Natural Language Processing

Word Meaning
Word Senses

Sudeshna Sarkar

22 August 2019

Word Sense Disambiguation (WSD)

- Given
  - A word in context
  - A fixed inventory of potential word senses
- Task: Create a system that automatically decides the correct sense

Why WSD?

- One of the central challenges in NLP.
- Needed in:
  - **Machine Translation**: For correct lexical choice.
  - **Information Retrieval**: Resolving ambiguity in queries.
  - **Information Extraction**: For accurate analysis of text.
- Computationally determining which ***sense*** of a word is activated by its use in a particular ***context***.
  - E.g. I am going to withdraw money from the ***bank.***
- A classification problem:
  - Senses → Classes

Two variants of WSD task

- Lexical Sample task
  - Small pre-selected set of target words (*line, plant*)
  - And inventory of senses for each word
  - **Supervised machine learning: train a classifier for each word**
- All-words task
  - Every word in an entire text
  - A lexicon with senses for each word
  - Data sparseness: can't train word-specific classifiers

WordNet Senses

- WordNets senses (like many dictionary senses) tend to be very fine-grained.
- “play” as a verb has 35 senses, including
  - play a role or part: “Gielgud played Hamlet”
  - pretend to have certain qualities or state of mind: “John played dead.”
- Difficult to disambiguate to this level
- Not clear such fine-grained senses are useful for NLP.
- Several proposals for grouping senses into coarser, easier to identify senses (e.g. homonyms only).

Approaches to WSD

- **Knowledge Based Approaches**
  - Rely on knowledge resources like WordNet, Thesaurus etc.
  - May use grammar rules for disambiguation.
  - May use hand coded rules for disambiguation.
- **Machine Learning Based Approaches**
  - Rely on corpus evidence.
  - Train a model using tagged or untagged corpus.
  - Probabilistic/Statistical models.
- **Hybrid Approaches**
  - Use corpus evidence as well as semantic relations form WordNet.
- **Semi-supervised Approaches**

Knowledge based WSD using Selection Preferences

Sense 1

- This airlines ***serves*** dinner in the evening flight.
- serve (Verb)
  - agent
  - object – edible

Sense 2

- This airlines ***serves*** the sector between Agra & Delhi.
- serve (Verb)
  - agent
  - object – sector

Requires exhaustive enumeration of:

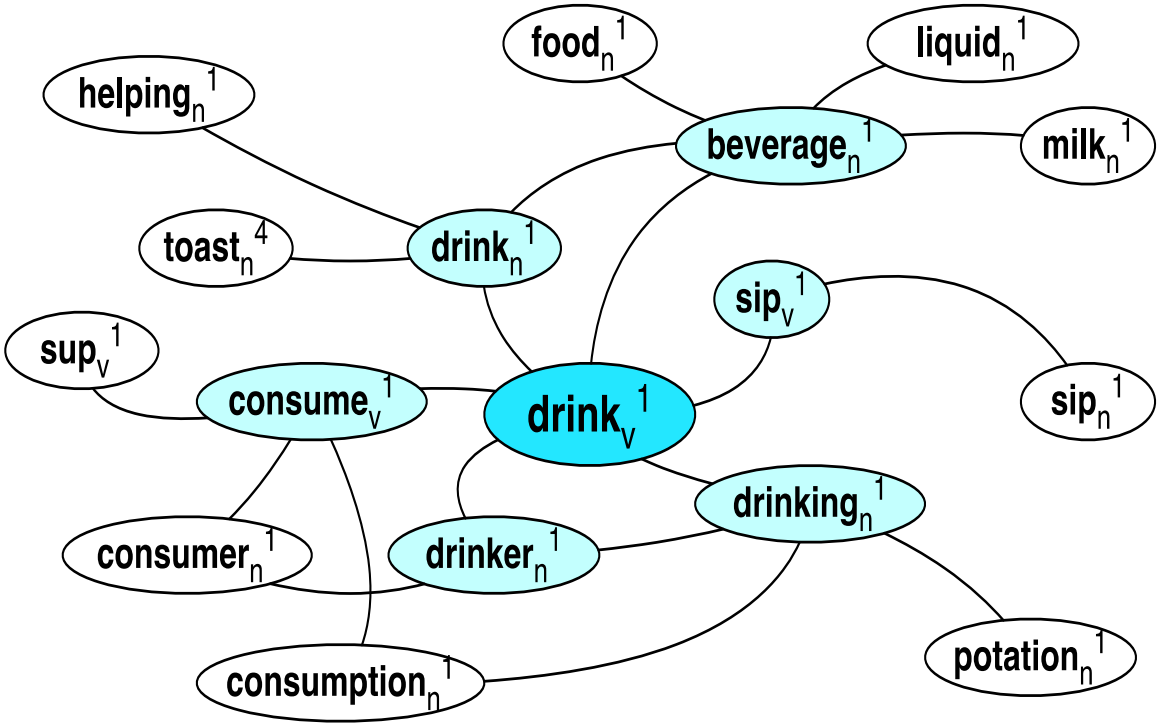
- Argument-structure of verbs.
- Selectional preferences of arguments.
- Description of properties of words such that meeting the selectional preference criteria can be decided.

