

Word Sense Disambiguation

CSC485 Lecture 8

Fishing for Bass

Word Sense Disambiguation

- Word sense disambiguation (WSD)
- Lexical disambiguation
- Resolving lexical ambiguity
- Lexical ambiguity resolution



How big is the problem?

- Most words of English have only one sense
 - 62% in Longman's Dictionary of Contemporary English
 - 79% in WordNet
- But the others tend to have several senses
 - Avg 3.83 in LDOCE
 - 2.96 in WordNet
- Ambiguous words are more frequently used
 - In British National Corpus, 84% of instances have more than one sense in WordNet
- Some senses are more frequent than others.



Words occurring in the British National Corpus are plotted on the horizontal axis in rank order by frequency in the corpus. Number of WordNet senses per word is plotted on the vertical axis. Each point represents a bin of 100 words and the average number of senses of words in the bin.



In each column, the senses are ordered by frequency, normalized per word, and averaged over all words with that number of senses.

Edmonds, Philip. "Disambiguation, Lexical." Encyclopedia of Language and Linguistics (second edition), Elsevier, 2006, pp 607–623.

Sense inventory of a word

- Dictionaries, WordNet list senses of a word.
- Often, no agreement on proper sense-division of words.
- Don't want sense-divisions to be too coarse-grained or too fine-grained.
 - Frequent criticism of WordNet

trench (trěnch) n. **1.** A deep furrow or ditch. **2.** A long, narrow ditch embanked with its own soil and used for concealment and protection in warfare. **3.** A long, steep-sided valley on the ocean floor. —trench v. trenched, trench·ing, trench·es. —tr. **1.** To

The American Heritage Dictionary of the English Language (3rd edition)

trench /trent∫/ *n* ditch dug in the ground, eg for drainage or to give troops shelter from enemy fire: *irrigation trenches* ○ *The workmen dug a trench for*

Oxford Advanced Learner's Dictionary (encyclopedic edition)

lit ter (lit 'ar) n. 1.a. A disorderly accumulation of objects; a pile. b. Carelessly discarded refuse, such as wastepaper: the litter in the streets after a parade.
2. The offspring produced at one birth by a multiparous mammal. See Synonyms at flock ¹. 3.a. Material, such as straw, used as bedding for animals. b. An absorbent material, such as granulated clay, for covering the floor of an animal's cage or excretory box.
4. An enclosed or curtained couch mounted on shafts and used to carry a single passenger.
5. A flat supporting framework, such as a piece of canvas stretched between parallel shafts, for carrying a disabled or dead person; a stretcher.
6. The uppermost layer of the forest floor consisting chiefly of fallen leaves and other decaying organic matter. — litter

litter /'Irtə(r)/ n **1** (**a**) [U] light rubbish (eg bits of paper, wrappings, bottles) left lying about, esp in a public place: *Please do not leave litter*. \Rightarrow article at ENVIRONMENT. (**b**) [sing] state of untidiness: *Her desk was covered in a litter of books and papers*. \circ *His room was a litter of old clothes, dirty crockery and broken furniture*. **2** [U] straw, etc used as bedding for animals. **3** [CGp] all the young born to an animal at one time: *a litter of puppies*. **4** [C] (**a**) type of stretcher(1). (**b**) (formerly) couch carried on men's shoulders or by animals as a means of transport.

AHDEL

OALD

What counts as the right answer?

- Often, no agreement on which sense a given word-token is.
- Some tokens seem to have two or more senses at the same time.

Which senses are these?

• Image

1. a picture formed in the mind;

- 2. a picture formed of an object in front of a mirror or lens;
- 3. the general opinion about a person, organization, etc, formed or intentionally created in people's minds;

[and three other senses]

"… of the Garonne, which becomes an unforgettable image. This is a very individual film, mannered, …"

Example from: Kilgarriff, Adam. "Dictionary word sense distinctions: An enquiry into their nature." *Computers and the Humanities*, 26: 365–387, 1993. Definitions from *Longman Dictionary of Contemporary English*, 2nd edition, 1987.

Which senses are these?

distinction

1. the fact of being different;



2. the quality of being unusually good; excellence.

"... before the war, shares with Rilke and Kafka the *distinction* of having origins which seem to escape ..."

