Lexical information and NL applications

- NL applications often need to know the meaning of words at least.
- Wrong meaning is tricky, messy stuff!
- IBM Watson: “Words by themselves have no meaning.”
- Many word strings express apparently unrelated senses / meanings, even after their POS has been determined.
- Well-known examples: BANK, STOCK, RIGHT, SET, STOCK.
- Homonymy affects the results of applications such as IR, and machine translation.
- The opposite case of different words with the same meaning (SYNONYMY) is also important.
- NOT BOOK/LAPTOP
  - E.g., for IR systems (synonym expansion)

Homonymy, homography, homophony

- HOMONYMY: Words strings like STOCK are used to express apparently unrelated senses / meanings, even in contexts in which their part-of-speech has been determined.
- Other well-known examples: BANK, RIGHT, SET, SCALE.
- HOMOPHONES: BASS
  - The expert angle from Dona. No was fly-casting or BASS rather than the traditional food.
  - The question is the sound of angry dogs baying and amorous BASS chords sounding.
  - Problems caused by homophony: text to speech synthesis.
- Many spelling errors are caused by HOMOPHONES – distinct lexemes with a single pronunciation.
  - Knit vs. knit
  - Weather vs. whether
  - Their vs. there

Polysemy vs Homonymy

- In cases like BANK, it’s fairly easy to identify two distinct senses (etymology also different). But in other cases, distinctions more questionable.
- E.g., senses 0.00 and 0.00 of stock clearly related, like 0.00 and 0.00, or 0.00 and 1.00.
- POLYSYNONYMOUS WORDS: meanings are related to each other.
- Commonly the result of some kind of metaphorical extension.
- In some cases, syntactic tests may help.
- Game can mean, e.g., sports, etc. over polysemy not homonymy
  - In general, distinction between HOMONYMY and POLYSYNONYMY not always easy.

Meaning in MRs, 2: SYNONYMY

- Two words are SYNONYMS if they have the same meaning at least in some contexts.
- E.g., PRICE and RARE; CHEAP and INEXPENSIVE; LAPTOP and NOTEBOOK; HOME and HOUSE.
- I’m looking for a CHEAP FLIGHT (INEXPENSIVE) FLIGHT.
- From Roger’s thesaurus.
- ORTHOGRAHY: ensure, cancellation, deletion.
- But very few words are truly synonymous in ALL contexts.
- HOME/HOUSE is where the heart is.
- The flight was CANCELLED/YY/DELETED.
- Knowing about synonyms may help in IR.
- NOT BOOK vs. LAPTOP, as well.
- CHEAP PRICE vs. INEXPENSIVE PRICE.
Hyponymy and Hypernymy

- **Hyponymy** is the relation between a subclass and a superclass:
  - Car and Vehicle
  - Dog and Animal
  - RUGGED and HUNGRY
- Generally speaking, a hypernymy relation holds between X and Y whenever it is possible to substitute Y for X.
  - That is a X.
  - That is a Y.
- **Hypernymy** is the opposite relation.

**Knowledge about TAXONOMIES useful to classify web pages**
- By Semantic Web: ISA relation of Y
- This information not generally contained explicitly in a traditional or machine-readable dictionary (MAD)

The organization of the lexicon: Synonymy

- Word-Strings
- Lexemes
- Senses

A free, online, more advanced lexical resource: WordNet

- **WordNet**
  - A lexical database created at Princeton
  - Freely available for research from the Princeton site
  - http://wordnet.princeton.edu
  - Information about a variety of SEMANTICAL RELATIONS
  - Three sub-databases supported by psychological research as early as (Fillenbaum and Jones, 1943)
  - 20K nouns
  - 30K verbs
  - 30K adjectives and adverbs
  - Each entry consists of a broad-taxonomic-synset
  - Each synset consists of any number of lexicalized concepts

The noun database

- About 90,000 forms, 116,000 senses
- Relations:
<table>
<thead>
<tr>
<th>Hypernym</th>
<th>Hyponym</th>
<th>HasMember</th>
<th>MemberOf</th>
<th>PartOf</th>
<th>WholeOf</th>
<th>Antonym</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breakfast -&gt; meal</td>
<td>meal -&gt; lunch</td>
<td>Faculty -&gt; professor</td>
<td>Professor -&gt; faculty</td>
<td>Course -&gt; meal</td>
<td>Meal -&gt; dinner</td>
<td>Owner -&gt; follower</td>
</tr>
</tbody>
</table>