  E.g. This flight serves the "**region"** between Mumbai and Delhi

  How do you decide if "region" is compatible with "sector"

Graph-based methods

- First, WordNet can be viewed as a graph
  - senses are nodes
  - relations (hypernymy, meronymy) are edges
  - Also add edge between word and unambiguous gloss words



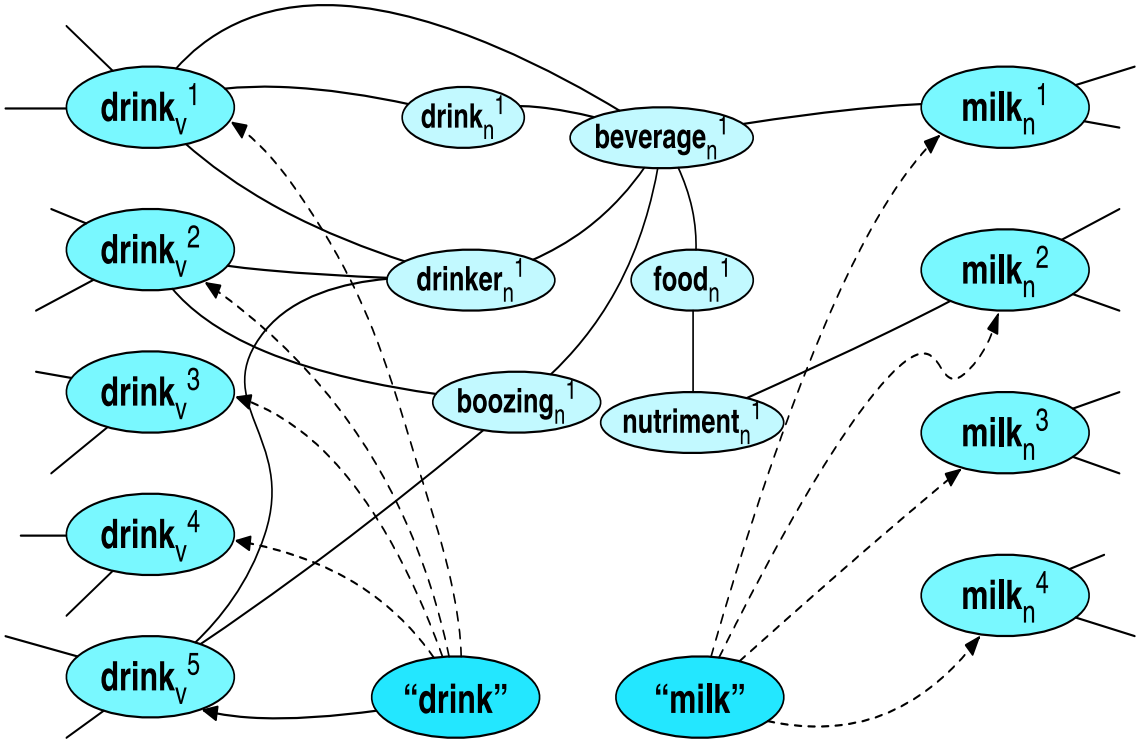
Use the graph for WSD

- Insert target word and words in its sentential context into the graph, with directed edges to their senses

"She drank some milk"

- Now choose the *most central* sense

Add some probability to "drink" and "milk" and compute node with highest "pagerank"



Supervised Machine Learning Approaches

- Supervised machine learning approach:
  - a training corpus of words tagged in context with their sense
  - used to train a classifier that can tag words in new text
- Summary of what we need:
  - the **tag set** (“sense inventory”)
  - the **training corpus**
  - A set of **features** extracted from the training corpus
  - A **classifier**

