Word sense disambiguation (WSD)

- WSD: problem of picking the right sense for a word occurrence from a given set of senses.
- Homonymy vs Polysemy. Homonymy refers to words that are the same auditorily and/or textually but have completely different meanings. (e.g. bank: financial institution vs strip of land on the side of a river vs to tilt; son vs sun - auditorily same). Polysemy refers to slightly different senses of the same general sense (e.g. loan bank, piggy-bank, gene bank etc.). Linguists distinguish the two.
- In NLP - WSD assumes a set of given senses from which the right sense should be chosen. So polysemy refers to simply the set of senses.

Examples

- Chair - furniture or person or position.
  `- He sat in the chair with a sigh of relief.`
  `- The department is without a chair. - The Lucasian Chair at Cambridge was occupied by Stephen Hawking.`
- Child - human offspring or young person.
  `- The children crossed the street.`
  `- Women in developing countries have more children on average.`
- Title: ownership right or ownership document.
  `- The title of cultivable land should belong to the cultivator and not the landlord.`
  `- The title to the land was in a language he could not read.`

Applications

- **Machine translation:**
  the bank -

Applications

- **Machine translation:**
  the bank - (in French) la banque | la rive | la berge.
- **Information retrieval:**
  logic -

Applications

- **Machine translation:**
  the bank - (in French) la banque | la rive | la berge.
- **Information retrieval:**
  logic - mathematical | hardware
- **Knowledge Base construction:**
  `Kalyanpur will have elections on 1 Mar. 2017.`

Applications

- **Machine translation:**
  the bank - (in French) la banque | la rive | la berge.
- **Information retrieval:**
  logic - mathematical | hardware
- **Knowledge Base construction:**
  `Kalyanpur will have elections on 1 Mar.  2017.`
  Kalyanpur Near IITK | Near Fatehpur - many other Kalyanpurs in India.
- **Question answering:**
  Who is a good match for Andre Agassi?

Applications

- **Machine translation:**
  the bank - (in French) la banque | la rive | la berge.
- **Information retrieval:**
  logic - mathematical | hardware
- **Knowledge Base construction:**
  `Kalyanpur will have elections on 1 Mar. 2017.`
  Kalyanpur Near IITK | Near Fatehpur - many other Kalyanpurs in India.
- **Question answering:**
  Who is a good match for Andre Agassi?
  Steffi Graf | Pete Sampras.

Coarse vs Fine grained WSD

- Disambiguation can be at the level of homographs/homonymy (coarse grained) or polysemy (fine grained).
- Polysemy is harder.

Approaches to WSD

- Using a resource like a dictionary, thesaurus or other (Wordnet).
- Using ML - supervised. Requires corpus with tagged senses.
- Unsupervised - word sense discrimination. Creates clusters of words with similar senses. So, not strictly disambiguation.

Lesk's algorithm

Uses a dictionary, thesaurus or Wordnet as resource.

Algorithm 0.1: $\textsc{Lesk}(\mathit{word}, \mathit{sent})$

```
comment: Uses a dictionary as resource.  Finds best
         sense for word in sentence sent.
sense ← mostFrequentSense(word);
cntxt ← neighboursOfWord(word, sent);
maxOverlap ← 0;
for each (s ∈ sensesInDict(word))
  do { neighboursInDict ← findNeighboursInExamplesDefn(s);
       overlap ← calcOverlap(cntxt, neighboursInDict);
       if (overlap > maxOverlap)
         then { sense ← s;
                maxOverlap ← overlap;
return (sense)
```

Supervised approaches

- Features used vary. Most detailed features include PoS and position information for $\pm m$ ($m \approx 50$) neighbours. Nouns tend to require many more neighbours. Verbs are more localized.
- Standard ML algorithms like: DTs, k-NN, cosine-similarity, naive Bayes, SVMs, Adaboost, ensemble models are in use.
- The right choice of features seems more important than choice of classifier. Morphology helps in some cases e.g. word is singular/plural.
- Data sets are not very numerous. SenseEval, SemEval data sets are available. Brown corpus also has senses marked. Ontonotes is another useful data set.

WSD using word-class disambiguation

- Consider the word `crane` - bird | machine. Often it is enough to know the word-class to infer the word-sense given the context.
- For word-class disambiguation it is enough to create a BIRDlist and a MACHINElist from the corpus. A fully labelled corpus is not needed. Even if the lists contain some examples of polysemous words (e.g. eagle, hawk in BIRDlist) these will typically be much more rare and the secondary senses are often evenly distributed.
- The sense can be found by $P(s_i|cntxt) = \sum_{j=1}^{N} P(s_i|class_j)P(class_j|cntxt)$
- A variant uses a single monosemous word closest to each class as a training proxy. For e.g. `heron`, `derrick` for `crane`.

Using hierarchical classes and selection constraints

- Define a probability distribution over a word hierarchy.

Word		**Wordnet Hypernyms**
`gin`	$\rightarrow$	(beverage, device, game)
`vodka`	$\rightarrow$	(beverage)
`pint`	$\rightarrow$	(volSpec)
`coffee`	$\rightarrow$	(beverage, seed, tree, colour)
`cup`	$\rightarrow$	(volSpec, vessel, trophy)

- Compute verb-object pair probablities e.g. drink(gin), play(gin), drink(coffee), grind(coffee) etc. from a parsed corpus. From this one can calculate $P(WordnetClass|drink)$ e.g. $P(beverage|drink)$. A small probability is uniformly assigned to each option and the rest is based on frequencies.
- These probabilities can then be used either directly or together with context words to disambiguate.

Graph approaches to WSD

- Use small amount of seed data (labelled data plus dictionary) and a few senses to bootstrap and tag all instances of words.
- Based on two empirical observations: a) a word mostly has only one sense per collocation b) a word mostly has only one sense per discourse.
- Starting from small labelled sets a classifier is learned and this is used to label some unknown instances that have labels with high confidence and that are within a collocation/discourse. This process is then iterated.

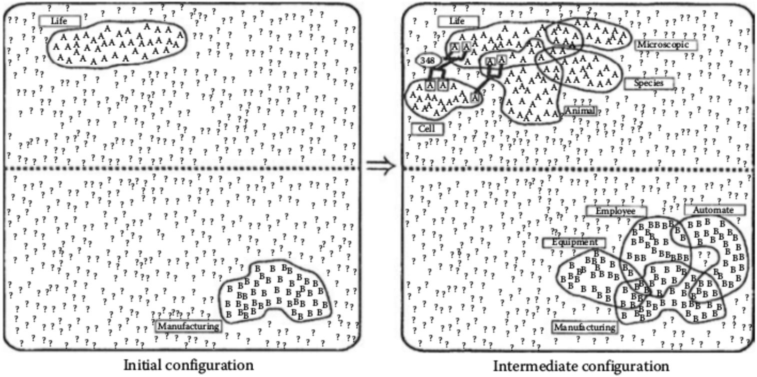

Figure: Iterative boostrapping for two senses of word `plant`. Source: Yarowsky, WSD, in Indurkhya, NLP Handbook, 2nd Ed.