Natural Language Processing: Word Sense Disambiguation

Based on slides by

Raymond J. Mooney

University of Texas at Austin

Jimmy Lin, University of Maryland

Lexical Ambiguity

- Many words in natural languages have multiple possible meanings.
  - "pen" (noun)
    - The dog is in the pen.
    - The ink is in the pen.
  - "take" (verb)
    - Take one pill every morning.
    - Take the first right past the stoplight.
- Context greatly helps disambiguation
- Syntax helps distinguish meanings for different parts of speech of an ambiguous word.
  - "conduct" (noun or verb)
    - John's conduct in class is unacceptable.
    - John will conduct the orchestra on Thursday.

Motivation for Word Sense Disambiguation (WSD)

- Many tasks in natural language processing require disambiguation of ambiguous words.
  - Question Answering
  - Information Retrieval
  - Machine Translation
  - Text Mining
  - Phone Help Systems
- Understanding how people disambiguate words is an interesting problem that can provide insight in psycholinguistics.

Sense Inventory

- What is a “sense” of a word?
  - Homonyms (disconnected meanings)
    - bank: financial institution
    - bank: sloping land next to a river
  - Polysemes (related meanings with joint etymology)
    - bank: financial institution as corporation
    - bank: a building housing such an institution
- Sources of sense inventories
  - Dictionaries
  - Lexical databases (WordNet)

  Entity/concept disambiguation in Wikipedia!

  Gloss: concise description of word sense (Human)

WordNet-English and Arabic

- A detailed database of semantic relationships between words (English/Arabic).
- Developed by famous cognitive psychologist George Miller and a team at Princeton University.
- About 155,287/23481 English/Arabic words (11/19/2017).
- Nouns, adjectives, verbs, and adverbs grouped into about 117,659/11,269 synonym sets called *synsets*.
- Each expressing a distinct concept.
- Synsets are interlinked by means of conceptual-semantic and lexical relations

WordNet Synset Relationships

- Antonym: front → back
- Attribute: benevolence → good (noun to adjective)
- Pertainym: alphabetical → alphabet (adjective to noun)
- Similar: unquestioning → absolute
- Cause: kill → die
- Entailment: breathe → inhale
- Holonym: chapter → text (part to whole)
- Meronym: computer → cpu (whole to part)
- Hyponym: plant → tree (specialization)
- Hypernym: apple → fruit (generalization)

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 for people and computers. Only expert lexicographers are perhaps able to reliably differentiate senses.
- 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).

WDS from WNet

Noun

- {pipe, tobacco pipe} (a tube with a small bowl at one end; used for smoking tobacco)
- {pipe, pipage, piping} (a long tube made of metal or plastic that is used to carry water or oil or gas etc.)
- {pipe, tube} (a hollow cylindrical shape)
- {pipe} (a tubular wind instrument)
- {organ pipe, pipe, pipework} (the flues and stops on a pipe organ)

Verb

- {shriek, shrill, pipe up, pipe} (utter a shrill cry)
- {pipe} (transport by pipeline) “pipe oil, water, and gas into the desert”
- {pipe} (play on a pipe) “pipe a tune”

- **Noun**
- {كرم} (Generosity)
- {كرم} (Grapevine)
- {كرم} (Masculine name)

- **Verb**
- {كرم} (Honor)
- {كرم} (Made an act of generosity)

Senses Based on Needs of Translation

- Only distinguish senses that are translate to different words in some other language.
  - play: مسرحية vs. يلعب
  - :bank ضفة vs. مصرف
  - leave: يغادر vs اوراق
  - take: ياخذvs. مرة
  - ذهب gold vs past of Go
  - غادر Left vs traitorous
- May still require overly fine-grained senses
  - river in French is either:
    - fleuve: flows into the ocean
    - rivière: does not flow into the ocean

How big is the problem?

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

Baseline + Upper Bound

- Baseline: most frequent sense
  - Equivalent to "take first sense" in WordNet
  - Does surprisingly well!

62% accuracy in this case!

Freq	Synset	Gloss
338	plant[1], works, industrial plant	buildings for carrying on industrial labor
207	plant[2], flora, plant life	a living organism lacking the power of locomotion
2	plant[3]	something planted secretly for discovery by another
0	plant[4]	an actor situated in the audience whose acting is rehearsed but seems spontaneous to the audience

- Upper bound:
  - Fine-grained WordNet sense: 75-80% human agreement
  - Coarser-grained inventories: 90% human agreement possible
- What does this mean?