Corpus for Supervised WSD

- Lexical sample task:
  - *Line-hard-serve* corpus - 4000 examples of each
  - *Interest* corpus - 2369 sense-tagged examples
- All words:
  - **Semantic concordance**: a corpus in which each open-class word is labeled with a sense from a specific dictionary/thesaurus.
    - SemCor: 234,000 words from Brown Corpus, manually tagged with WordNet senses
    - SENSEVAL-3 competition corpora - 2081 tagged word tokens

SemCor

<wf pos=PRP>**He**</wf>

<wf pos=VB lemma=recognize wnsn=4 lexsn=2:31:00::>**recognized**</wf>

<wf pos=DT>**the**</wf>

<wf pos=NN lemma=gesture wnsn=1 lexsn=1:04:00::>**gesture**</wf>

<punc>.</punc>

Learning for WSD

- Assume **part-of-speech** (POS) determined.
- Treat as a classification problem with the appropriate potential senses for the target word given its POS as the categories.
- Encode context using a set of features to be used for disambiguation.
- Train a classifier on labeled data encoded using these features.
- Use the trained classifier to disambiguate future instances of the target word given their contextual features.

Feature Engineering: Contextual Features

- Surrounding bag of words
- POS of neighboring words
- Local collocations
- Syntactic relations

Surrounding Bag of Words

- Unordered individual words near the ambiguous word.
- Words in the same sentence.
- May include words in the previous sentence or surrounding paragraph.
- Gives general topical cues of the context.
- May use feature selection to determine a smaller set of words that help discriminate possible senses.
- May remove common “stop words” such as articles, prepositions, etc.

POS of Neighboring Words

- Use POS of immediately neighboring words.
- Provides evidence of local syntactic context.
- P_{-i} is the POS of the word *i* positions to the left of the target word.
- P_i is the POS of the word *i* positions to the right of the target word.
- Typical to include features for:

  $P_{-3}, P_{-2}, P_{-1}, P_1, P_2, P_3$

Local Collocations

- Specific lexical context immediately adjacent to the word.
- For example, to determine if “interest” as a noun refers to “readiness to give attention” or “money paid for the use of money”, the following collocations are useful:
  - “in the interest of”
  - “an interest in”
  - “interest rate”
  - “accrued interest”
- $C_{i,j}$ is a feature of the sequence of words from local position *i* to *j* relative to the target word.
  - $C_{-2,1}$ for “in the interest of” is “in the of”
- Typical to include:
  - Single word context: $C_{-1,-1}$, $C_{1,1}$, $C_{-2,-2}$, $C_{2,2}$
  - Two word context: $C_{-2,-1}$, $C_{-1,1}$,$C_{1,2}$
  - Three word context: $C_{-3,-1}$, $C_{-2,1}$, $C_{-1,2}$, $C_{1,3}$

Syntactic Relations (Ambiguous Verbs)

- For an ambiguous verb, it is very useful to know its direct object.
  - "played the game"
  - "played the guitar"
  - "played the risky and long-lasting card game"
  - "played the beautiful and expensive guitar"
  - "played the big brass tuba at the football game"
  - "played the game listening to the drums and the tubas"
- May also be useful to know its subject:
  - "The game was played while the band played."
  - "The game that included a drum and a tuba was played on Friday."

Syntactic Relations (Ambiguous Nouns)

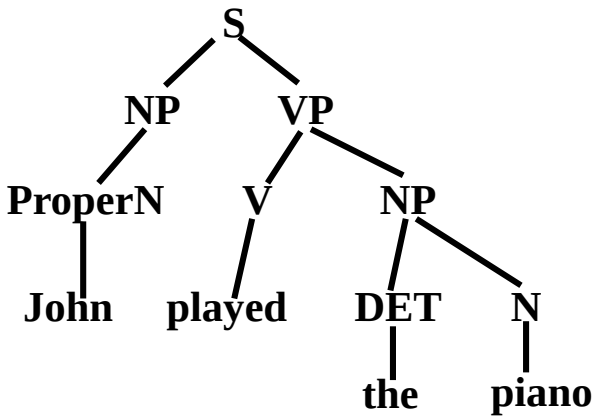
- For an ambiguous noun, it is useful to know what verb it is an object of:
  - "played the piano and the horn"
  - "wounded by the rhinoceros' horn"
- May also be useful to know what verb it is the subject of:
  - "the bank near the river loaned him $100"
  - "the bank is eroding and the bank has given the city the money to repair it"

Syntactic Relations (Ambiguous Adjectives)

- For an ambiguous adjective, it useful to know the noun it is modifying.
  - "a brilliant young man"
  - "a brilliant yellow light"
  - "a wooden writing desk"
  - "a wooden acting performance"

Using Syntax in WSD

- Produce a parse tree for a sentence using a syntactic parser.



