Knowledge-Lean Approaches to Word Sense Disambiguation (2008)

Michael Gabilondo

Project

Wrote reviews for five different papers

Four papers on word sense induction (WSI), also called word sense discrimination (knowledge-free approaches to disambiguation)

One paper on word sense disambiguation (a knowlege-lean approach; the Yarowsky 1995 paper)

Word Sense Disambiguation (WSD)

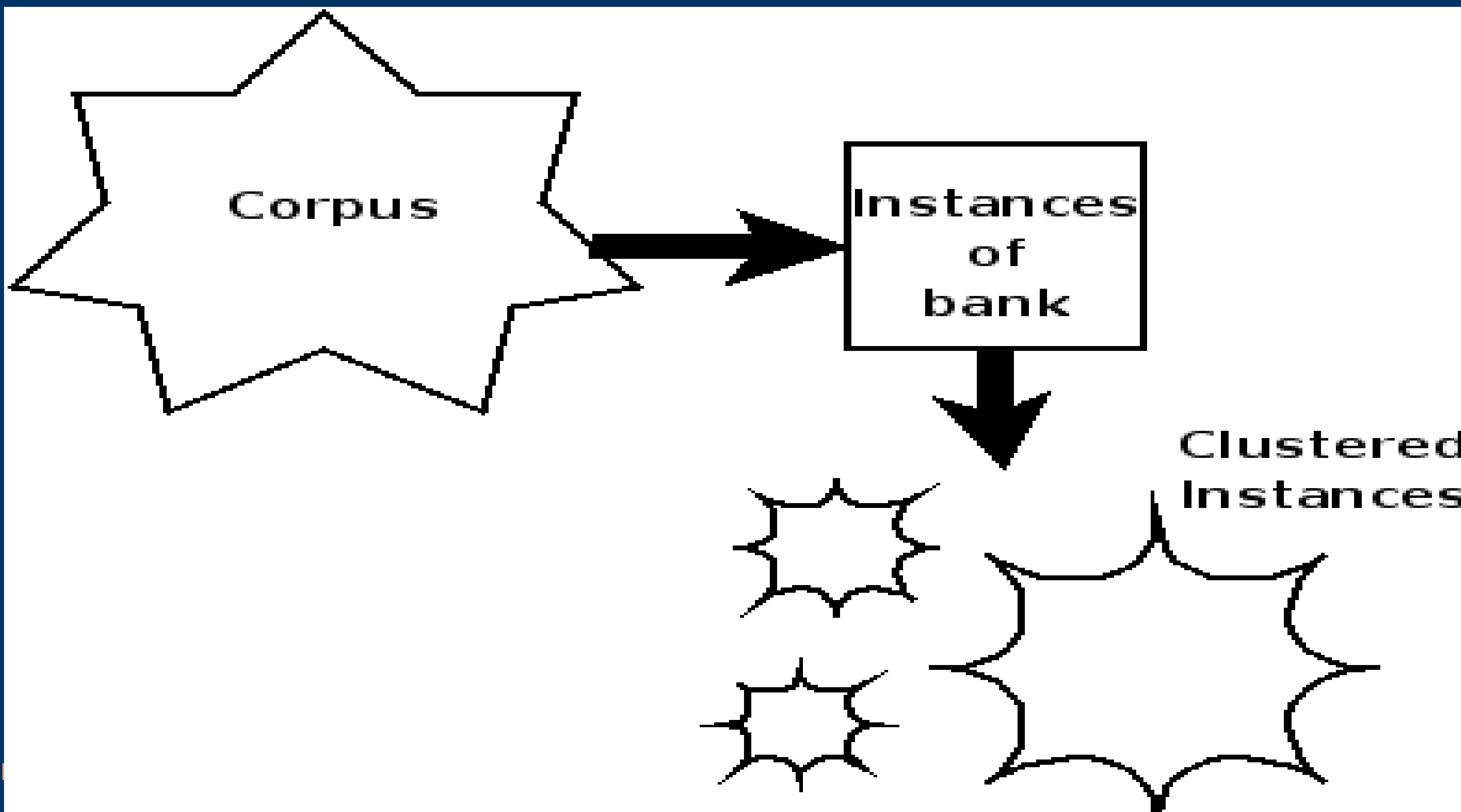
Given a polysemous word instance, classify as one of the dictionary senses

