

Introduction to Corpus-based Computational Semantics: *Word Senses*

Alexander Koller
Columbia University

Sabine Schulte im Walde
University of Stuttgart

Introductory Course
ESSLLI 2007
Trinity College, Dublin, Ireland
August 6-10, 2007

Word Senses



Lexical Semantics and Word Senses

- **Lexical semantics** is the study of how and what the words of a language denote.
- Lexical semantics involves the meaning of each individual word.
- A **word sense** is one of the meanings of a word.
- A word is called **ambiguous** if it can be interpreted in more than one way, i.e., if it has multiple senses.
- **Disambiguation** determines a specific sense of an ambiguous word.

Homonymy and Polysemy

- A **homonym** is a word with multiple, unrelated meanings. A homonym is a word that is spelled and pronounced the same as another but with a different meaning.
 - bank* → financial institution
 - slope of land alongside a river
- A **polyseme** is a word with multiple, related meanings.
 - school* → *He goes to school every day.* (institution)
 - *The school has a blue facade.* (building)
 - *The school is on strike.* (teacher)
- **Regular polysemy** performs a regular induction of a word sense on the basis of another, cf. *school / office*.

Word Sense Ambiguity: Examples

- Homonymy (nouns): *bank, sentence*
- Homonymy (nouns vs. verbs): *saw, watch, rose*
- Polysemy: *concrete floor vs. school floor*
- Polysemy: *pine cone vs. ice cream cone*
- Multiple ambiguous words: *time flies like an arrow*

Metonymy

- Metonymy is substituting a word for another word closely associated with it, or suggested by the original.
- Greek: from *meta*, "change" and *onoma*, "name"
- Examples:
 - » *crown* for *king*:
The power of the crown was mortally weakened.
 - » author for his works:
I'm studying Shakespeare.
 - » noun for noun purpose (Lapata and Lascarides, 2003):
Mary finished the cigarette.

Vagueness

- A word is **vague** to the extent that it has borderline cases.
- Examples: *heap, red, bold, big, delicious*
- Words might be **vague** and **ambiguous**:
 - » *child* is ambiguous between *offspring* and *immature offspring*.
 - » The reading *immature offspring* is vague because there are borderline cases of immature offspring.

Definition and Usage of Word Senses

- Lexicography
- Computational Lexicography
- Corpus Linguistics
- Word Sense Disambiguation (WSD)
- Natural Language Processing (NLP) applications

What goes into the lexicon?