- For ambiguous verbs, use the head word of its direct object and of its subject as features.
- For ambiguous nouns, use verbs for which it is the object and the subject as features.
- For ambiguous adjectives, use the head word (noun) of its NP as a feature.

Evaluation of WSD

- **Intrinsic evaluation**
  - Corpus developed in which one or more ambiguous words are labeled with explicit sense tags according to some sense inventory.
  - Corpus used for training and testing WSD and evaluated using accuracy (percentage of labeled words correctly disambiguated).
    - Use most common sense selection as a baseline.
- **extrinsic ('end-to-end', `task-based') evaluation**
  - Embed WSD algorithm in a task and see if you can do the task better!
  - such as machine translation, information retrieval, or question answering.

Bootstrapping

- For `bass`
  - Rely on “One sense per collocation” rule
    - A word reoccurring in collocation with the same word will almost surely have the same sense.
  - the word `play` occurs with the music sense of bass
  - the word `fish` occurs with the fish sense of bass

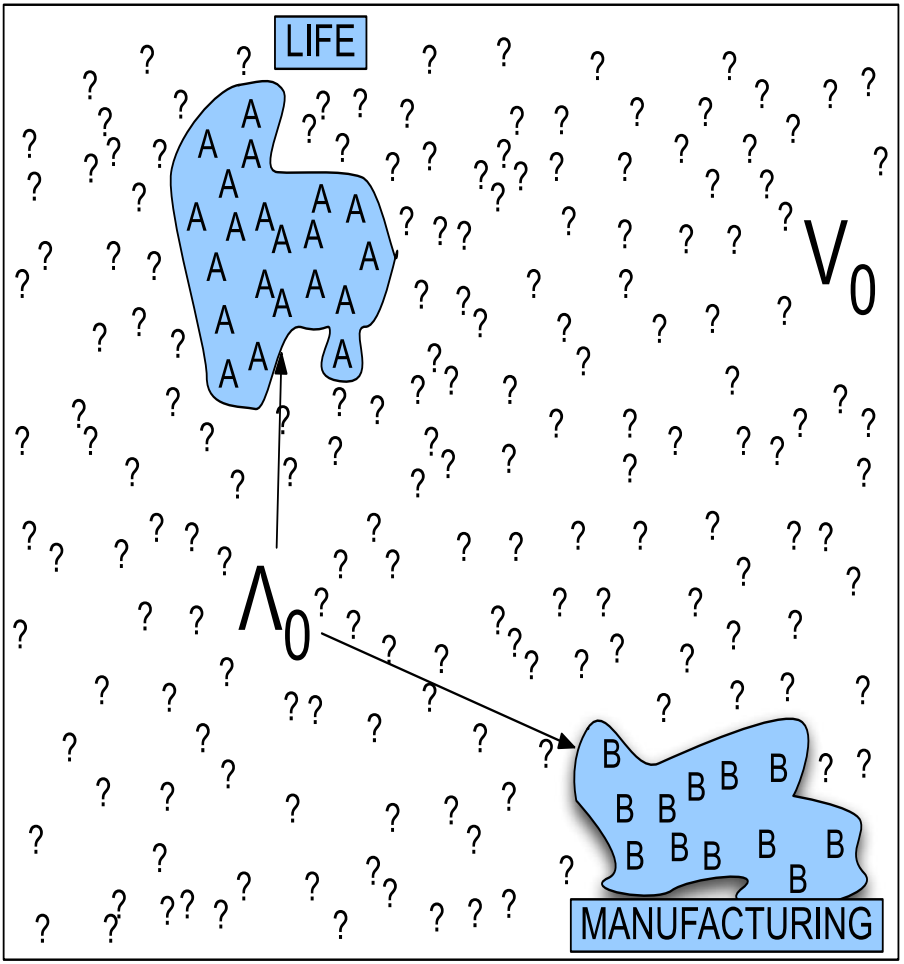
Sentences extracting using ***fish*** and ***play***

We need more good teachers – right now, there are only a half a dozen who can **play** the free **bass** with ease. An electric guitar and **bass play**er stand off to one side, not really part of the scene, just as a sort of nod to gringo expectations perhaps.
The researchers said the worms spend part of their life cycle in such **fish** as Pacific salmon and striped **bass** and Pacific rockfish or snapper. And it all started when **fish**ermen decided the striped **bass** in Lake Mead were too skinny.