Can infer the correct sense by using the features of the local context

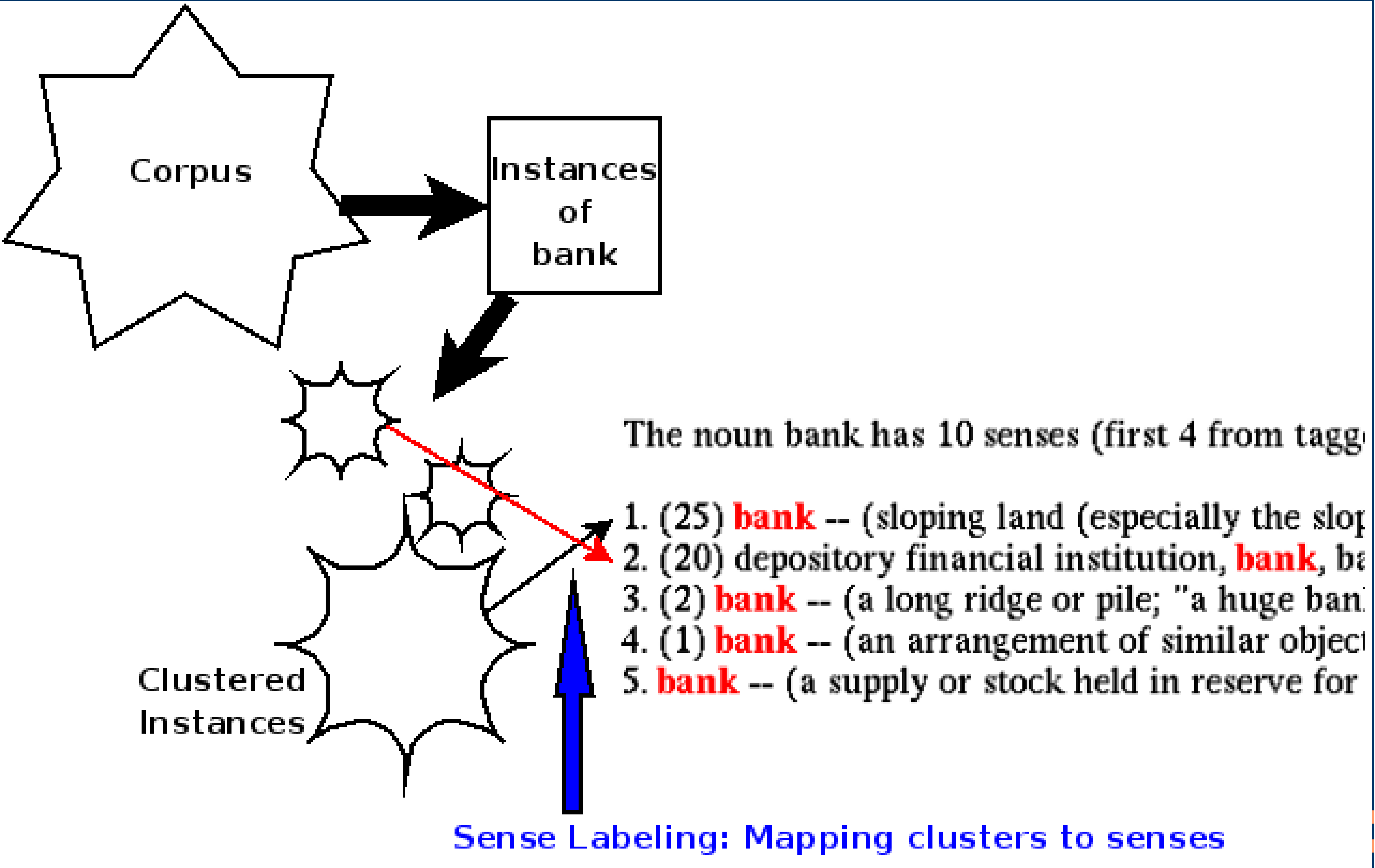
Features can be word cooccurrences, syntactic dependencies, ...