Example from: Kilgarriff, Adam. "Dictionary word sense distinctions: An enquiry into their nature." *Computers and the Humanities*, 26: 365–387, 1993. Definitions from *Longman Dictionary of Contemporary English*, 2nd edition, 1987.

What counts as the right answer?

- Therefore, hard to get a definitive sense-tagged corpus.
- And hard to get human baseline for performance.
 - Human annotators agree about 70–95% of the time. [Depending on word, sense inventory, context size, discussions, etc.]

Baseline algorithms

- Assume that input is PoS-tagged.
- Obvious baseline algorithm: Pick most-likely sense (or pick one at random).
- Accuracy: 39–62%

Baseline algorithms

- Simple tricks (1): Notice when ambiguous word is in unambiguous fixed phrase.
 - private school, private eye.
 - All right?



Baseline algorithms

- Simple tricks (2): "One sense per discourse"
 A homonymous word is rarely used in more than one sense in the same text.
 - sense in the same text.
 - If word occurs multiple times, ...
 - Not true for **polysemy**.
- Simple tricks (3): Lesk's algorithm (see below).

"Context"

- Meaning of word in use depends on (determined by) its context.
 - Circumstantial context.
 - Textual context.
 - Complete text.
 - Sentence, paragraph.
 - Window of n words.

"Context"

- Words of context are also ambiguous; need for mutual constraints; often ignored in practice.
- "One sense per collocation".
- Collocation: words that *tend* to co-occur together.

Selectional preferences

- Constraints imposed by one word meaning on another especially verbs on nouns.
 - I don't mind washing **dishes** now and then.
 - It was the most popular **dish** <u>served</u> in the Ladies' Grill.
- Some words select more strongly than others.
 see (weak) drink (moderate) elapse (strong)

see you

with eyes? meet?

time elapse

Limitations of selectional preferences

• Negation:

You can't <mark>eat good intentions</mark>. It's nonsense to say that a <mark>book elapsed</mark>.

• Odd events:

Los Angeles secretary Jannene Swift **married** a 50-pound pet rock in a formal ceremony in Lafayette Park. (Newspaper report)

• Metaphor:

The issue was acute because the exiled Polish Government in London, supported in the main by Britain, was still competing with the new Lublin Government formed behind the Red Army. More time was spent in trying to **marry** these incompatibles than over any subject discussed at Yalta. ... The application of these formulae could not please both sides, for they really attempted to **marry** the impossible to the inevitable.

Limitations of selectional preferences

- In practice, attempts to induce selectional preferences or to use them have not been very successful.
 - Apply in only about 20% of cases, achieve about 50% accuracy. (Mihalcea 2006, McCarthy & Carroll 2003)
 - At best, they are a coarse filter for other methods.

Lesk's algorithm

• Sense s_i of ambiguous word w is likely to be the intended sense if many of the words used in the dictionary definition of s_i are also used in the definitions of words in the context window.

... the <u>keyboard</u> of the **terminal** was ...

More overlap, more likely to be the sense!

terminal

- 1. a point on an electrical device at which electric current enters or leaves.
- 2. where transport vehicles load or unload passengers or goods.
- 3. an input-output device providing access to a **computer**.

keyboard

- 1. set of keys on a piano or organ or typewriter or typesetting machine or **computer** or the like.
- 2. an arrangement of hooks on which keys or locks are hung.





Lesk's algorithm

- Sense s_i of ambiguous word w is likely to be the intended sense if many of the words used in the dictionary definition of s_i are also used in the definitions of words in the context window.
- For each sense s_i of w, let D_i be the bag of words in its dictionary definition.
- **Bag of words**: unordered set of words in a string, excepting those that are very frequent (stop list).
- Let *B* be the bag of words of the dictionary definitions of all senses of all words $v \neq w$ in the context window of *w*. (Might also (or instead) include all *v* in *B*.)
- Choose the sense s_i that maximizes overlap (D_i, B) .



Lesk's algorithm

• Many variants of overlap score, but most common are based on cosine similarity of vectors that count occurrences of each word.

