Lexical Semantics

Martin Rajman
&
Jean-Cédric Chappelier

Overview

- **Basic concepts**
- **Semantic relations**
- **Resources for Lexical Semantics: Wordnet**
- **Applications of Lexical Semantics**

Basic concepts

Lexical Semantics vs. Compositional Semantics

- **Lexical semantics**: The study of the meaning of *words*

- **Compositional Semantics**: the study of the meaning of *sentences* (or more complex linguistic entities: paragraphs, texts, …)

Compositional Semantics

- ***Compositional Semantics*** is the study of the meaning of complex linguistic units such as sentences, paragraphs, or documents

- A standard approach for exploring compositional semantics with human subjects are ***reading tests***

Reading tests

- Consider the following text:

"Under Peter's supervision, John is participating to an experiments consisting in placing on a table blocks with various shapes and colors initially lying on the floor.

The first day, he puts two triangle blocks on the table, one red and one green.

The second day, he replaces the red triangle block by a square block of the same color, and added a green triangle block."

- Answer the following questions:
  1. Who is manipulating the blocks during the experiment?
  2. How many blocks are on the table at the end of the experiment?
  3. What is the shape of the red block(s) on the table at the end of day 1?
  4. How many triangles have been manipulated during the whole experiment?

Reading tests (2)

- The test may seem trivial to (almost any, at least English speaking) human subject... however, it requires a lot of knowledge to be successfully passed!
  - Knowledge about involved objects: What is a block? What is a shape? What is a color? What is a table? What is a floor?
  - Knowledge about involved actions: What is participate? Consist? Lie? ...
  - Knowledge about people who are referred to: Who is John? Who is Peter?
  - Knowledge about the language: syntactic analysis (e.g. in “blocks (...) initially lying on the floor”, what is the subject of lying?); anaphora resolution (who is the pronoun “he” in the second sentence referring to?)
  - Knowledge about the real world: e.g. when a block is put on a table, it stays there (while a drop of water may evaporate or a feather may be blown away) or if somebody is participating to an experiments, s/he is performing the actions during this experiment, not the person who is supervising it! ...

Usual representations

- Symbolic representations:
  - various formal logics: the meaning is expressed as a logical formula that can then be manipulated through various inferential mechanisms;
  - various graph based representations: the meaning is expressed as a graph that can then be manipulated through various graph transformations;

- Vectorial representations:
  - typically approaches based on “distributional semantics” (e.g. Word embeddings): the meaning is represented as a vector in a (usually high dimension) vector space and can then be manipulated through vector based operations (e.g. weighted sums, projections, etc.)

Usual representations (2)

- Currently, only vectorial representations can be deployed at a large scale because:
  - it is extremely difficult (if not impossible) to guarantee the consistency of large sets of logical propositions derived from textual input, which often makes the inferential mechanisms very hard to use;
  - there isn't yet a consensus neither on which are the most suitable graph based representations (semantic nets? Conceptual graphs? ...) for expressing the meaning of linguistic entities, nor on which are the proper operations to be applied to these representations;
- ... but the currently associated vector based operations seems to be too simplistic for suitably mimicking the transformations that are required to manipulate linguistic meaning.

Intermediate conclusion

- Large scale Compositional Semantics is still out of reach,

and

- This lecture will therefore restrict to a simpler form of semantics, the semantics of individual words, e.g. ***Lexical Semantics***

Lexical Semantics

- ***Lexical Semantics*** is the study of the meaning of words (i.e. of the simplest linguistic units)
- A standard approach for exploring lexical semantics for human subjects are ***dictionaries*** (not to be confused with encyclopedias which are not concerned with word meanings but with comprehensive information about subjects/topics/fields from the real world)

Note: In this course, a dictionary (especially when tailored for some automated processing) will also often be called a *lexicon*

Word sense

- A word sense can be represented, for example, as :
  - A definition in natural language
  - A definition based on it's relationship (e.g."is a", "has a") to other word senses
  - A set of synonyms ("synset")
- Dictionaries usually define word senses/meanings … However, different dictionaries often use different definitions (different content and/or different granularity)