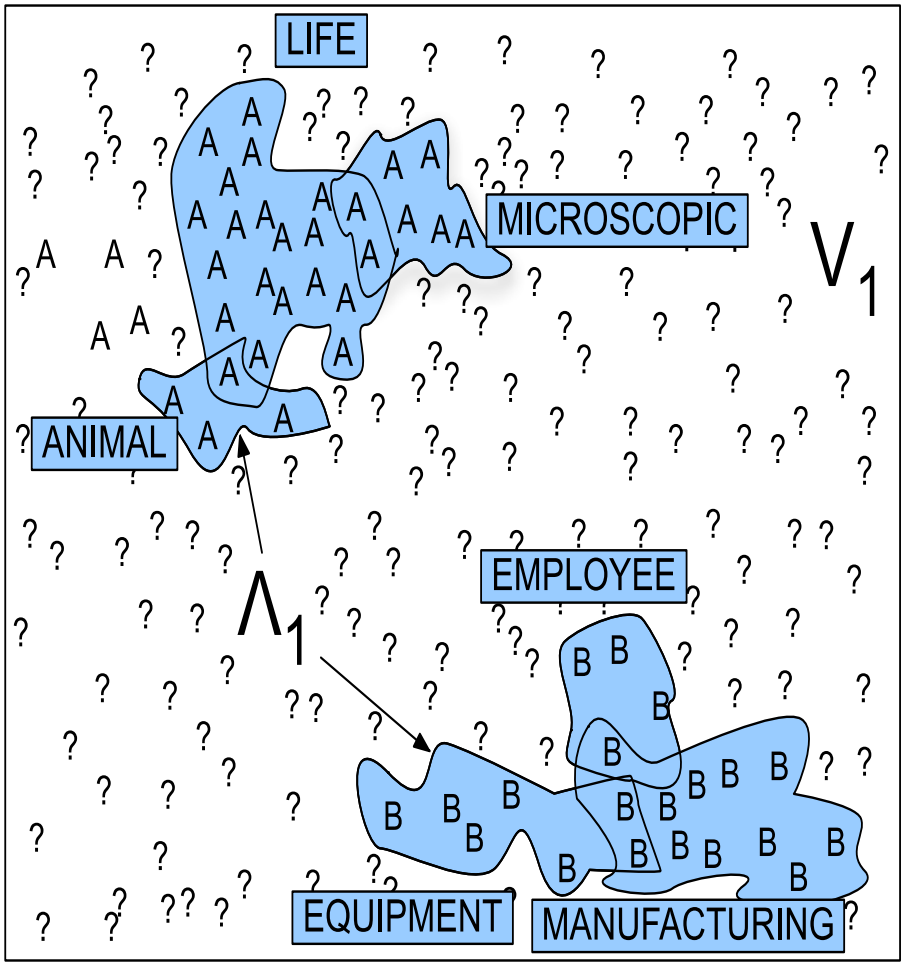
Summary: generating seeds

1) Hand labeling
2) "One sense per collocation":
   - A word reoccurring in collocation with the same word will almost surely have the same sense.
3) "One sense per discourse":
   - The sense of a word is highly consistent within a document - Yarowsky (1995)
   - (At least for non-function words, and especially topic-specific words)

Stages in the Yarowsky bootstrapping algorithm for the word “plant”



(a)



(b)

SenseEval

- Standardized international “competition” on WSD.
- Organized by the Association for Computational Linguistics (ACL) Special Interest Group on the Lexicon (SIGLEX).
- After 2007, evolved in broader “SemEval” competition.

Senseval 1: 1998

- Datasets for
  - English
  - French
  - Italian
- Lexical sample in English
  - Noun: accident, behavior, bet, disability, excess, float, giant, knee, onion, promise, rabbit, sack, scrap, shirt, steering
  - Verb: amaze, bet, bother, bury, calculate, consumer, derive, float, invade, promise, sack, scrap, sieze
  - Adjective: brilliant, deaf, floating, generous, giant, modest, slight, wooden
  - Indeterminate: band, bitter, hurdle, sanction, shake
- Total number of ambiguous English words tagged: 8,448

Senseval 1 English Sense Inventory

- Senses from the HECTOR lexicography project.
- Multiple levels of granularity
  - Coarse grained (avg. 7.2 senses per word)
  - Fine grained (avg. 10.4 senses per word)