• Results:

- Simple versions of Lesk achieve accuracy around 50–60%;
- Lesk plus simple smarts gets to nearly 70%.
- Many variants possible on what is included in D_i and B.
 - E.g., include the examples in dictionary definitions.
 - E.g., include other manually tagged example texts.
 - PoS tags on definitions.
 - Give extra weight to infrequent words occurring in the vectors.

Cosine Similarity Score (bag of words)



Yarowsky 1995 Unsupervised decision-list learning

• **Decision list**: ordered list of strong, specific clues to senses of homonym.*



*Yarowsky, never taken CSC485, calls them "polysemous words".



Fishing for Bass

Decision list for <i>bass</i> :			
	LogL	Context	Sense
	10.98	<i>fish</i> in ± <i>k</i> words	FISH
	10.92	striped bass	FISH
	9.70	<i>guitar</i> in ± <i>k</i> words	MUSIC
	9.20	bass player	MUSIC
	9.10	<i>piano</i> in ± <i>k</i> words	MUSIC
	8.87	sea bass	FISH
	8.49	play bass	MUSIC
	8.31	<i>river</i> in ± <i>k</i> words	FISH
	7.71	on bass	MUSIC
	5.32	bass are	FISH

Yarowsky: Basic ideas

- Separate decision list learned for each homonym.
- Bootstrapped from seeds, very large corpus, heuristics.
 - One sense per discourse.
 - One sense per **collocation**.
- Uses supervised classification algorithm to build decision-list.
- Training corpus: 460M words, mixed texts.

Yarowsky: Method

• 1–2. Get data (instances of target word); choose seed rules; apply them.



used to strain microscopic plant life from the zonal distribution of plant life . close-up studies of plant life and natural too rapid growth of aquatic plant life in water the proliferation of plant and animal life establishment phase of the plant virus life cycle that divide life into plant and animal kingdom many dangers to plant and animal life mammals . Animal and plant life are delicately



automated manufacturing plant in Fremont vast manufacturing plant and distribution chemical manufacturing plant, producing viscose keep a manufacturing plant profitable without computer manufacturing plant and adjacent discovered at a St. Louis plant manufacturing copper manufacturing plant found that they copper wire manufacturing plant, for example s cement manufacturing plant in Alpena vinyl chloride monomer plant , which is
molecules found in plant and animal tissue
Nissan car and truck plant in Japan is
and Golgi apparatus of plant and animal cells
union responses to plant closures .
cell types found in the plant kingdom are
company said the plant is still operating
Although thousands of plant and animal species
animal rather than plant tissues can be





Figure from Yarowsky 1995.

Initial state after use of seed rules

Yarowsky: Method

- 3. Iterate:
 - 3a. Create a new decision-list classifier: supervised training with the data tagged so far. Looks for collocations as features for classification.
 - 3b. Apply new classifier to whole data set, tag some new instances.
 - 3c. Optional: Apply one-sense-per-discourse rule wherever one sense now dominates a text.





Yarowsky: Method

- 4. Stop when converged.
 - (Optional: Apply one-sense-per-discourse constraint.)
- 5. Use final decision list for WSD.

Yarowsky: Evaluation

- Experiments: 12 homonymous words.
 - 400–12,000 hand-tagged instances of each.
 - Baseline (most frequent sense) = 63.9%.
- Best results, avg 96.5% accuracy.
 - Base seed on dictionary definition; use one-sense-perdiscourse heuristic.
 - As good as or better than supervised algorithm used directly on fully labelled data.

Yarowsky: Discussion

- Strength of method:
 - The one-sense heuristics.
 - Use of precise lexical and positional information.
 - Huge training corpus.
 - Bootstrapping: Unsupervised use of supervised algorithm.
- Disadvantages:
 - Train each word separately.
 - Homonyms only. Why?

Yarowsky: Discussion

- Not limited to regular words; e.g., in speech synthesis system:
 - "/" as fraction or date:

 $3/4 \rightarrow$ "three-quarters" or "third of April".

• Roman number as cardinal or ordinal:

chapter VII → "chapter seven"; *Henry VII* → "Henry the seventh".

Yarowsky, David. "Homograph disambiguation in speech synthesis." In Jan van Santen, Richard Sproat, Joseph Olive and Julia Hirschberg (eds.), *Progress in Speech Synthesis*. Springer-Verlag, pp. 159–175, 1996.