Word Senses in Lexicography

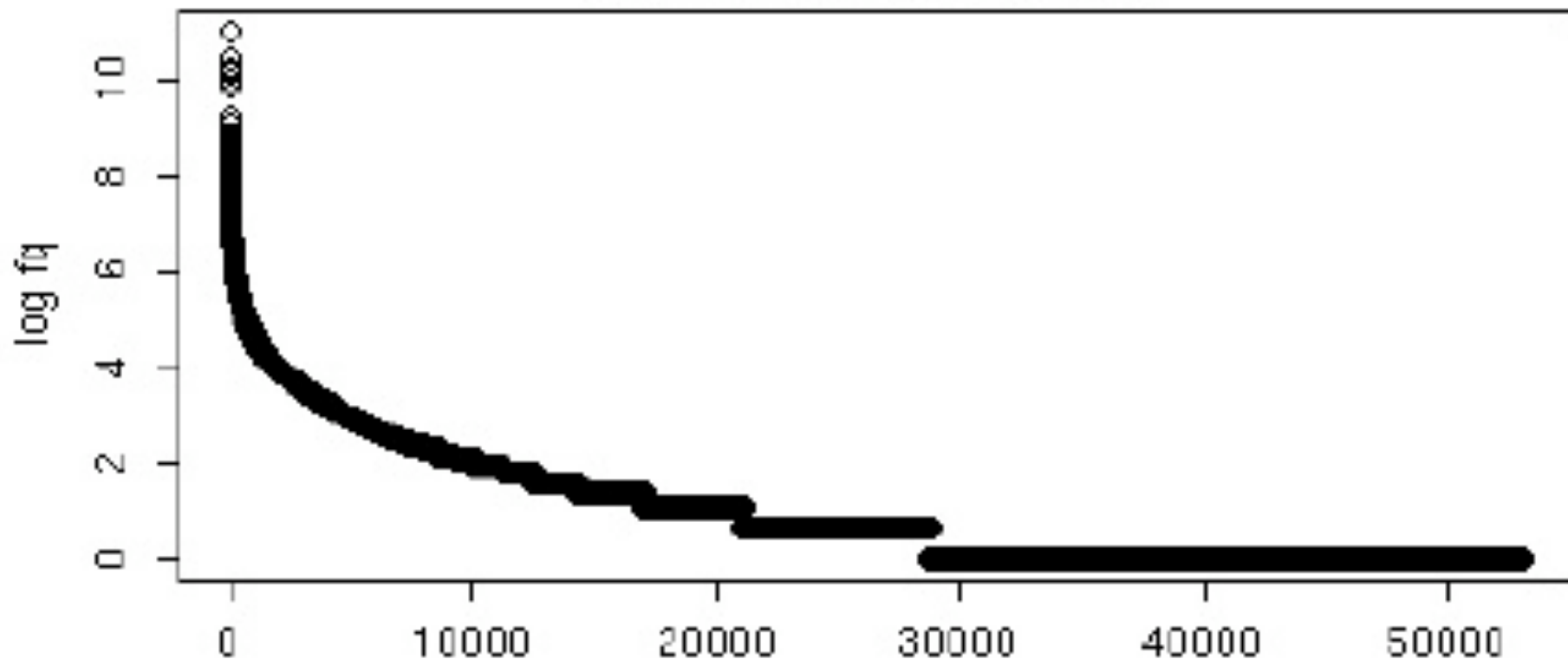
- Any working lexicographer is well aware that they are making **subjective decisions on senses**.
- Any particular dictionary is written with a particular target **audience** in mind, and with a particular **editorial philosophy**.
- Publishers are expected to publish, bookshops are expected to sell, and buyers are expected to buy and use dictionaries. So **policies are not easily changed**.
- Lexicographers write dictionaries rather than writing about writing dictionaries.

Word Senses in Corpus Linguistics

- **Data-driven** rather than theory-driven
- Based on **concordances** of words
- Divide concordances into clusters and work out what it is that makes its members belong together
- Knowledge sources: **frequency** and **predictability** (standard meaning, general linguistic and world knowledge, collocations, etc.)

Word Senses and Frequency (Zipf)

- Frequency as **non-linearly decreasing function of rank**.
 - » corpus frequencies of word types
 - » word sense distributions / pre-dominant sense



Multi-Word Expressions

- **Multiword expressions**: any phrase that is not entirely predictable on the basis of standard rules and lexica
- **Anomalous collocations**: *by and large, on show*
- **Formulae**: *alive and well, as good as gold*
- **Metaphors**: *pack one's bags, kick the bucket*
- **Collocations**: *rancid butter, too ... to ...*

Lexical Semantic Relations

- Relations between word senses
- **Synonymy**: **total** (*Samstag - Sonnabend*) vs. **near/partial** (default case)
- **Hypernymy/Hyponymy**: *is-a* relation (*animal - dog*)
- **Antonymy**: opposition; various types;
 - » **reversives** (*rise - fall*)
 - » **converse** (*buy - sell*)
 - » **contradiction** (*succeed - fail, above - below, (fe)male*)
 - » **co-hyponymy** (*cat - dog, sit - stand, blue - white*)
- **Holonymy/Meronymy**: *part-of* relation (*house - door*)
- *etc.*

WordNet



WordNet – Overview

- Online lexical reference system
- Design inspired by psycholinguistic theories of **human lexical memory**.
- English nouns, verbs, adjectives and adverbs are organised into **synonym sets** (synsets).
- Each synset represents one underlying **lexical concept**.
- Different **(paradigmatic) relations** link the synonym sets.
- WordNet was developed by Princeton University, under the direction of George A. Miller.
- WordNets now exist for many languages.