Senseval Metrics

- Fixed training and test sets, same for each system.
- System can decline to provide a sense tag for a word if it is sufficiently uncertain.
- Measured quantities:
  - A: number of words assigned senses
  - C: number of words assigned correct senses
  - T: total number of test words
- Metrics:
  - Precision = C/A
  - Recall = C/T

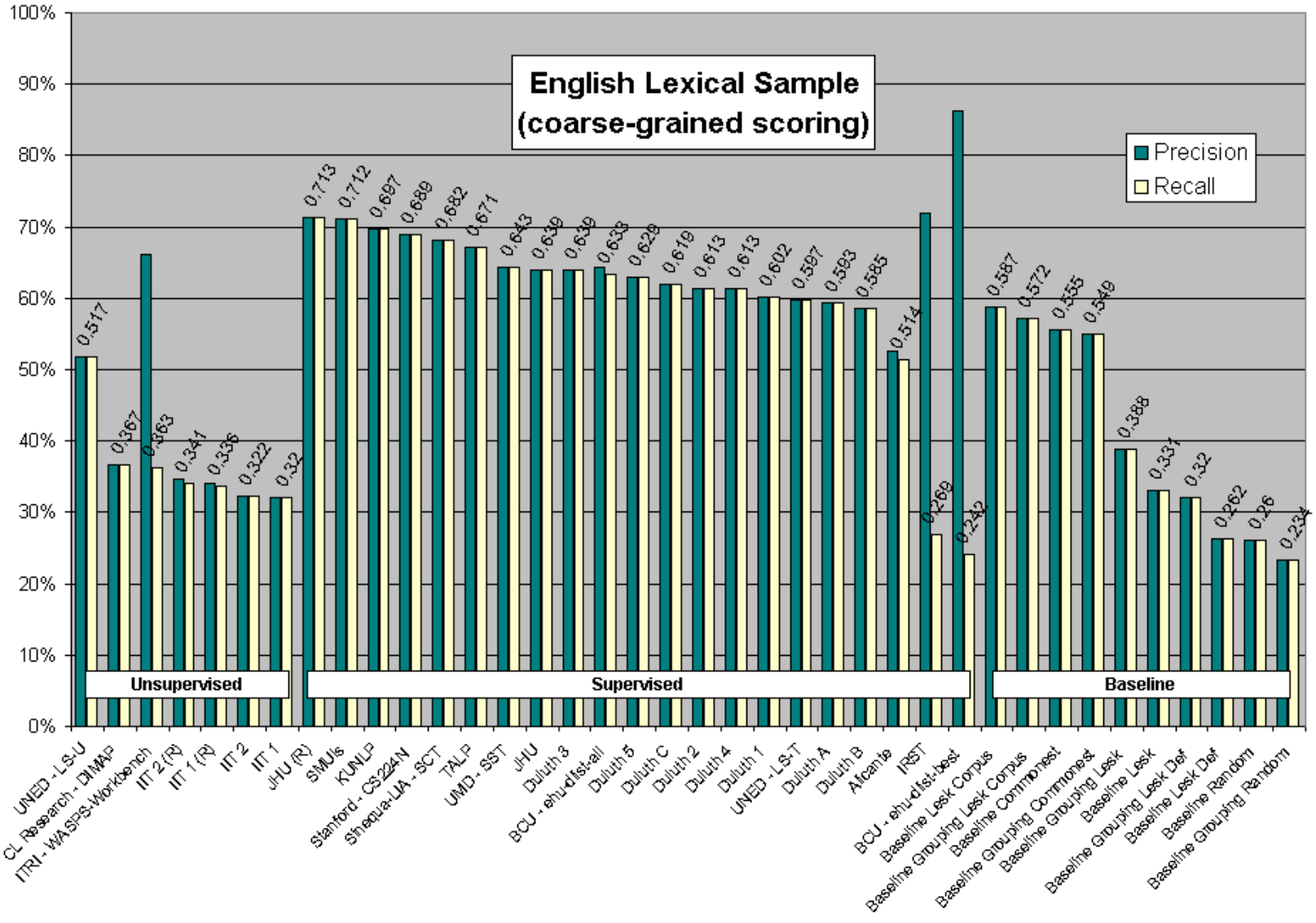
Senseval 1 Overall English Results

	Fine grained precision (recall)	Course grained precision (recall)
Human Lexicographer Agreement	97% (96%)	97% (97%)
Most common sense baseline	57% (50%)	63% (56%)
Best system	77% (77%)	81% (81%)