(a bit) more formally: Lexemes

- Lexeme:
  - An individual entry in a lexicon/dictionary
  - A pairing of a particular orthographic and phonological form with some meaning representation

Orthographic form	Phonological form	Meaning
1. bass	[beys]	*adj.* low in pitch; a bass instrument
2. bass	[bas]	*n.* (…) freshwater or marine fishes (…)
3. wood	[woo d]	*n.* (…) substance of a tree (…)
4. would	[woo d]	*v.* A pt. and pp. of WILL

- … And, the sense of a word is then the meaning component of the associated lexeme…

Dictionary definitions

- Propose a definition for the word “bee”…

By Bartosz Kosiorek Gang65 - Own work, CC BY-SA 3.0,
https://commons.wikimedia.org/w/index.php?curid=1992636

Dictionary definitions (2)

- Definition of “bee” (according to the English Wiktionary):

  “A flying insect, of the superfamily Apoidea, known for its organised societies and for collecting pollen and (in some species) producing wax and honey.”

- The definition requires the meaning of the words it contains...
  - Apoidea: A taxonomic superfamily within the order Hymenoptera – the bees and some wasps.
  - To fly: To travel through the air, another gas or a vacuum, without being in contact with a grounded surface.
  - Insect: An arthropod in the class Insecta, characterized by six legs, up to four wings, and a chitinous exoskeleton.

Word sense definition in Natural Language

- If the different meanings (aka senses) of a words are defined by well chosen definitions in natural language (as it is the case in dictionaries), we are faced with a vicious circle:
  - understanding the meaning (i.e. making it exploitable) of the different senses of a word (lexical semantics) requires to understand the meaning of the associated definitions and thus the availability of some form of compositional semantics...
- To break this vicious circle, natural language cannot be used to define the various meanings of a word and some more formal representations must be used instead; in this course, we will consider two types of formalisms:
  - semantic relations, and
  - synsets (see the slides on Wordnet)

Semantic Relations

Many different types of relations

- Relations characterizing word meanings:
  - Homonymy / Homophony / Homography
  - Polysemy
  - Synonymy

- Relation connecting word meanings:
  - Hyponymy/Hypernymy
  - Meronymy/Holonymy

Homonymy, homophony, homography

- **Homophony**: two distinct words are homophones is they have the same pronunciation (i.e. the same "phonological form")

Example: "die" and "dye"

- **Homography**: two words are homographs if they are spelled the same (i.e. have the same "orthographic form") but not pronounced the same

Example: "bass" (the fish) and "bass" (the guitar)

- **Homonymy**: two words are homonyms if they are spelled and pronounced the same, but do not have the same meaning

Example: "bat" (the wooden club) and "bat" (the flying mammal)

Polysemy

- A relation that holds between multiple *related* meanings within a single lexeme

Orthographical form	Meaning
Crown	1. Headgear worn by a monarch 2. The highest part of anything, e.g. a tree 3.The part of a tooth that is covered by enamel …

Homonymy vs. Polysemy

- Both homonyms and polysems are spelled and pronounced the same but …
- homonyms have a different etymology and usually correspond to two distinct entries in a lexicon, while polysems share the same etymology but correspond to two different meaning of the same lexicon entry

Example:

➢ “bat” (the flying mammal) comes from a dialectal variant of the Middle English “bakke”, while “bat” (the wooden club) comes from the Old English “batt”

but

➢ “crown” (the headgear) and “crown” (the highest part) both come from the Anglo-Norman “coroune”