WordNet Synsets

- Synsets are sets of **synonymous words**.
- Polysemous words appear in multiple synsets.
- Examples:

{coffee, java}

noun example

{coffee, coffee tree}

{coffee bean, coffee berry, coffee}

{chocolate, coffee, deep brown, umber, burnt umber}

{cold}

adjective example

{aloof, cold}

{cold, dry, uncordial}

{cold, unaffectionate, uncaring}

Synset Description

- Synset number (= offset)
- List of words
- Relation pointers to other synsets
- Glosses: *coffee* - a beverage consisting of an infusion of ground coffee beans
- Examples: *coffee* - "he ordered a cup of coffee"
- Subcategorisation frames

WordNet Relations

Within synsets:

- **Synonymy**, such as {coffee, java}

Between synsets / parts of synsets:

- **Antonymy**: opposition, such as {cold} — {hot}
- **Hypernymy / Hyponymy**: is-a relation, such as {coffee, java} — {beverage, drink, potable}
- **Meronymy / Holonymy**: part-of relation, such as {coffee bean, coffee berry, coffee} — {coffee, coffee tree}

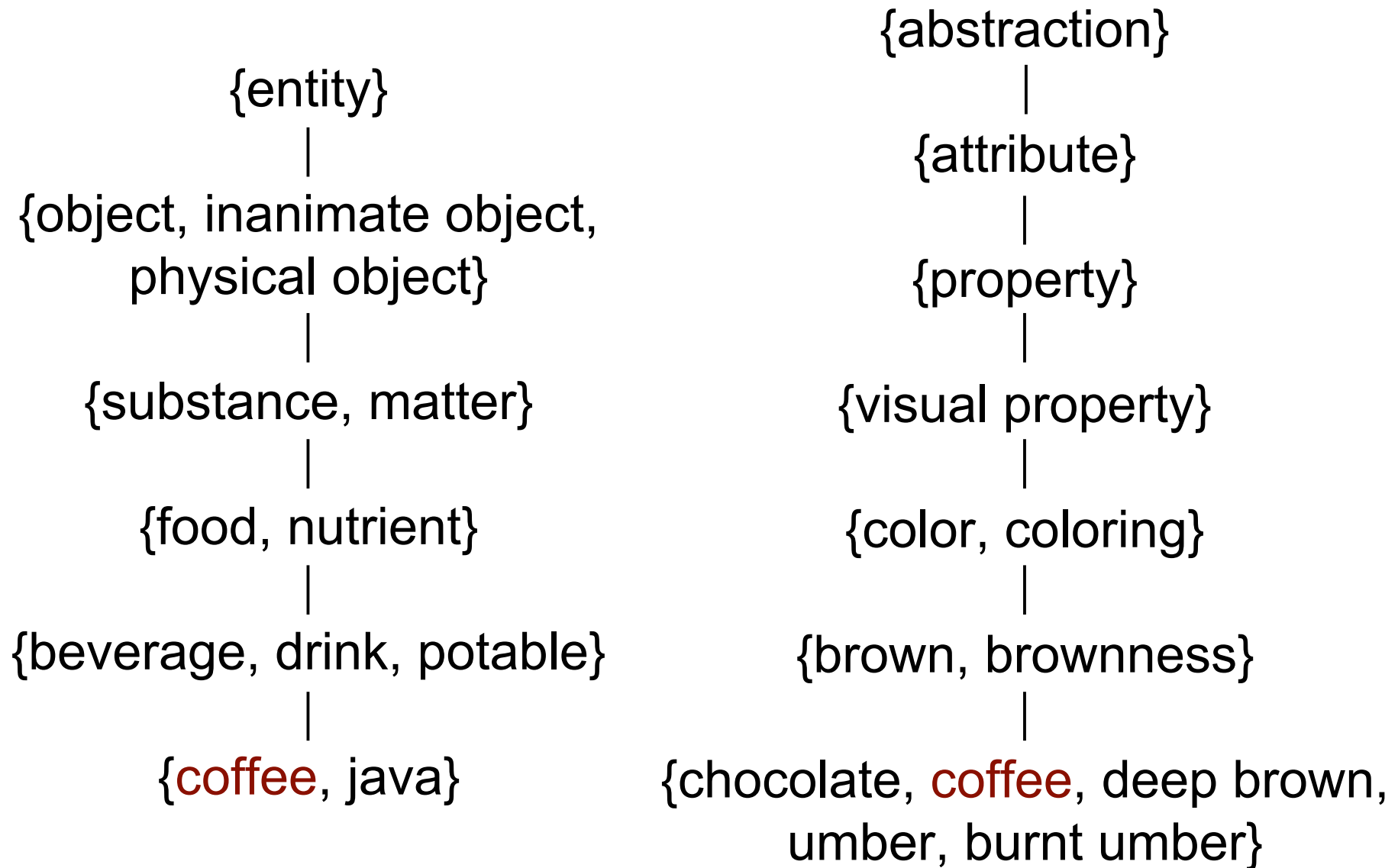
Morphology:

- **Compounds**: arabian coffee, coffee break, coffee table

WordNet Hierarchy

- Depending on the part-of-speech, different relations are defined for a word. For example, the **core relation** for **nouns** is **hypernymy**, the core relation for **adjectives** is **antonymy**.
- Hypernymy imposes a hierarchical structure on the synsets.
- The most general synsets in the hierarchy consists of a number of **pre-defined disjunctive top-level synsets**:
 - nouns → {entity}, {abstraction}, {psychological}, etc.
 - verbs → {move}, {change}, {get}, {feel}, etc.

Wordnet Hierarchy: Example



WordNet Statistics

wnstats - WordNet 3.0 database statistics

DESCRIPTION

Number of words, synsets, and senses

POS	Unique Synsets		Total
	Strings		Word-Sense Pairs
Noun	117097	81426	145104
Verb	11488	13650	24890
Adjective	22141	18877	31302
Adverb	4601	3644	5720
Totals	155327	117597	207016

Polysemy information

POS	Monosemous	Polysemous	
	Words and Senses	Words	Senses
Noun	101321	15776	43783
Verb	6261	5227	18629
Adjective	16889	5252	14413
Adverb	3850	751	1870
Totals	128321	27006	78695

WordNet Family

- Current status: WordNets for 38 languages
- WordNets in the world:
http://www.globalwordnet.org/gwa/wordnet_table.htm
- Integration of WordNets into multi-lingual resources:
 - » EuroWordNet: English, Dutch, Italian, Spanish, German, French, Czech and Estonian
 - » BalkaNet: Bulgarian, Czech, Greek, Romanian, Turkish, Serbian

Word Sense Disambiguation



Word Sense Disambiguation (WSD)

Nancy Ide and Jean Véronis (1998): “Introduction to the special issue on Word Sense Disambiguation: The state of the art”. *Computational Linguistics* 24(1):1-40.

- Automatic disambiguation of word sense
- Intermediate task (Wilks and Stevenson, 1996), which is not an end itself
- Necessary to accomplish NLP tasks: machine translation, information retrieval, etc.

Word Sense Disambiguation: Task

- Association of a given word in a text or discourse with a definition or meaning (sense) which is distinguishable from other meanings potentially attributable to that word
 1. Determination of all different senses
 2. Means to assign each occurrence of a word to the appropriate sense
- Major sources of information:
 - » **context**, together with extra-linguistic information about the text such as situation → data-driven
 - » **external knowledge sources** → knowledge-driven

Word Sense Induction: Sources

- Dictionaries and thesauri
- WordNet
- Automatic, corpus-based; apply heuristics and statistics
- Variation or combination of above

Word Sense Induction: Heuristics

Which factors influence word meaning?

- Context, collocations, syntagmatic embedding
- Discourse
- Domain
- Probability and prototypicality ~ statistics
- World knowledge

Sense Inventory

Adam Killgarriff (1997): “I don’t believe in word senses”. In *Computers and the Humanities* 31: 91-113.

- There is no reason to expect a single set of word senses to be appropriate for different applications.
- Different corpora and different purposes will lead to different senses.
- If we are to know what word senses are, we need operational criteria for distinguishing them.

Sense Granularity

Nancy Ide and Yorick Wilks (2006): “[Making sense about sense](#)”. In Eneko Agirre and Philip Edmonds, editors: *Word Sense Disambiguation: Algorithms and Applications*, chapter 3.