Synsets

- **Senses** (or "lexicalized concepts") are represented in WordNet by the set of words that can be used in AT LEAST ONE CONTEXT to express that sense / lexicalized concept
  - the SNIF
  - E.g.,
    - (chump, fool, fool, gull, mark, mark, fall guy, sucker, shenmel, soft touch, mug)
    - (gloss: person who is gullible and easy to take advantage of)
Hyperonyms

<table>
<thead>
<tr>
<th>Synonym</th>
<th>Antonym</th>
</tr>
</thead>
<tbody>
<tr>
<td>robe</td>
<td></td>
</tr>
</tbody>
</table>
Identifying the sense of a word in its context

Bar-Hillel

“Let me state rather dogmatically that there exists at this moment no method of reducing the ambiguity of a short, twenty-word Russian sentence in a scientific article below a remainder of 1; I would estimate, at least five or six words, with many possible English renderings, which would not seriously endanger the quality of the machine output. Many tend to believe that by reducing the number of initially possible renderings of a twenty-word Russian sentence from a few tens of thousands, to 1, and the assumption that each of the twenty Russian words has two renderings on the average, while seven or eight of them have only one rendering, to some eighty which would be the number of renderings on the assumption that all ten words are uniquely rendered, you have three renderings, namely, forgetting now about all the other aspects, such as change of word order, etc. The main bulk of this kind of work has been achieved, the remainder requiring only some slight additional effort.” (Bar-Hillel, 1950, p. 163).

Identifying the sense of a word in its context

Philosophy

“You shall know a word by the company it keeps.”
—J. R. R. Tolkien (1937)

“You say: the point isn’t the word, but its meaning, and you think of the meaning as a thing of the same kind as the word, though also different from the word. Here the word, then the meaning, The money, and the cow that you can buy with it. (But contrast: money, and its use.)”
—Wittgenstein, Philosophical Investigations

“You might, in a great many cases, though not for all—in which we employ the word “meaning” it can be defined thus: the meaning of a word is its use in the language.”
—Wittgenstein, Philosophical Investigations

U.S. Supreme Court

“[T]he fundamental principle of statutory construction [and, indeed, of language itself] that the meaning of a word cannot be determined in isolation, but must be drawn from the context in which it is used.”
Corpora used for word sense disambiguation work

- **Sense Annotated** (Difficult and expensive to build)
  - Semcor (200,000 words from Brown)
    - 362,000 semantically annotated occurrences of 121 nouns and 70 verbs.
    - Classic words: interest, link, ... (Training data for Senseval competitions (lexical samples and running text))
- **Non Annotated** (Available in large quantity)
  - Brown, newspaper, Web

Dictionary-based approaches

- **Latek (1986):**
  1. Retrieve from WordNet all sense definitions of the word to be disambiguated
  2. Compare with sense definitions of words in context
  3. Choose sense with most overlap

  **Example:**
  - PINE
    - 1 kinds of evergreen tree with needle-shaped leaves
    - 2 waste away through center of bricks
    - 3 solid body which rotates to a point
    - 4 something of this shape whether solid or hollow
  - Disambiguate: PINE CONE

Frequency-based word sense disambiguation

- If you have a corpus in which each word is annotated with its sense, you can collect class-based unigram statistics (count the number of times each sense occurs in the corpus)
  - **POS(NEW)**
    - **POS(NEW) / POS(OLD)**
  - E.g., if you have
    - 5465 uses of the word bridge
    - 5461 cases in which it is tagged with the sense STRUCTURE
    - 14 instances with the sense ORTAL-DEVICE

  - Frequency-based WSJ can get about 60-70% correct!
  - To improve upon these results, need context

Traditional selectional restrictions

- One type of contextual information is the information about the type of arguments that a verb takes — its *selectional restriction.*
  - AGENT (A Agent FOOD-TO-INF A)
  - AGENT (A Agent DRINK-TO-inf A)

  **Examples:**
  - Which airline serves DENVER
  - Which airline serves BOULDER

  **Limitations:**
  - In 1975, championship battle, Mr. Hubert ATE GLASS on an empty stomach, accompanied only by water and tea.
  - But it fell apart in 1981, perhaps because people realized that you can’t EAT GLASS for lunch if you’re hungry.