Source for polysemy

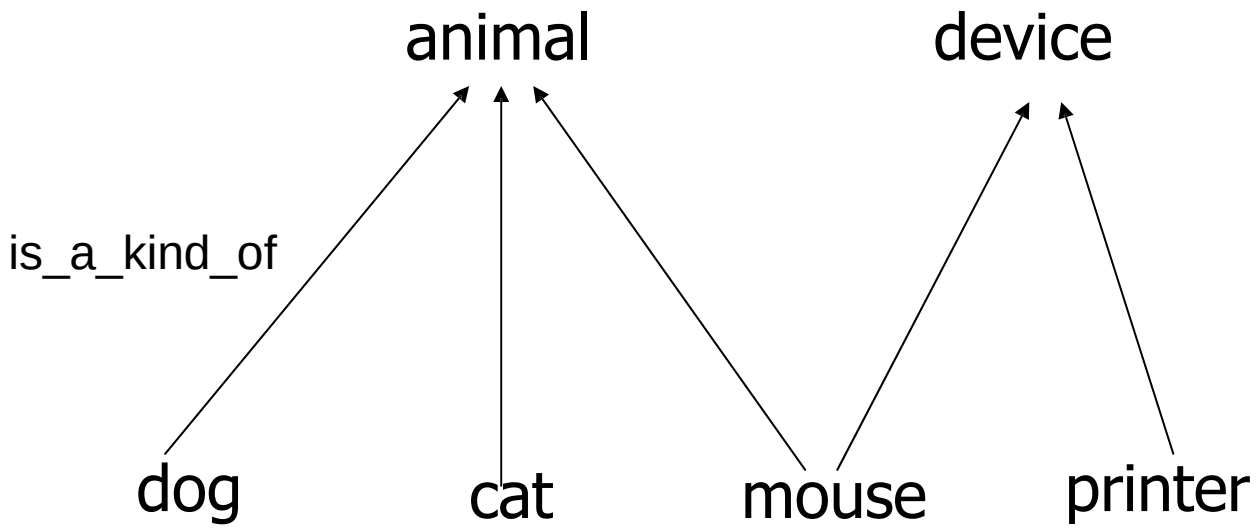
- Metaphor
  - “Germany will pull Slovenia out of its economic slump”
  - “I spent 2 hours on that homework”
- Metonymy
  - “The White House announced yesterday”
  - “This chapter talks about part-of-speech tagging”
  - Bank (building) and bank (financial institution)

Synonymy

- Two words are synonymous if they have the same sense
- Criteria for synonymy:
  - They have the same value for all their semantic features
  - They map to the same concept
  - They satisfy the Leibniz substitution theory
    - The substitution of one for the other never changes the truth value of a sentence in which the substitution is made
- Example of non-synonyms:
    - Tony is the **big** brother
    - Tony is the **large** brother

Hyponymy/Hypernymy

A ***hyponym*** is a word whose meaning contains the entire meaning of another, known as the *superordinate* or ***hypernym***.



Meronymy/Holonymy

- A word w1 is a meronym of another word w2 (the holonym) if the relation **is-part-of** holds between the meaning of w1 and w2.
  - Meronymy is transitive and asymmetric
  - A meronym can have many holonyms
  - Meronyms are distinguishing features that hyponyms can inherit.
    - Ex. If “beak” and “wing” are meronyms of “bird”, and if “canary” is a hyponym of “bird”, then (by inheritance), “beak” and “wing” must be meronyms of “canary”.
  - Limited transitivity:
    - Ex. “A house has a door” and “a door has a handle”, then “a house has a handle” (?)

Different type of meronymic (part-whole) relationships

- Component-object (branch/tree)
- Member-collection (tree/forest)
- Portion-mass (slice/cake)
- Stuff-object (aluminium/airplane)
- Feature-activity (paying/shopping)
- Place-area (Lausanne/Vaud)
- Phase-process (addolescence/growing up).

Defining word senses with semantic relations

- A standard way of defining word senses with semantic relations is to follow the Aristotelian principle of "Genus-Differentia":
  - Genus: each word meaning is first associated to a hypernym through a "hyponymy/hypernymy" relation (this is equivalent to defining the superclass associated with a given class in an object oriented model)
  - Differentia: each word meaning is then uniquely differentiated from the other hyponyms of its hypernym by additional relations (e.g. Meronymy/Holonymy) associating it with other words meanings
- Of course, to make this type of approach realistic on a large scale, more than two semantic relations are required!