Word Sense Induction (WSI)

WSI or word sense discrimination is the task of generating sense distinctions from a corpus instead of working with a set of dictionary senses --- And then assigning words from a test corpus to the clusters that represent senses



Sense Labeling



Contextual Hypothesis for Senses

Why should any cluster map directly to a dictionary sense?

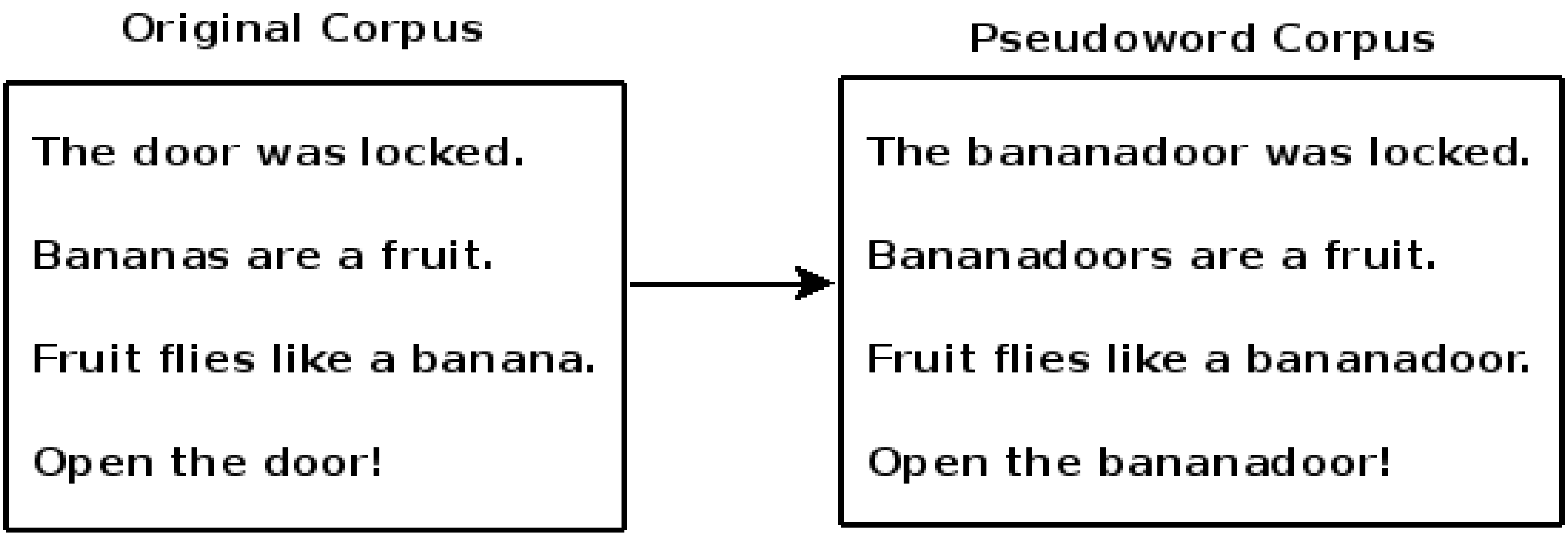
Contextual Hypothesis for Senses

Two occurrences of an ambiguous word belong to the same sense to the extent that their contextual representations are similar.

Evaluation of WSI

- Makes sense clusters without guidance from lexicon sense distinctions
- Evaluation usually tests how well these clusters map to course-grained sense distinctions made by judges
- Evaluation can be automated by using sense-tagged text (tests contextual hypothesis) or by pseudoword evaluation (does not need to map clusters to dictionary senses)

Pseudoword Evaluation



lexicon entry Bananadoor

pseudoword.
sense 1: a door
sense 2: a banana

Disambiguate:* *Bananadoors are a fruit. ***Easy to check correctness with original corpus! ****

**** only train on portion of pseudoword corpus and set some aside for evaluation.***

Schütze (1998)

Context-Group discrimination

Words in the corpus have cooccurrence vectors (word vectors); made up of words they coccur with often

Features of local context are the word vectors of the words in the context (not the words themselves: second order)

Local context represented by centroid of feature vectors, called context vector

Context vectors are clustered; centroids of clusters are called sense vectors, represent senses