  Resnik (1998): 44% with these methods.

Context in general

- But it’s not just classic selectional restrictions that are useful context.
  - Often simply knowing the topic is really useful.
**Supervised approaches to WSD: the rebirth of Naïve Bayes in CompLing**

- A Naïve Bayes Classifier chooses the most probable sense for a word given the context:
  \[ s = \arg \max_s P(s|C) \]
- As usual, this can be expressed as:
  \[ P(C|s_i)P(s_i) \]
- The "NAÏVE BAYES ASSUMPTION: all the features are independent:
  \[ P(C|s_i) = \prod_i P(C|s_i) \]

**Gale et al: words as contextual clues**

- Gale et al. view a context as a substring of words of some length (how long?) at some distance away (how far?)
- Good clues for different senses of DRUG:
  - Medication, poison, prescription, patient, receive, consumer
  - Legal substance, abuse, paraphernalia, illicit, cocaine, trafficking
  - To determine which interpretation is more likely, extract words (e.g., ABUSES) from context, and use P(ABUSES|medication), P(ABUSES|drogue) estimated as smoothed relative frequency
  - Gale et al. (1992) disambiguation system using this algorithm correct for about 90% of occurrences of ambiguous nouns in the Hansard corpus
  - July, drug, land, language, position, sentence

**An example of use of Naïve Bayes classifiers: Gale et al (1992)**

- Used this method to disambiguate word senses using an ALIGNED CORPUS (Hansard) to get the word senses

<table>
<thead>
<tr>
<th>English</th>
<th>French</th>
<th>Sense</th>
<th>Number of examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>July</td>
<td>calendar</td>
<td>tax obligation</td>
<td>687</td>
</tr>
<tr>
<td>drug</td>
<td>medicament</td>
<td>medical</td>
<td>2292</td>
</tr>
<tr>
<td>land</td>
<td>water pays</td>
<td>properly</td>
<td>198</td>
</tr>
</tbody>
</table>

**Gale, Church, and Yarowsky (1992): Even Remote Context is Informative**

The figure below shows the distance of context words from an ambiguous word, while the second axis shows the nearest context words. The figure shows that the nearest words are informative, while the second context words are not as informative.

**Gale, Church, and Yarowsky (1992): Wide Contexts are Useful**

The figure below shows the number of training instances needed to achieve a certain level of accuracy. The figure shows that increasing the context length increases the accuracy, but also increases the computational cost.

**Gale, Church, and Yarowsky (1992): Even just a few training instances help**

The figure below shows the number of training instances needed to achieve a certain level of accuracy. The figure shows that increasing the context length increases the accuracy, but also increases the computational cost.
Other methods for WSD

- **Supervised**
  - Bein et al. 1993: using mutual information to combine senses into group
  - Tovey et al. 1992: using a lexicon and a topic-based corpus
  - Not recently, any multiclass learning method when names you know
- **Unsupervised**
  - Discourse shine (unsupervised based on things, USA)
  - WordNet's 1.05 bootstrapping algorithm
  - Discourse shine, example of using context and content to control each other. More realistic

Principles:

- One sense per collocation
- One sense per discourse
- Be ad context vs. collocations: both are useful

---

Evaluation

- **Baseline**: is the system an improvement?
  - Unsupervised: Random, Simple-Link
  - Supervised: Most, Frequent, Lexicon plus corpus

- **Upper bound**: agreement between humans?

---

SENSEVAL

Goals:

- Provide a common framework to compare WSD systems
- Standardise the test (especially evaluation procedure)
- Build and distribute new lexical resources

Website: [http://www.senseval.org](http://www.senseval.org)

- There are now many computer programs for a statistically determining the
  sense of words in context. Word Sense Disambiguation at NIST. The paper is
  very large. It gives new ideas about the question of how best to approach the
  problem with respect to different kinds, different domains of language, and
different sources of language.

- ACL SIGLEX workshop (293): Tannen and Markey paper
- SENSEVAL-1 (1998); SENSEVAL-2 (Toshiba, 2001)
- Usual SENSEVAL and All Words

---

SENSEVAL-II

Choosing the right sense for a word among those of WordNet.