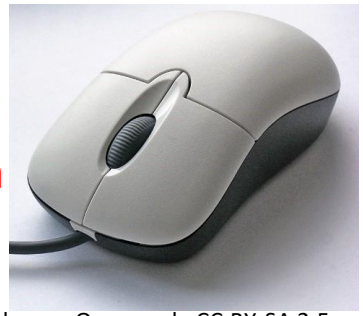
Hands-on…

Lexical Semantics with semantic relations

- Consider the following meanings of the word "mouse":
  1. Any small rodent of the genus Mus.

https://commons.wikimedia.org/w/index.php?curid=28335

  2. An input device that is moved over a pad or other flat surface to produce a corresponding movement of a pointer on a graphical display.



By Darkone - Own work, CC BY-SA 2.5
https://commons.wikimedia.org/w/index.php?curid=235633

How could you use semantic relations to distinguish between these two meanings?

Lexical semantics with semantic relations (2)

- Mouse:
  1. hyponym of “rodent”
  2. hyponym of “device”

Lexical Semantics with semantic relations (3)

- Consider the following meanings of the word "wood":
  1. The substance making up the central part of the trunk and branches of a tree.
     example: this table is made of wood
  2. A forested or wooded area.
     example: he got lost in the wood
  3. A type of golf club, the head of which was traditionally made of wood.
     example: he played golf with a wood

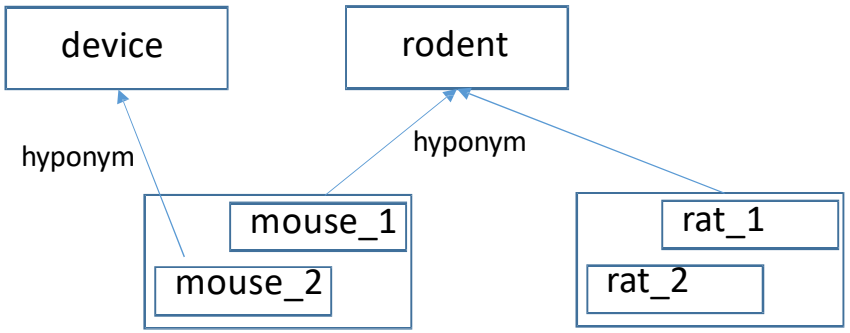
How could you use semantic relations to distinguish between these two meanings?

Lexical semantics with semantic relations (4)

- Wood:
  1. hyponym of “substance”
  2. hyponym of “area”
  3. hyponym of “club”

Let us go further!...

- The definitions based on semantic relations given so far are good enough for distinguishing the meanings of various polysemic words but they do not allow to distinguish between the hyponyms of a given hypernym!...



But how to distinguish between mouse_1 and rat_1?

Let us go further!... (2)

- Let us recall the definitions of mouse_1 and rat_1:

➢ mouse_1: Any small rodent of the genus Mus.

➢ rat_1: Any medium-sized rodent belonging to the genus Rattus.

- To distinguish between mouse_1 and rat_1, additional semantic relations may be used...

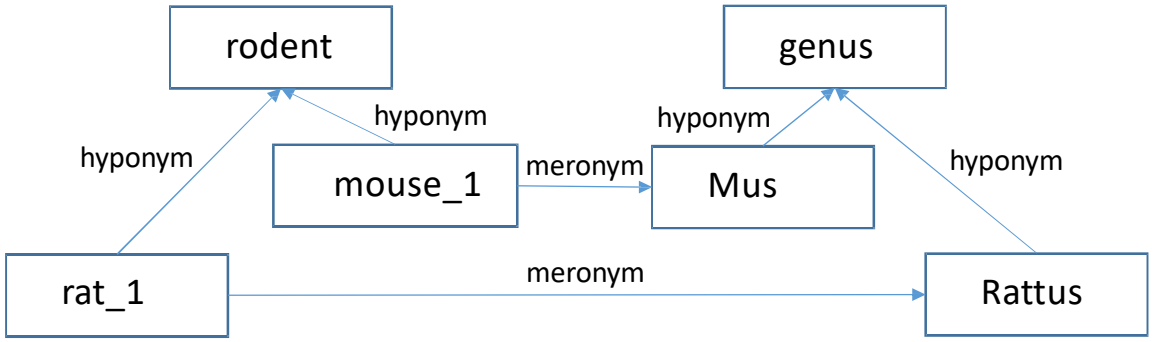
Let us go further!... (3)