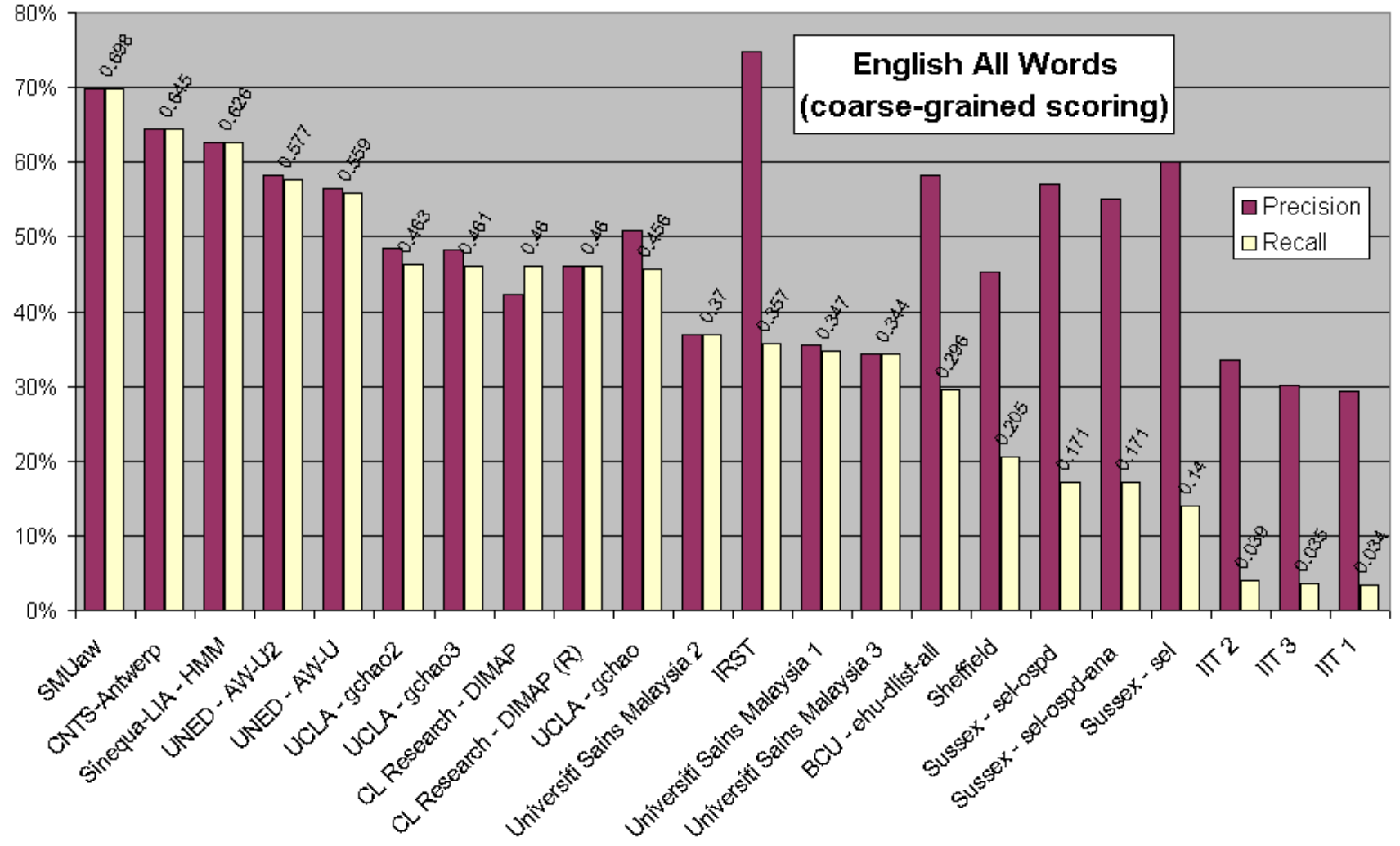
Senseval 2: 2001

- More languages: Chinese, Danish, Dutch, Czech, Basque, Estonian, Italian, Korean, Spanish, Swedish, Japanese, English
- Includes an “all-words” task as well as lexical sample.
- Includes a “translation” task for Japanese, where senses correspond to distinct translations of a word into another language.
- 35 teams competed with over 90 systems entered.