- **Pre-defined, enumerated sense lists** from any source have proven to be **problematic for WSD**.
- Human annotators cannot distinguish well between finer-grained senses in LDOCE (Kilgarriff, 1993).
- Different NLP applications require different degrees of disambiguation (e.g., MT vs. IR).
- SensEval adopted a full and partial scoring scheme, where sub-senses are collapsed.

Sense Granularity

- Example: *She sat on her bicycle and rode away.*
→ 150 senses of bike (= distinct parts) vs. vagueness
- Suggestion: NLP needs a **level of sense-discrimination that corresponds to homographs.**
- Psycholinguistic evidence: the **mental lexicon holds only a core meaning of a word**, and polysemous extensions are computed on the fly from contextual features.
- But: how to identify the homograph-level distinctions that are useful for NLP, since they are not explicitly identified as such in any existing resource?

WSD: Knowledge-based Methods

- Large-scale lexical resources
- **Machine-readable dictionaries:**
 - » Ready-made source of information about word senses
 - » Text of dictionary definitions provides lexical and semantic knowledge
- **Thesauri:**
 - » Information about relationships among words
 - » Words in the same category are semantically related
- **Problems:**
 - » Inconsistencies
 - » Created for human use, not for machine exploitation

WSD: Data/Corpus-driven Methods

- Context and co-occurrence
- Bag of words vs. relational information
- Empirical methods for the study on words:
bank of samples
- Frequency and polysemy (Zipf, 1945)
- Supervised learning from sense-tagged corpora
- Senses for clusters vs. word occurrences
- Bilingual corpora, using translations

Sense Disambiguation with a MRD

Michael Lesk (1986): “Automatic sense disambiguation using machine-readable dictionaries: How to tell a pine cone from an ice cream cone”. In *Proceedings of the SIGDOC Conference*. Toronto, Canada.

- Idea: decide automatically which sense of a word is intended within a context
 - » by using machine-readable dictionaries, and
 - » looking for words in the sense definitions that
 - » overlap words in the definition of nearby words
- Non-syntactic approach
- Not dependent on prototypical information

Example

- pine*** 7 kinds of evergreen tree with needle-shaped evergreen tree
- 0 waste away through sorrow or illness
- 0 pine for sth; pine to do sth; have a pine
- cone*** 0 solid body which narrows to a point ...
- 0 sth of this shape whether solid or hollow
- 8 fruit of certain evergreen trees (fir, pine, ...)

One Sense per Discourse

William A. Gale, Kenneth W. Church, David Yarowsky (1992): “**One sense per discourse**”. In *Proceedings of the DARPA Speech and Natural Language Workshop*.

- Assumption: if a **polysemous word appears two or more times in a well-written discourse**, it is extremely likely that they will share the **same sense**
- Tendency to share sense in the same discourse: **98%**
- Implication: tag senses on a per-discourse basis

Predominant Senses in Untagged Text

Diana McCarthy, Rob Koeling, Julie Weeds, John Carroll (2004): “[Finding predominant senses in untagged text](#)”. In *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics*.

- Distribution of the senses of a word is often skewed.
 - Heuristic of **choosing the most common sense in WSD** is not optimal, but extremely powerful.
 - Heuristic assumes some quantity of hand-tagged data, and depends on genre and domain.
- Use a **thesaurus** and **distributional similarity** to find predominant senses in untagged text.
Example: *star* → *sky, winner, hero, sun, ...*

Sense Ranking of Thesaurus Neighbours

- The senses of each word w are the senses in WordNet.
- Each neighbor n_k of a word w is associated with a **distributional similarity score**: $dss(w, n_k)$.
- For each sense of w ($ws_i \in \text{senses}(w)$), a ranking score is obtained by summing over the $dss(w, n_k)$ of each neighbour n_k , multiplied by a weight.
- The weight is the WordNet similarity score between the target sense ws_i and the sense of n_k that maximises this score.
- Example: *star* \rightarrow *sky, winner, hero, sun, ...*