- For example:

➢ mouse_1: hyponym of “rodent” and meronym of “Mus”

➢ rat_1: hyponym of “rodent” and meronym of “Rattus”

which, if we add the fact that “Mus” and “Rattus” are both hyponyms of “genus” would lead to the following graph based representation:



Let us go further!... (5)

- Exercise: Apply the Genus-Differentia approach to differentiate:

➢ wood_1: The substance making up the central part of the trunk and branches of a tree.

from

➢ stone_1: A hard earthen substance that can form large rocks.

Intermediate conclusion (2)

- In a relation based approach to Lexical Semantics, the word meanings are defined as the nodes of a directed graph the arcs of which correspond to various semantic relations
- The targeted semantic graph is built with the main purpose of correctly differentiating the various meanings of the words (which is one of the primary objectives of Lexical Semantics)
- Most often, pure lexical semantic models are not sophisticated enough to be fully adequate for more advanced exploitations (such as the automated generation of the answers to the questions asked in the simple reading test given at the beginning of the lecture)
- For such advanced applications, lexical semantic models will have to be embedded in more complex ones providing some (possibly limited) semantic representation for linguistic units larger than words (Compositional Semantics)

Resources for Lexical Semantics

WordNet

http://wordnetweb.princeton.edu/perl/webwn

"WordNet® is a large lexical database of English. Nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms (**synsets**), each expressing a distinct concept."

WordNet Search - 3.1

- WordNet home page - Glossary - Help

Word to search for: mouse Search WordNet

Display Options: (Select option to change) Change

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations

Display options for sense: (gloss) "an example sentence"

Display options for word: word#sense number

Noun

- S: (n) **mouse#1** (any of numerous small rodents typically resembling diminutive rats having pointed snouts and small ears on elongated bodies with slender usually hairless tails)
- S: (n) shiner#1, black eye#1, **mouse#2** (a swollen bruise caused by a blow to the eye)
- S: (n) **mouse#3** (person who is quiet or timid)
- S: (n) **mouse#4**, computer mouse#1 (a hand-operated electronic device that controls the coordinates of a cursor on your computer screen as you move it around on a pad; on the bottom of the device is a ball that rolls on the surface of the pad) *"a mouse takes much more room than a trackball"*

Verb

- S: (v) sneak#1, **mouse#1**, creep#2, pussyfoot#1 (to go stealthily or furtively) *"..stead of sneaking around spying on the neighbor's house"*
- S: (v) **mouse#2** (manipulate the mouse of a computer)

WordNet Search - 3.1

- WordNet home page - Glossary - Help

Word to search for: mouse Search WordNet

Display Options: (Select option to change) Change

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations

Display options for word: word#sense number

Noun

- S: (n) **mouse#1**
- S: (n) shiner#1, black eye#1, **mouse#2**
- S: (n) **mouse#3**
- S: (n) **mouse#4**, computer mouse#1

Verb

- S: (v) sneak#1, **mouse#1**, creep#2, pussyfoot#1
- S: (v) **mouse#2**

Synsets

- Hypothesis: A *synonym* is often sufficient to identify a sense.
  - Example
    - **“board”** means 1) piece of lumber 2) group of people assembled for some reason
    - Sense 1: {board, **plank**} Sense 2: {board, **committee**}

  (Note that this is true for English which is rich in synonyms but may not be true for all languages…)

- A synset is a set of words (surface forms) that share a given sense/meaning/concept

- Synsets do not explain what senses are, they simply express that they exist, and allow to differentiate them from each other

How is meaning represented?