**Sense 1**: horses, horse shrill -- [traditionally] predicate-adjacent deontic modal sense (pragmatic sense)
**Sense 2**: horses -- is a padded gymnastic apparatus
**Sense 3**: cavalry, horse cavalry, horses -- (temporal tensed to right on homogenous: "All horses lead the way"
**Sense 4**: equestrian, horses, equestrian, horse -- is
**Sense 5**: horse, horses -- is another in the shape of a horse's head, not even legs against horizontally and use1 altively (or vice versa)
**Sense 6**: horse, horses, equestrian, horse -- is
**Sense 7**: horse, horses, equestrian, horse -- is

---

English All Words: All N, V, Adj, Adv

- **Data**: 3 tests for a total of 1770 words
- **average": polysemy": 6.5
- **Example": Calm off Tote 1

The art of change–singing is peculiar to the English and, like many English peculiarities, unmistakable to the rest of the world. -- Dorothy L. Sayers, *The Nine Tailors* - ALAN Turing, *England*: Of all senses that

---

English All Words (2-Dimensional scaling)
English Lexical Sample

- Data: 8699 texts for 73 words
- Average WN polysemy: 9.22
- Training Data: 8566 (average 118/word)
- Baseline (commonest): 0.47 precision
- Baseline (Lesk): 0.31 precision

Quiz

Which of the following was not one of Church and Gale’s (1992) claims about WSD?

a) Context helps, even if very distant
b) Context helps, even if very wide
c) Context helps, even if POS tags are used instead of words
d) Context helps, even if only a few training examples

2004/05 ANLE 45

LEXICAL ACQUISITION: Lexical and Distributional notions of meaning similarity

Thesaurus-based word similarity

- How can we work out how similar in meaning words are?
- We could use anything in the thesaurus
  - Meronymy
  - hypernymy
  - Sample sentences
    - By “thesaurus-based” we usually just mean
      - Using the in-building hypernymy hierarchy
    - Word similarity with-in-word relationships
    - Similar words are near-antonyms
    - Not related could be related any way
    - Or, opposite, related, not similar
    - Or, opposite, similar

Path based similarity

- Two words are similar if nearby in thesaurus hierarchy (i.e. short path between them)
Problem with basic path-based similarity

- Assumes each link represents a uniform distance
- Nickel to money seems closer than nickel to standard instead.
- We ask a metric which attributes represent the cost of each edge independently
- There have been a whole slew of methods that augment thesaurus with notions from a corpus (Salton, Lin, ...)
- But we won’t cover them here.

The coverage problem

- Sampson (1989): tested coverage of Oxford ALO (~70,000 entries) looking at a 45,000-token subsample of the LOB. About 3% of tokens not listed in dictionary
- Examples:

<table>
<thead>
<tr>
<th>Type of problem</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proper noun</td>
<td>Caracas, Chateau-Chiat</td>
</tr>
<tr>
<td>Foreign word</td>
<td>pavarotti</td>
</tr>
<tr>
<td>Trans-derived words</td>
<td>nauseability</td>
</tr>
<tr>
<td>Code</td>
<td>R01</td>
</tr>
<tr>
<td>Non-standard English</td>
<td>yours</td>
</tr>
<tr>
<td>Hyphen omitted</td>
<td>black/white</td>
</tr>
<tr>
<td>Technical vocabulary</td>
<td>nomenclature</td>
</tr>
</tbody>
</table>

The limits of hand-encoded lexical resources

- Manual construction of lexical resources is very costly
- Because language keeps changing, these resources have to be continuously updated
- Some information (e.g., about frequencies) has to be computed automatically anyway

Vector-based lexical semantics

- Very old idea in NL engineering: the meaning of a word can be specified in terms of the values of certain “features” (COMPONENTAL SEMANTICS)
- dog: animate, eat, SAT/OMIT, social: +
- horse: animate+, eat, social: -
- cat: animate-, eat, social: +
- Similarity measured: proximity in feature space

General characterization of vector-based semantics

- Vectors as models of concepts
- The CLUSTERING approach to lexical semantics:
  1. Define properties one cares about, and give values to each property (generally numerical)
  2. Create a vector of length n for each item to be classified
  3. Viewing the n-dimensional vector as a point in n-space, cluster points that are near one another
- What changes between models:
  1. The properties used in the vector
  2. The distance metric used to decide if two points are “close”
  3. The algorithm used to cluster
- For similarity based approaches, skip the 3rd step
**Distributional Similarity: Using words as features in a vector-based semantics**