Disambiguation: Context vector is created for that instances and gets assigned to cluster with closest sense vector

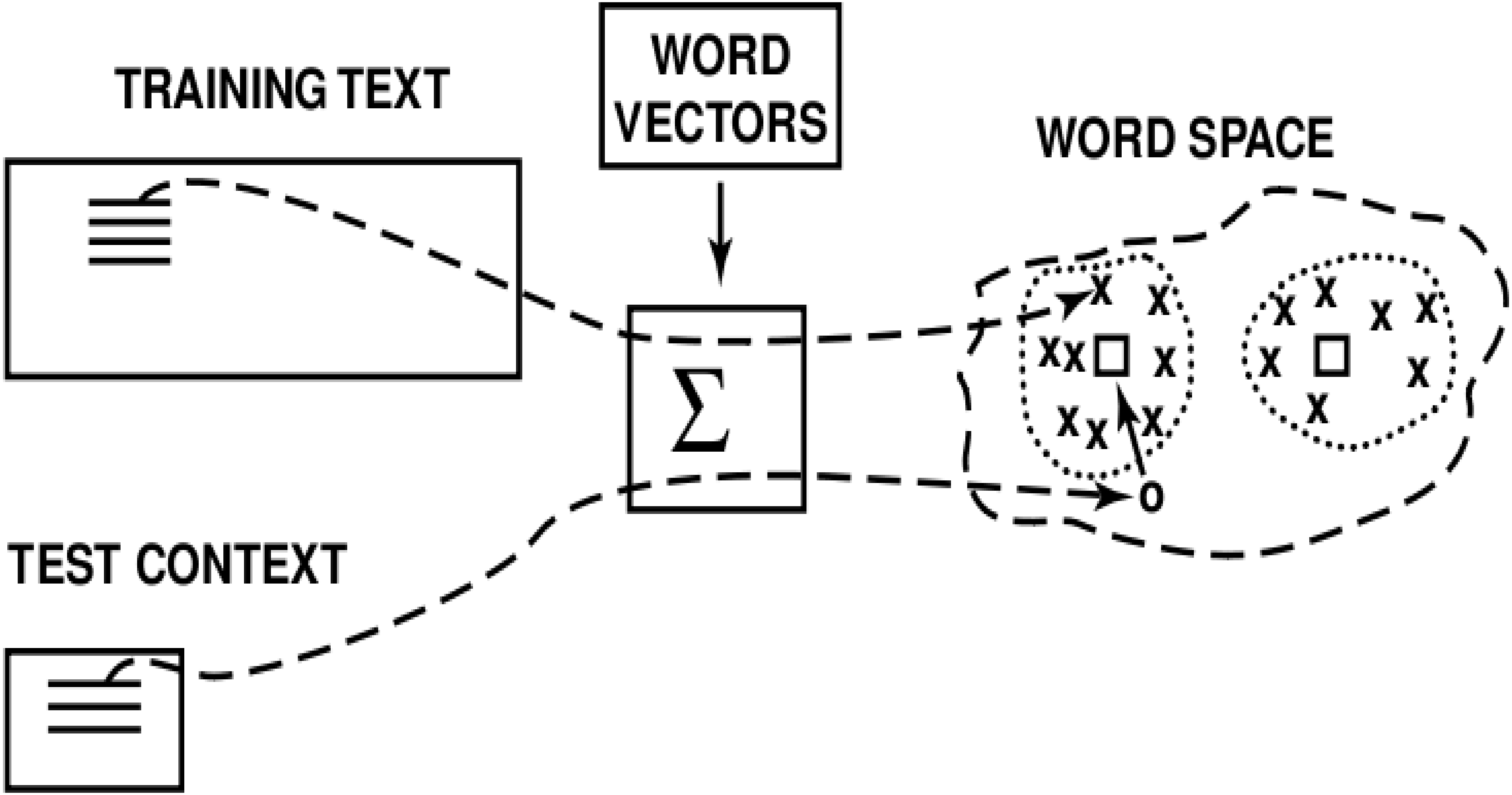


Figure 1

The basic design of context-group discrimination. Contexts of the ambiguous word in the training set are mapped to context vectors in Word Space (upper dashed arrow) by summing the vectors of the words in the context. The context vectors are grouped into clusters (dotted lines) and represented by sense vectors, their centroids (squares). A context of the ambiguous word ("test context") is disambiguated by mapping it to a context vector in Word Space (lower dashed arrow ending in circle). The context is assigned to the sense with the closest sense vector (solid arrow).