- Wordnet (differential approach)
  - Meanings (concepts) are represented as **a list of word forms** (the synsets) that *distinguish* them from other meanings
  - No two synsets should contain exactly the same set of words

- Dictionaries (conventional approach)
  - Meanings are represented as natural language expressions
  - The expressions have to contain enough information to accurately *define* the concepts
    - Not so easy, definitions are often cyclic
      - **Tree**: “a plant having a permanently **woody** main stem or trunk…”
      - **Wood**: “the hard, fibrous substance composing most of the stem and branches of a **tree”**
    - Conventional dictionaries rarely meet this requirement

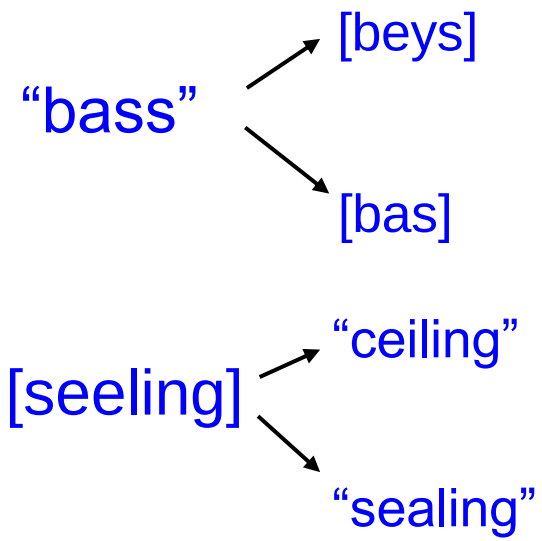
Word categories in Wordnet

- Nouns
  - Organised as topical hierarchies with lexical inheritance (hyponymy/hyperymy and meronymy/holonymy).
- Verbs
  - Organised by a variety of entailment relations
- Adjectives
  - Organised on the basis of bipolar opposition (antonymy relations)
- Adverbs
  - Like adjectives

Application of lexical semantics in language engineering

Lexical semantics in Speech Processing

- Text to speech
  - Choose the right pronunciation of a word depending on the word's meaning
- Speech recognition
  - Choose the right word among possible words with the same pronunciation (homophones)



Lexical semantics for Spelling Error Correction

- In some cases a spelling error can result in a real word in the lexicon and therefore cannot be detected by a conventional spell checker

- Such errors can be detected by using word meanings to compute lexical cohesion and identify tokens that are semantically unrelated to their context

- Examples:
  - It is my sincere hole [hope] that you will recover soon
  - The boss [toss] of the coin

Lexical semantics in Information Retrieval

- Semantic indexing
  - Indexing word senses instead of words
  - Improves
    - Recall by handling synonymy
    - Precision by handling homonymy and polysemy

Example 1: Different indexing of the term “Java”