- The old compositional semantics approach requires:
  1. Specifying the features.
  2. Characterizing the value of these features for each sentence.
- A simpler approach is to use features that WORDS occur in proximity of.
- E.g., You shall know a word by the company it keeps.
- More specifically, you can calculate various statistics of these features.
- The idea is that words occur in the vicinity of other words whose meanings are similar.
- For instance, the probabilities that these words occur in the vicinity of each other can give good correlation.
- Lexical associations learned this way correlate very well with priming experiments. (Lund et al, 1997).
- We can extract the following co-occurrence matrix:

<table>
<thead>
<tr>
<th>Word</th>
<th>astronaut</th>
<th>moon</th>
<th>car</th>
<th>truck</th>
</tr>
</thead>
<tbody>
<tr>
<td>astronaut</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>moon</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>car</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>truck</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

**Using neighboring words to specify the meaning of words**

- Take, e.g., the following corpus:
  1. John ate a banana.
  2. John ate an apple.
  3. John drove a car.
- We can extract the following co-occurrence matrix:

<table>
<thead>
<tr>
<th>Word</th>
<th>astronaut</th>
<th>moon</th>
<th>car</th>
<th>truck</th>
</tr>
</thead>
<tbody>
<tr>
<td>astronaut</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>moon</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>car</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>truck</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

**Acquiring lexical vectors from a corpus (Schutze, 1991; Burgess and Lund, 1997)**

- To construct vectors C(w) for each word w:
  1. Scan a text.
  2. Whenever a word w is encountered, increment all cells of C(w) corresponding to the words in the vicinity of w.
- Differences among methods:
  - Subset windows
  - Weighted / dampened or not
- For instance, words in the vocabulary counts as dimension (including function words such as the and of) or whether instead only some specially chosen words are used (typically the most common content words in the corpus, or perhaps modifiers only. The words chosen as dimensions are often called CONTEXT WORDS.
- Whether dimensionality reduction methods are applied.

**Variant: using modifiers to specify the meaning of words**

- ... The Soviet cosmonaut ... The American astronaut ... The red American car ... the red truck ... the spacewalking cosmonaut ... the full Moon ...

---

**Measures of semantic similarity**

- **Euclidean distance:**
  \[ d = \sum_{i=1}^{n} (x_i - y_i)^2 \]
- **Cosine:**
  \[ \cos (\theta) = \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2} \sqrt{\sum_{i=1}^{n} y_i^2}} \]
- **Manhattan (L1) Metric:**
  \[ d = \sum_{i=1}^{n} |x_i - y_i| \]

**The HAL model (Burgess and Lund, 1995, 1997)**

- A 160 million words corpus of articles extracted from all newsgroups containing English dialogue.
- Context words: the 70,000 most frequently occurring symbols within the corpus.
- Window size: 10 words to the left and 10 words to the right.
- Measure of similarity: cosine.
Latent Semantic Analysis (LSA) (Landauer et al., 1997)

- Goal: extract relations of expected contextual usage from passages
- Steps:
  1. Build a word/document co-occurrence matrix
  2. "Weight" each cell (e.g. 1/df)
  3. Perform a DIMENSIONALITY REDUCTION with SVD
- Argued to correlate well with humans on a number of tests

Detecting hyponymy and other relations

- Could we discover new hyponyms, and add them to a taxonomy under the appropriate hyponym?
- Why is this important?
  - "healthy" and "hypersensitive" occur in WordNet 2.1, but "health" and "hypersensitive" are not.
  - "clickability" and "inability" but not "clickability", "inability", or "actability".
  - HTML and "SQL", but not "XML or "HTML".
- "Google" and "Facebook", but not "Microsoft" or "IBM".
- This unknown word problem occurs throughout NLP.

Hearst (1992) Approach

- Agar is a substance prepared from a mixture of red algae, such as Gelidium, for laboratory or industrial use.
- What does Gelidium mean? How do you know?

\[ NP_0 \text{ such as } NP_i \in \{ NP_i \ldots \text{and/or} \} \}] \geq 1 \]

implies the following semantics

\[ \forall NP_i \geq 1, \text{hyponym}(NP_i, NP_0) \]

allowing us to infer

hyponym(Gelidium, red algae)

Hearst’s hand-built patterns

- Np:NP \lor (\text{Prep}) \text{ Prep}_{-}\text{Prep}_{+} \text{ Prep}_{-}\text{Prep}_{+} \text{ Prep}_{-}\text{Prep}_{+} \text{ Prep}_{-}\text{Prep}_{+} \text{ Prep}_{-}\text{Prep}_{+} \text{ Prep}_{-}\text{Prep}_{+}

Each of Hearst’s patterns can be captured by a syntactic dependency path in MINIPAR:

Hearst Pattern

$Y \text{ such as } X$...