WSD Approaches

- Depending on use of manually created knowledge sources
  - Knowledge-lean
  - Knowledge-rich
- Depending on use of labeled data
  - Supervised
  - Semi- or minimally supervised
  - Unsupervised

Lesk's Algorithm

- Intuition: note word overlap between context and dictionary entries (glosses)
  - Unsupervised, but knowledge rich

bank can guarantee deposits will eventually cover future tuition costs because it invests in adjustable-rate mortgage securities.

bank[1]	Gloss:	a financial institution that accepts deposits and channels the money into lending activities
	Examples:	"he cashed a check at the bank", "that bank holds the mortgage on my home"
bank[2]	Gloss:	sloping land (especially the slope beside a body of water)
	Examples:	"they pulled the canoe up on the bank", "he sat on the bank of the river and watched the currents"

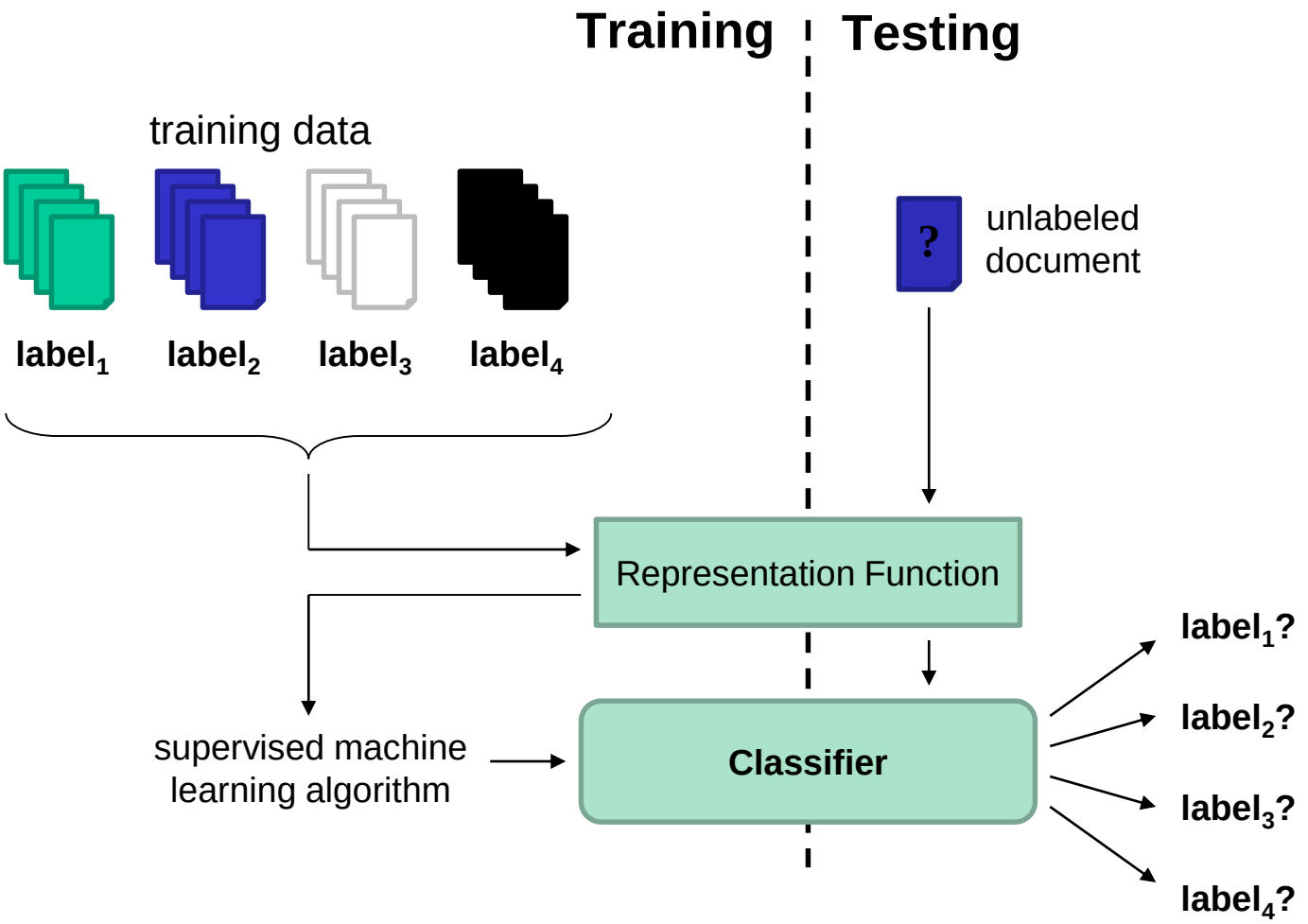
Lesk's Algorithm

- Simplest implementation:
  - Count overlapping content words between glosses and context
- Lots of variants:
  - Include the examples in dictionary definitions
  - Include hypernyms and hyponyms
  - Give more weight to larger overlaps (e.g., bigrams)
  - Give extra weight to infrequent words (e.g., *idf* weighting)
  - …
- Works reasonably well!

Supervised WSD: NLP meets ML

- WSD as a supervised classification task
  - Train a separate classifier for each word
- Three components of a machine learning problem:
  - Training data (corpora)
  - Representations (features)
  - Learning method (algorithm, model)

Supervised Classification



Care with Machine Learning

- Thou shalt not mingle training data with test data
- Have user annotated data: careful with your own
- Be objective!

Features

- Possible features
  - POS (Part Of Speech) and surface form of the word itself
  - Surrounding words and POS tag
  - Positional information of surrounding words and POS tags
  - Same as above, but with *n*-grams
  - Grammatical information
  - …
- Richness of the features?
  - Richer features = ML algorithm does less of the work
  - More impoverished features = ML algorithm does more of the work

Classifiers

- Once we cast the WSD problem as supervised classification, many learning techniques are possible:
  - Naïve Bayes (the thing to try first)
  - Decision trees
  - MaxEnt
  - Support vector machines
  - Nearest neighbor methods
  - …

Classifiers Tradeoffs

- Which classifier should I use?
- It depends:
  - Number of features
  - Types of features
  - Number of possible values for a feature
  - Noise,…..
- General advice:
  - Start with Naïve Bayes
  - Use decision trees/lists if you want to understand what the classifier is doing
  - SVMs often give state of the art performance

Learning for WSD