Evaluation

No distinctions between parts of speech, always between two senses (train had a verb & noun sense)

10 Natural ambiguous words

- Corpus labeled manually with senses, judged manually

10 Pseudowords

- Pseudoword evaluations

Average 89.7% accuracy using two clusters

Average baseline was 61.2%

In some cases it was below baseline

Application

Application where no mapping to external senses is required
Document-Query similarity
Documents that contain same words as the ones in the query but for different senses are filtered out
In an experiment they summarize by them they claim it improved accuracy

Could query user giving example contexts to choose the intended sense of the search term

Pedersen and Bruce (1998)

Used first order context representations
Many different types of features of local context
Parts of speech, positions of words, morphology (past tense, plural, ...)

Low results
Nouns only group that achieved above baseline, but only around 60% accuracy...

Yarowsky (1995)

(one sense per collocation, discourse)

WSD not WSI since it recieves guidance from the beginning as to what the predefined senses are.

Uses only a small amount of knowlede, so it is knowledge-lean

Knowlege required is small number of seed collocate terms representing each sense; i.e. for plant (living, harvest; manufacturing, cars)

Searches for instances with ”plant” that have a seed term and assigns a sense to plant based on the seed term

Finds more identifying collocate terms from those sentences and repeats the process, classifying more sentences

High precision of about 98.6% for plant and 93.6% for space (around 60% for these words in Schütze (1998))

A little guidance at the beginning goes a long way towards inducing proper sense distinctions

Bordag (2006)

Each content word in the corpus has a list of its 200 most significant cooccurrences

All Triplets of content words in the local context are created, and the intersections of their most-significant-cooccurrences are taken; these intersections are context representations

