scissors, scissor kick, pair of scissors)
Countability and the Syntax-Semantics Interface

- Semantic properties of a given noun are strong predictors of both its countability (lexical semantics) and surface manifestation (syntax):
  - (simple) enumerable $\leftrightarrow$ countable
  - usable as bare singular NP $\leftrightarrow$ uncountable

- I.e., syntax offers a powerful semantic validation tool
Differences in Conceptualisation

Knowing the referent is not enough to determine countability, e.g. *scales*

1. Thought of as being made of two arms: *(British)*
   *a pair of scales*
2. Thought of as a set of numbers: *(Australian)*
   *a set of scales*
3. Thought of as discrete whole objects: *(American)*
   *one scale/two scales*

Also *Lego* – countable or uncountable?
Looking at corpus data to determine countability leads to its own challenges, e.g. *enrichment*

*Education itself provides enrichment to ...*
*... would bestow great enrichment upon ...*
*Job enrichment is part of ...*

*It was a developmental enrichment.*
*... an enrichment of life.*
*... received many enrichments ...*
Basic vs. Derived Uses

- Countability categorisation is confused by the existence of highly-productive conversion rules, e.g.:

  the Universal Grinder: countable noun with individuated semantics $\rightarrow$ uncountable noun with “piecemeal” semantics (e.g. *the floor was littered with computer*)

  the Universal Packager: uncountable noun with substance semantics $\rightarrow$ countable noun with portion of substance semantics (e.g. *two beers*)

- Rather than consider all nouns as both countable and uncountable, we generally identify the “basic” uses of a given noun and derive alternate uses through the use of lexical rules (but consider *chicken* vs. *dog* vs. *worm*)

- Cf. regular/logical polysemy

Copestake and Briscoe (1995); Jackendoff (1991); Pustejovsky (1995)
There is growing awareness in lexical semantics of:

- context-sensitivity
- sense specificity
- basic vs. derived word usages (and the fuzziness of the boundary)
- difficulties in making categorical judgements for a word
Acknowledgments

➢ Slides incorporate material from Tim Baldwin and James Curran.

➢ No dogs were harmed in the making of these slides.
Bibliography
References


---

Representing Meaning