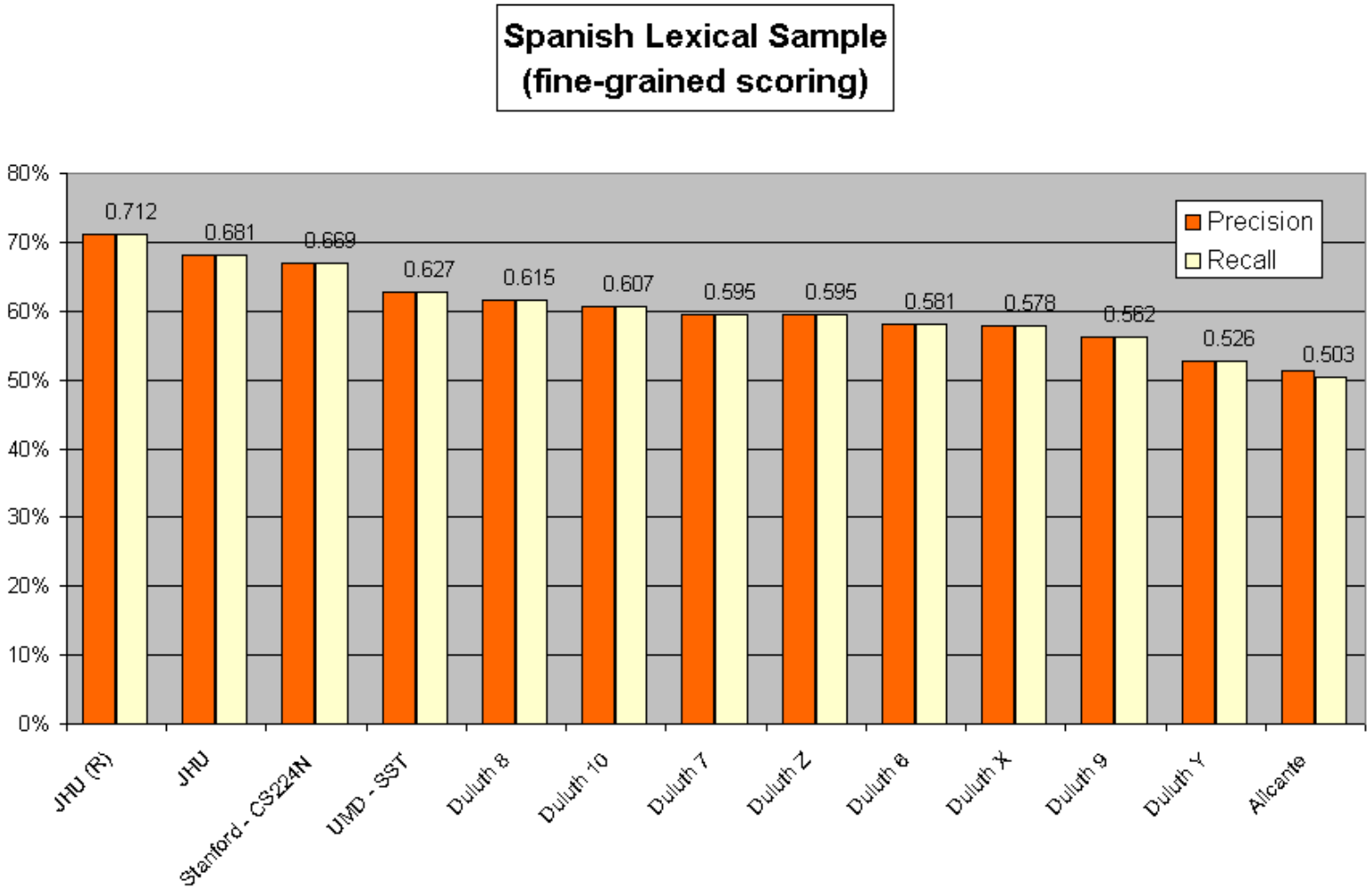
Senseval 2 Results



Senseval 2 Results



Senseval 2 Results



Issues in WSD

- What is the right granularity of a sense inventory?
- Integrating WSD with other NLP tasks
  - Syntactic parsing
  - Semantic role labeling
  - Semantic parsing
- Does WSD actually improve performance on some real end-user task?
  - Information retrieval
  - Information extraction
  - Machine translation
  - Question answering

Data

- Semcor: 200K+ words tagged with Wordnet senses.

http://www.cse.unt.edu/~rada/downloads.html#semcor

- WordNet

https://wordnet.princeton.edu/wordnet/download/