- Assume part-of-speech (POS), e.g. noun, verb, adjective, for the target word is determined:

ذهب being a verb may solve the WSD problem!

- 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 (same as do while testing).

Baseline + Upper Bound

- Baseline: most frequent sense
  - Equivalent to "take first sense" in WordNet
  - Does surprisingly well!

Freq	Synset	Gloss
338	plant[1], works, industrial plant	buildings for carrying on industrial labor
207	plant[2], flora, plant life	a living organism lacking the power of locomotion
2	plant[3]	something planted secretly for discovery by another
0	plant[4]	an actor situated in the audience whose acting is rehearsed but seems spontaneous to the audience

62% accuracy in this case!

- Upper bound:
  - Fine-grained WordNet sense: 75-80% human agreement
  - Coarser-grained inventories: 90% human agreement possible

Feature Engineering

- The success of machine learning requires instances to be represented using an effective set of features that are correlated with the categories of interest.
- Feature engineering can be a laborious process that requires substantial human expertise and knowledge of the domain.
- In NLP it is common to extract many (even thousands of) potential features and use a learning algorithm that works well with many relevant and irrelevant features.

Contextual Features

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

Experimental evaluations indicate that all of these features are useful; and the best results comes from integrating all of these cues in the disambiguation process.

This is for English. What about Arabic?

Surrounding Bag of Words

- Unordered individual words near the ambiguous word. E.g. Words in the same sentence.
- May include words in the previous sentence or surrounding paragraph (how far?).
- Gives general topical cues of the context.
- May use feature selection to determine a smaller set of words that help discriminate possible senses.
- May just remove common “stop words” such as articles, prepositions, etc.
- If have parallel data (English/Arabic): can use that!

POS of Neighboring Words

- Use part-of-speech 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**" [2 before to 1 after, No word]
- 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 [have POS!], it is very useful to know its direct object [play: instrument/game?].
  - "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)

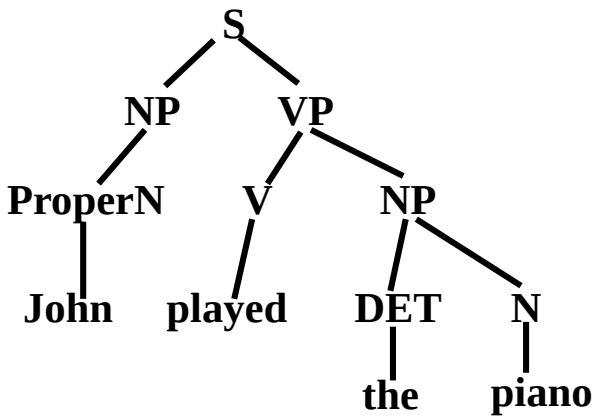
- For an ambiguous noun, it is useful to know what verb it is an object of:
  - "**played** the piano and the horn"
  - "**wounded** by the rhinoceros' horn"
- May also be useful to know what verb it is the subject of:
  - "the bank near the river **loaned** him $100"
  - "the bank is **eroding** and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