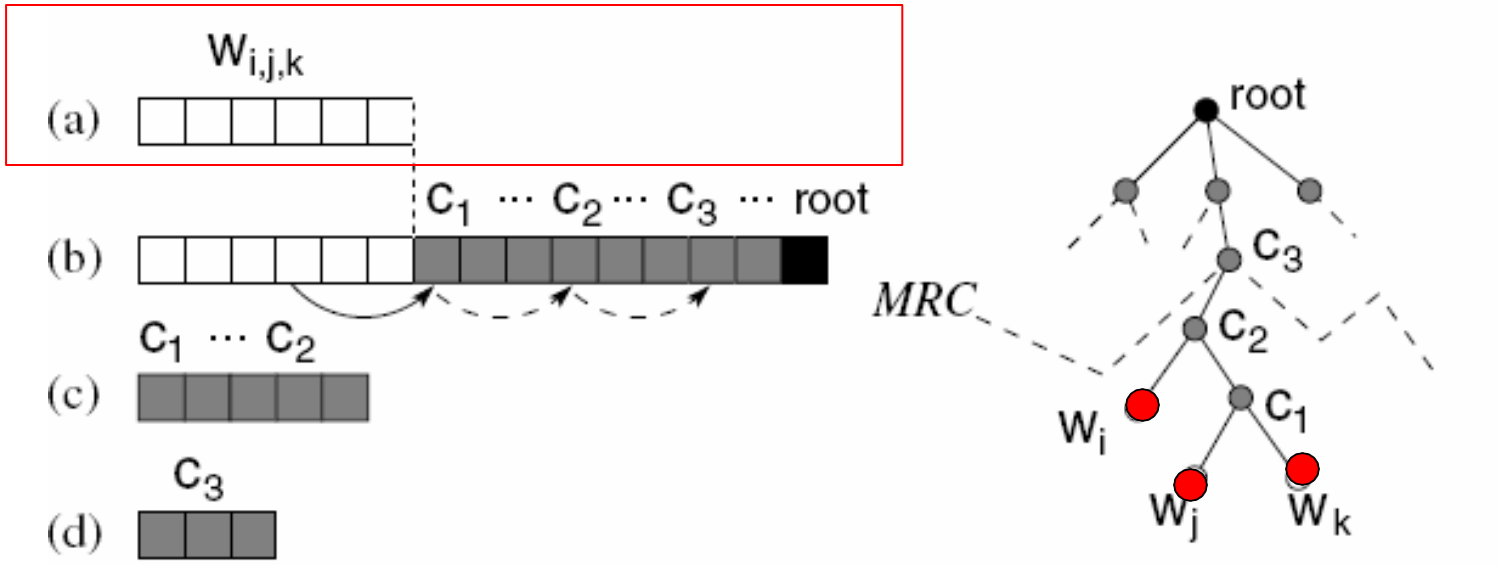
- Programming language
- Type of coffee
- Location

Example 2: a query for "cars" should also return a document that mentions only "automobiles"

Lexical semantics for Information Retrieval

- Indexing schemes

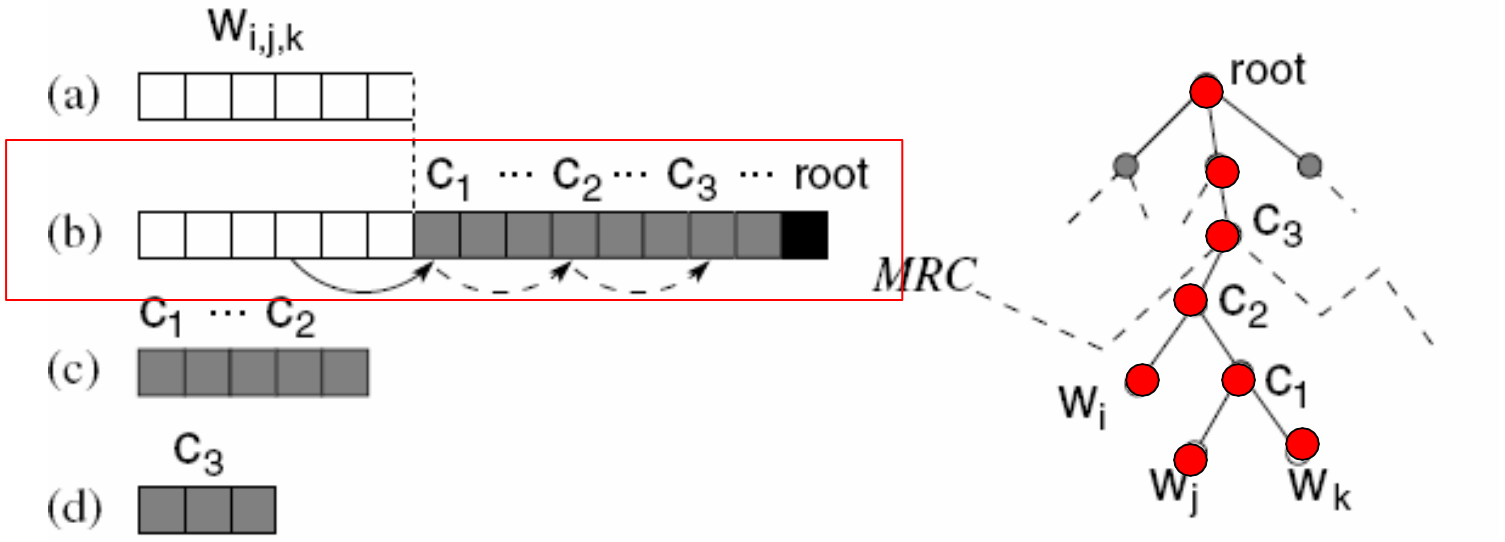
  a) Standard indexing with words (stems or lemmas)