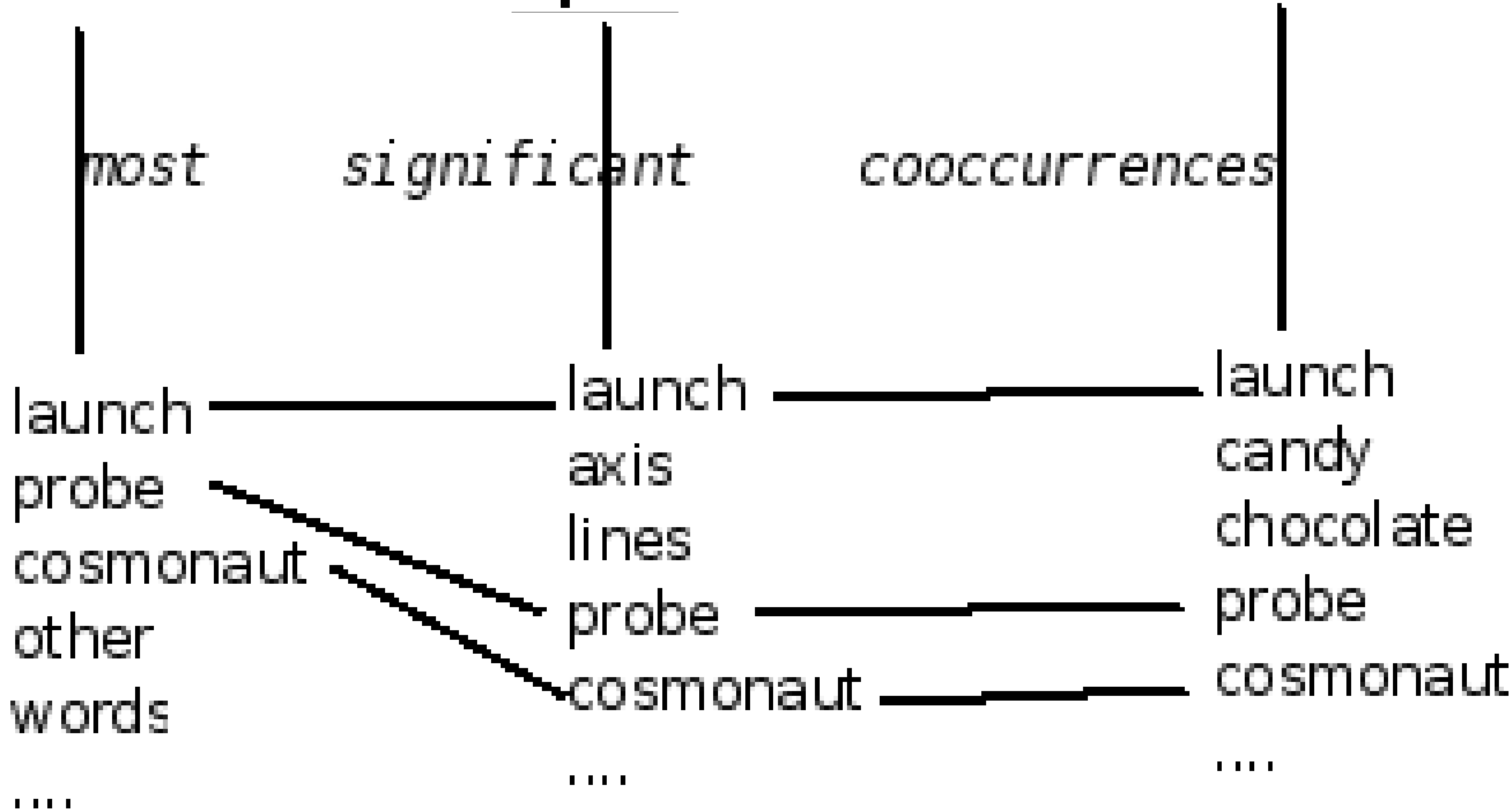
Context representations are clustered

Suggested different types of ambiguity: syntactic and semantic; hypothesized window sized played a role

Used pseudoword evaluation

About 78.66% accuracy

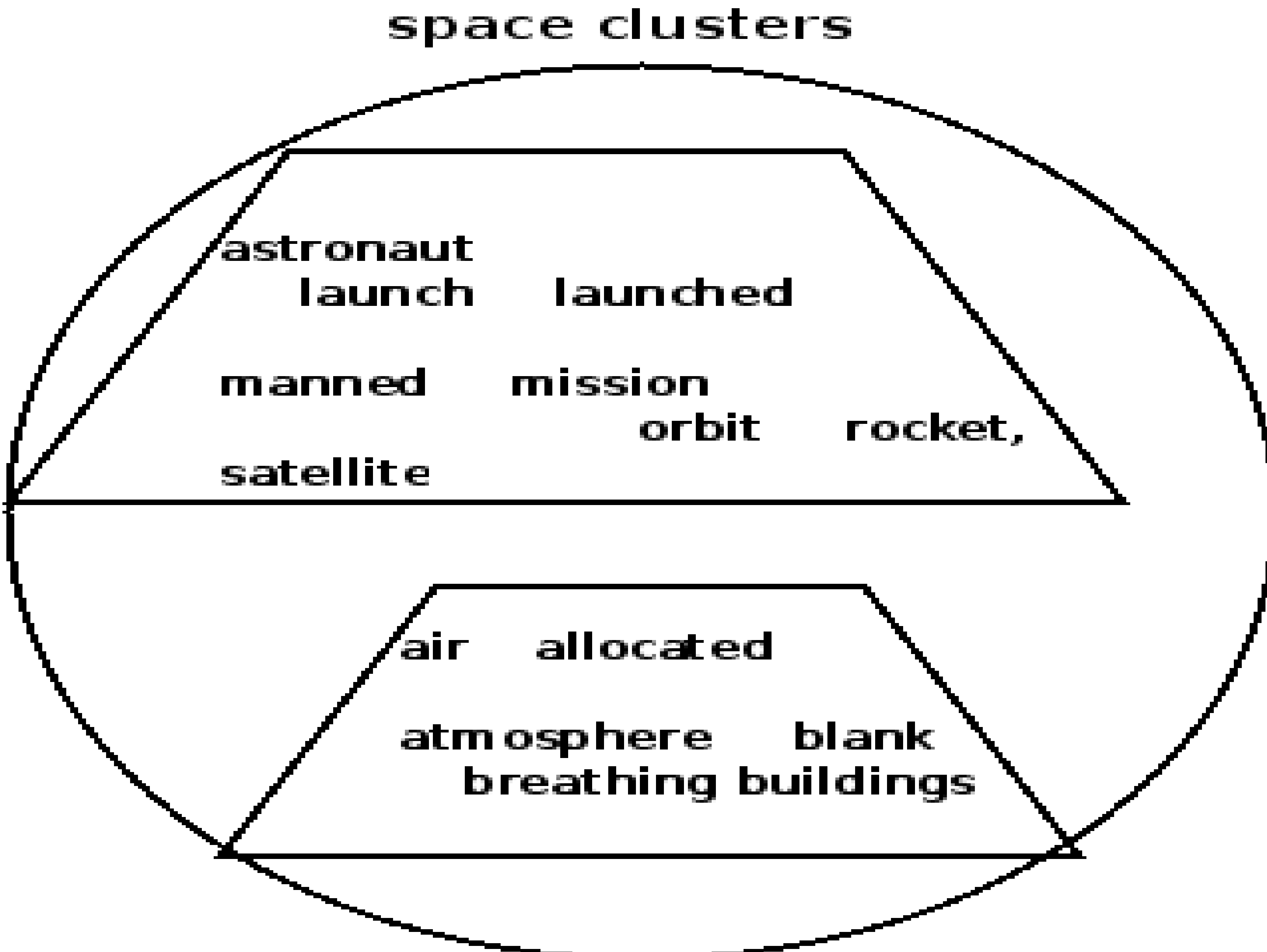
NASA wants a space mission to MARS.