- "In vitro":
  - 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.
- "In vivo":
  - Incorporate WSD system into some larger application system, such as machine translation, information retrieval, or question answering.
  - Evaluate relative contribution of different WSD methods by measuring performance impact on the overall system on final task (accuracy of MT, IR, or QA results).

Lexical Sample vs. All Word Tagging

- Lexical sample:
  - Choose one or more ambiguous words each with a sense inventory.
  - From a larger corpus, assemble sample occurrences of these words.
  - Have humans mark each occurrence with a sense tag.
- All words:
  - Select a corpus of sentences.
  - For each ambiguous word in the corpus, have humans mark it with a sense tag from an broad-coverage lexical database (e.g. WordNet).

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: semantics/meaning.
- Started with word senses, now to semantic role, coreference, smenatic relations and sentiment analysis
- Arabic appeared in Semeval2016 (https://en.wikipedia.org/wiki/SemEval)

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

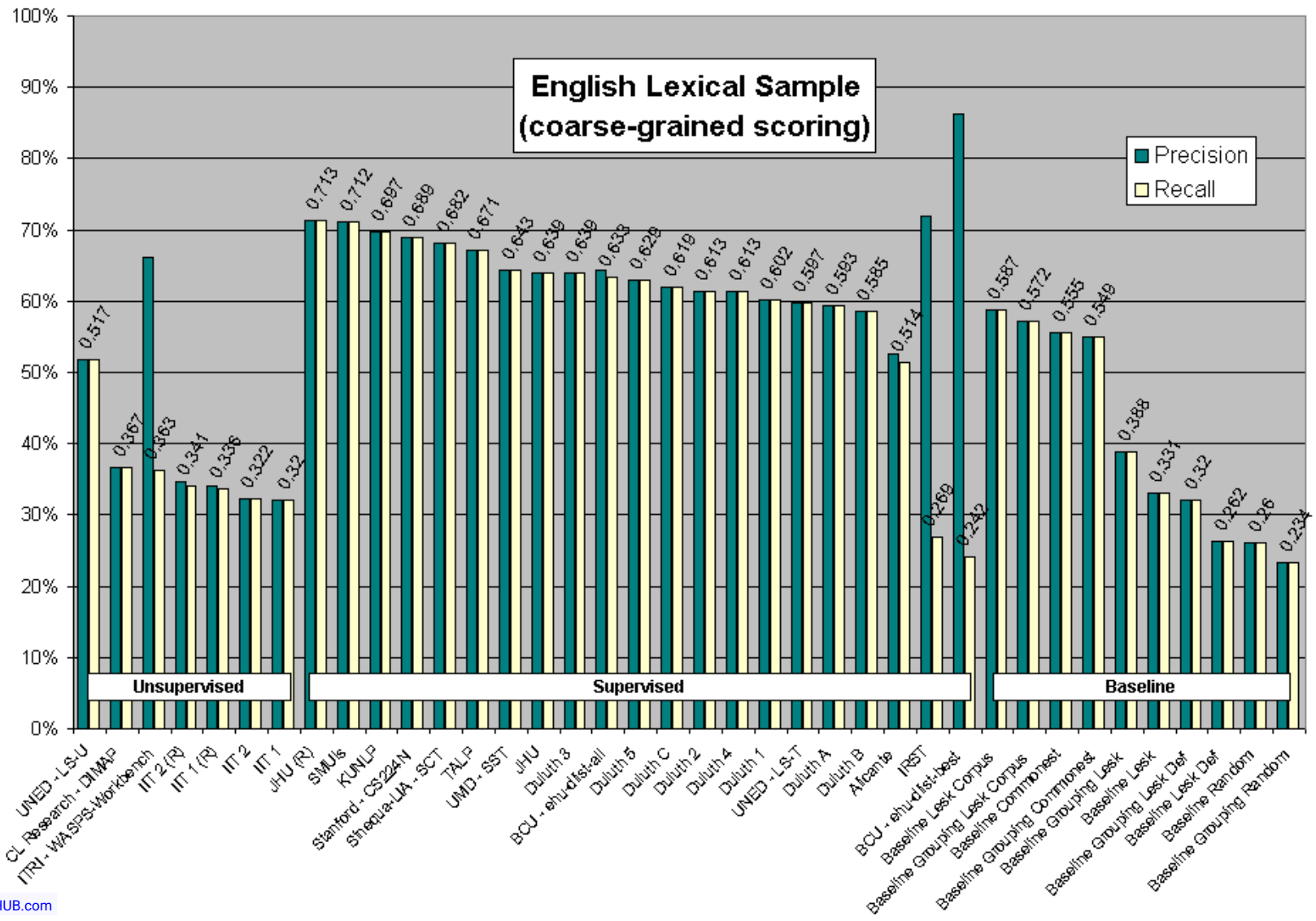
Senseval 1 Overall English Results

	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%)

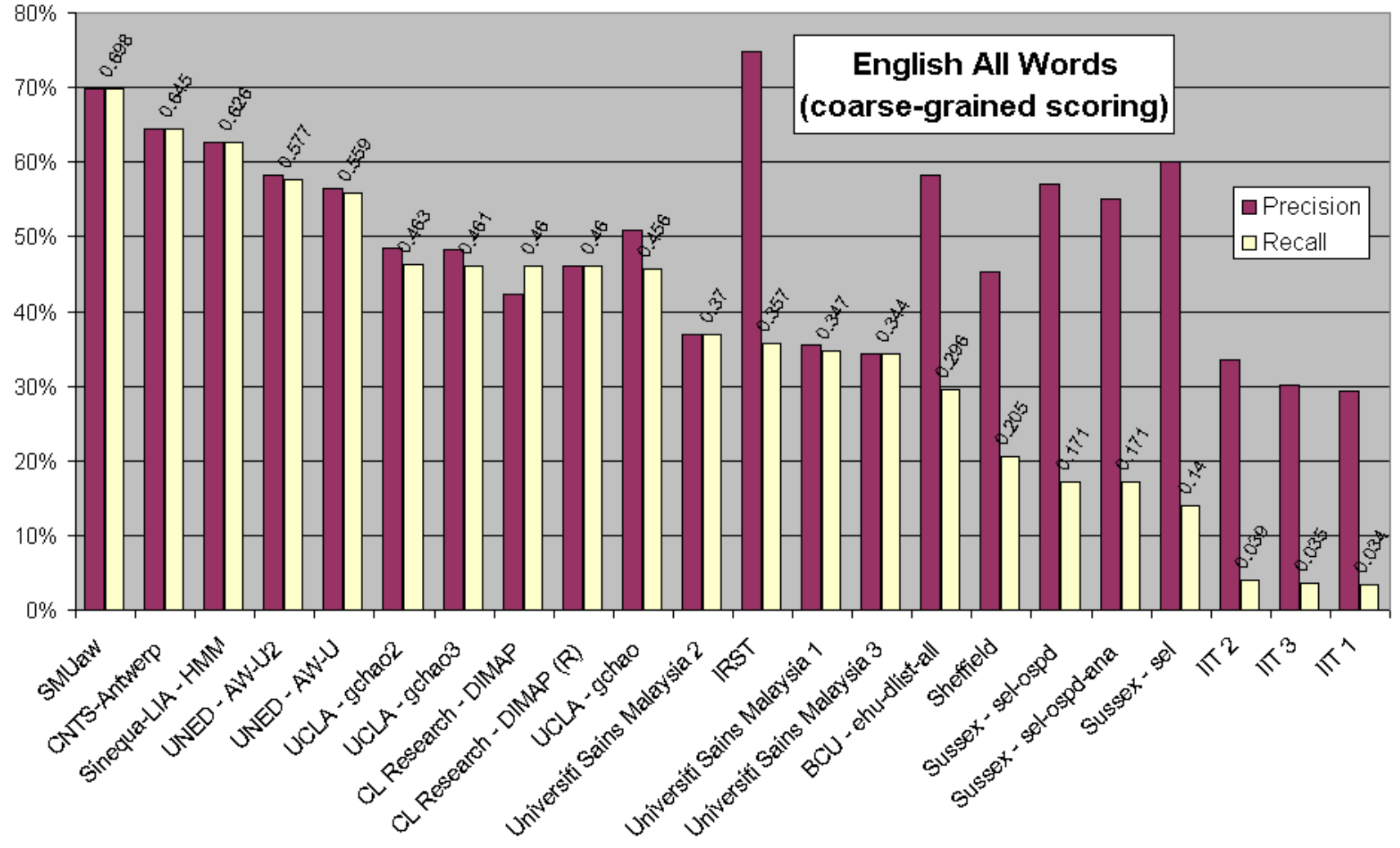
Senseval 2: 2001

- 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.