SensEval

- SensEval: [open evaluation exercise](#) for WSD
- SensEval is the SIGLEX international organisation devoted to the evaluation of WSD systems.
- Gold standard for word sense annotation: [SemCor](#) (a.o.)
- Sense inventory taken from [WordNet](#)
- Task: develop a [supervised/unsupervised system](#) for sense annotation (all words, lexical sample)
- Subtasks for subcategorisation acquisition, etc.
- Basis: mixture of labeled and unlabeled data

SensEval

- Purposes:
 - » agree on an explicit and detailed **definition of the task**
 - » produce a **gold standard corpus** of correct answers
 - » evaluate the strengths and weaknesses of WSD programs with respect to different words, different varieties of language, and different languages
 - » further the understanding of lexical semantics and polysemy
- Languages: English, Italian, Basque, Catalan, Chinese, Romanian, Spanish

References



Word Senses: References

- George K. Zipf (1949): „*Human behaviour and the principle of least-effort*“. Addison-Wesley, Cambridge, MA.
- Jurij D. Apresjan (1974): „*Regular polysemy*“. *Linguistics* 142: 5-32.
- Adam Killgarriff (1997): „*I don't believe in word senses*“. *Computers and the Humanities* 31: 91-113.
- Patrick Hanks (2000): „*Do word meanings exist?*“ *Computers and the Humanities* 34: 205-215.
- Harald Baayen (2001): „*Word frequency distributions*“. Kluwer Academic Publishers.
- Marco Baroni (2007?): „*Distributions in text*“. In: Anke Lüdeling and Merja Kytö, editors: „*Corpus Linguistics. An International Handbook*“.

WordNet: References

- George A. Miller, editor (1990): "[WordNet: An on-line lexical database](#)". Special issue of the *International Journal of Lexicography*, 3(4).
- Christiane Fellbaum, editor (1998): "[WordNet - An electronic lexical database](#)". MIT Press.
- Piek Vossen (2004): "[EuroWordNet: A multilingual database of autonomous and language-specific WordNets connected via an interlingual index](#)". *International Journal of Lexicography*, 17:161-173.
- WordNets online:
 - » WordNet: <http://wordnet.princeton.edu/>
 - » EuroWordNet: <http://www.ilc.uva.nl/EuroWordNet/>
 - » MultiWordNet: <http://multiwordnet.itc.it/>
 - » Global WordNet Association: <http://www.globalwordnet.org/>

WSD: References

- Nancy Ide and Jean Véronis (1998): “Introduction to the special issue on Word Sense Disambiguation: The state of the art”. *CL* 24(1):1-40.
- Hwee T. Ng, Chung Y. Lim, Shou K. Foo (1999): “A case study on the inter-annotator agreement for word sense disambiguation”. In *Proceedings of the ACL Workshop on Standardizing Lexical Resources*.
- Nancy Ide and Yorick Wilks (2006): “Making sense about sense”. In Eneko Agirre and Philip Edmonds, editors: *Word Sense Disambiguation: Algorithms and Applications*, chapter 3.
- Roberto Navigli (2006): “Reducing the granularity of a computational lexicon via an automatic mapping to a coarse-grained sense inventory”. In *Proceedings of the 5th LREC*. Genoa, Italy.
- Martha Palmer, Hoa Trang Dang, Christiane Fellbaum (2007): “Making fine-grained and coarse-grained sense distinctions, both manually and automatically”. *Natural Language Engineering* 13(2): 137-163.

WSD Approaches & SensEval: References

- Michael Lesk (1986): “Automatic sense disambiguation using machine-readable dictionaries: How to tell a pine cone from an ice cream cone”. In *Proceedings of the SIGDOC Conference*. Toronto, Canada.
- William A. Gale, Kenneth W. Church, David Yarowsky (1992): “One sense per discourse”. In *Proceedings of the DARPA Speech and Natural Language Workshop*.
- Adam Kilgarriff and Martha Palmer (2000): “Introduction to the special issue on SensEval”. *Computers and the Humanities*, 34(1-2):1-13.
- Philip Edmonds and Adam Kilgarriff (2002): “Introduction to the special issue on evaluating Word Sense Disambiguation systems”. *Journal of Natural Language Engineering*, 8(4).
- Diana McCarthy, Rob Koeling, Julie Weeds, John Carroll (2004): “Finding predominant senses in untagged text”. In *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics*.