MINIPAR Representation

$Y \text{ such as } X$

$\text{Prep}_{-}\text{Prep}_{+} \text{ Prep}_{-}\text{Prep}_{+} \text{ Prep}_{-}\text{Prep}_{+} \text{ Prep}_{-}\text{Prep}_{+} \text{ Prep}_{-}\text{Prep}_{+} \text{ Prep}_{-}\text{Prep}_{+}$

Such $Y$ as $X$...

$\text{Prep}_{-}\text{Prep}_{+} \text{ Prep}_{-}\text{Prep}_{+} \text{ Prep}_{-}\text{Prep}_{+} \text{ Prep}_{-}\text{Prep}_{+} \text{ Prep}_{-}\text{Prep}_{+} \text{ Prep}_{-}\text{Prep}_{+}$

$X \ldots \text{and other } Y$

$\text{Prep}_{-}\text{Prep}_{+} \text{ Prep}_{-}\text{Prep}_{+} \text{ Prep}_{-}\text{Prep}_{+} \text{ Prep}_{-}\text{Prep}_{+} \text{ Prep}_{-}\text{Prep}_{+} \text{ Prep}_{-}\text{Prep}_{+}$

$\text{and}\text{ punc}\text{ N}\text{ other}\text{ A}\text{ mod}\text{ N}$

$\text{and}\text{ punc}\text{ N}\text{ other}\text{ A}\text{ mod}\text{ N}$

$\text{and}\text{ punc}\text{ N}\text{ other}\text{ A}\text{ mod}\text{ N}$

$\text{and}\text{ punc}\text{ N}\text{ other}\text{ A}\text{ mod}\text{ N}$
**Algorithm (Snow, Jurafsky, and Ng 2005)**

- Collect noun pairs from corpora (732,311 pairs from 6 million words of newswire).
- Identify each pair as positive or negative example of hyperonym-hyponym relationship (14,387 yes, 737,924 no).
- Parse the sentences, extract patterns (69,592 dependency paths occurring in >5 pairs).
- Train a hyperonym classifier on these patterns. We could interpret each path as a binary classifier.
- Logistic regression with 69,592 features (actually converted to 974,288 bucketed binary features).

**Using Discovered Patterns to Find Novel Hyponym/Hyperonym Pairs**

Example of a discovered high-precision path:

```
Nobels/<category> Nobel: The annual award from the Nobel Institute.
```

- may be used to discover new hyperonym pairs not in WSDnet.
- useful for “composition” and “adjective” hyponymy.
- may find cases where other methods have other reasons for.

**Using each pattern/feature as a binary classifier: Hyperonym Precision / Recall**

**There are lots of fun lexical semantic tasks: Logical Metonymy**

- Additional meaning areas from characterization of an event:
  - Mary finished her dinner.
  - Mary finished eating her dinner.
- Medically, drinking her boss.
  - Mary finished drinking her boss.
- Mary finished her drink.
  - Mary finished her ale.
- How can we work out the implicit activities?

**Logical metonymy**

- Easy cake
- Easy cake to bake
- Good soup
- Good to eat NOT enjoyable to make
- Fast plane
- Flies fast NOT fast to construct

- There is a default interpretation, but it depends on context:
  - All the office personnel had left the company sports day last week.
  - One of the programmers was a good athlete, but the other was
  - Struggling to finish the event.
- The fast programmer came first in the 10km.
- Some cases seem to lack default metonymic interpretations:
  - John enjoyed the dictionary.

**How can you learn them? (Lapata and Lascarides 2000)**

- Corpus statistics
- But cases that fill in the metonymic interpretation (beginV NP orgive V NP)
  - and bare are not regularly used
- So just use general facts about verbs and complements
- The best result of overcome the problem to be independent of whether it is the
  - complement of another verb
- Examples learned by model
  - Began story: To begin a story
  - Begins story: To begin a story
  - The book looks: To enjoy reading books
- This doesn’t do context-based interpretation, of course!