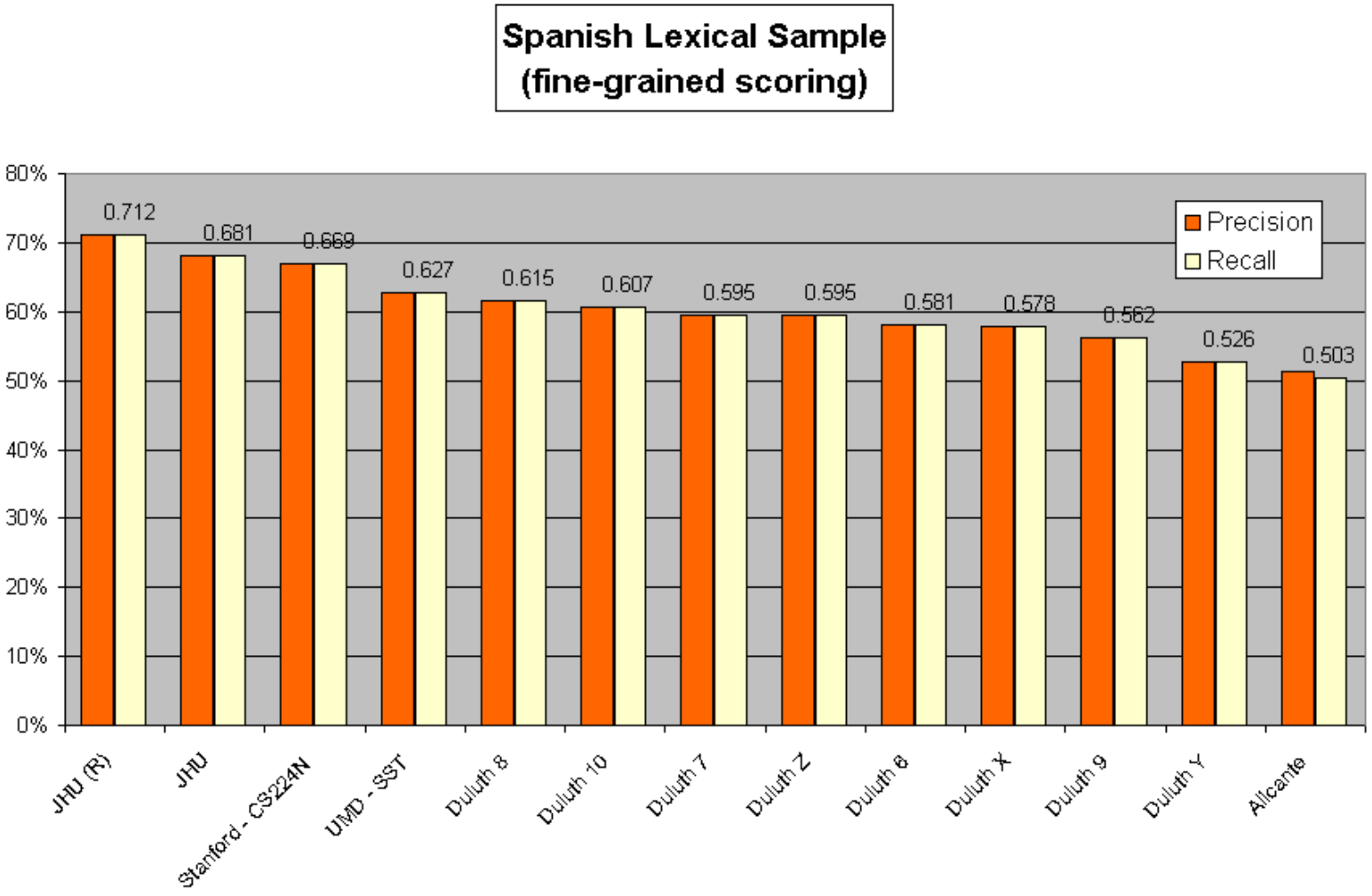
Senseval 2 Results



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