co(NASA) inter.s. co(space) inter.s. co(mars)

=

launch, probe, cosmonaut, ...


space clusters
astronaut
launch launched
manned mission
orbit rocket,
satellite
air allocated
atmosphere blank
breathing buildings

WSI pros and cons - Conclusions

WSI cons

- No guidance from beginning to map to certain dictionary senses (course distinctions may map somewhat, but finer distinctions less so; we saw a little bit of guidance from the beginning as in Yarowsky can be helpful)
- Mapping to dictionary senses is useful for machine translation
- Distinctions made depends on the corpus (neutral?)

WSI pros

- No external lexicon needed
- Could find different senses of highly domain specific terms
- Could still be useful for applications that don't need to work with a particular set of senses (i.e. the IR query-document similarity application)

References

Stefan Bordag. Word sense induction: Triplet-based clustering and automatic evaluation, 2006. URL `citeseer.ist.psu.edu/bordag06word.html`.

Christiane Fellbaum, editor. *WordNet: An Electronic Lexical Database*. The MIT Press, Cambridge, Massachusetts, 1998.

T. Pedersen and R. Bruce. Knowledge lean word sense disambiguation. In *Proceedings of the Fifteenth National Conference on Artificial Intelligence*, pages 800–805, Madison, WI, July 1998. URL `citeseer.ist.psu.edu/article/pedersen98knowledge.html`.

Ted Pedersen and Rebecca Bruce. Distinguishing word senses in untagged text. In Claire Cardie and Ralph Weischedel, editors, *Proceedings of the Second Conference on Empirical Methods in Natural Language Processing*, pages 197–207. Association for Computational Linguistics, Somerset, New Jersey, 1997. URL `citeseer.ist.psu.edu/552289.html`.

Amruta Purandare and Ted Pedersen. Word sense discrimination by clustering contexts in vector and similarity spaces. In *Proceedings of CoNLL-2004*, pages 41–48, 2004.

Hinrich Schütze. Automatic word sense discrimination. *Computational Linguistics*, 24(1):97–124, 1998. URL `citeseer.ist.psu.edu/article/schutze98automatic.html`.

Hinrich Schütze and Jan O. Pedersen. Information retrieval based on word senses, 1995. URL `citeseer.ist.psu.edu/sch95information.html`.

David Yarowsky. Unsupervised word sense disambiguation rivaling supervised methods. In *Meeting of the Association for Computational Linguistics*, pages 189–196, 1995. URL `citeseer.ist.psu.edu/yarowsky95unsupervised.html`.