Lexical semantics for Information Retrieval

- Indexing schemes

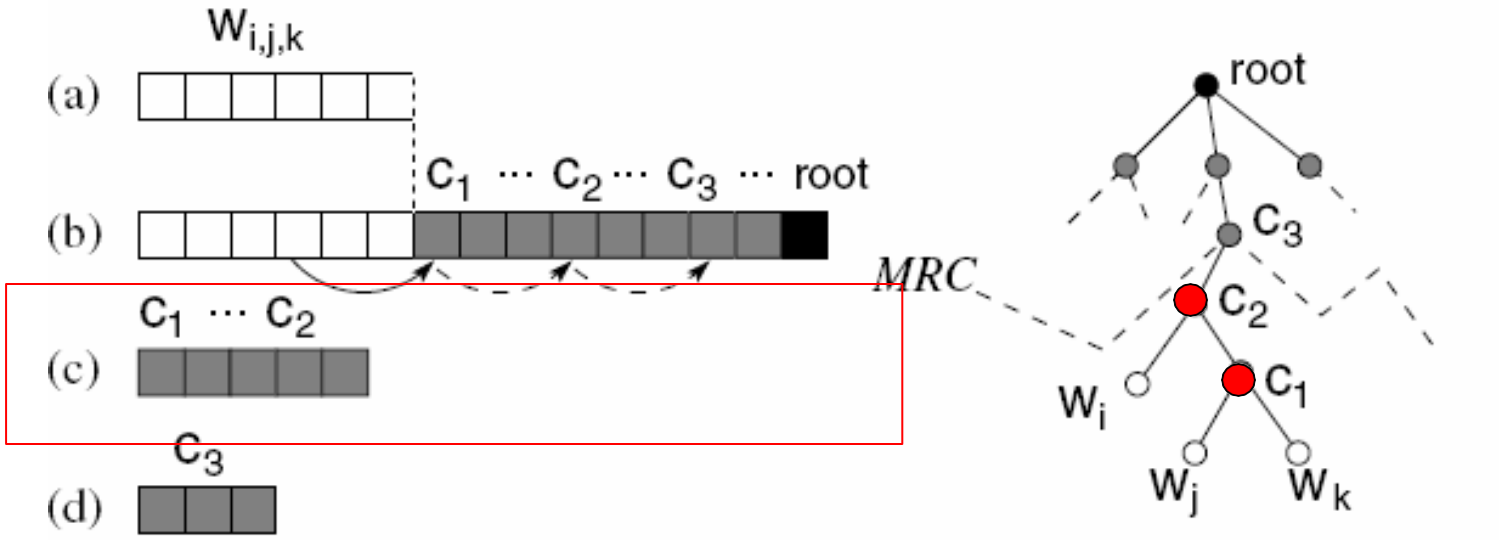
  b) Indexing with a semantic ontology, each indexing term is extended with all the hypernym senses



Lexical semantics for Information Retrieval

- Indexing schemes

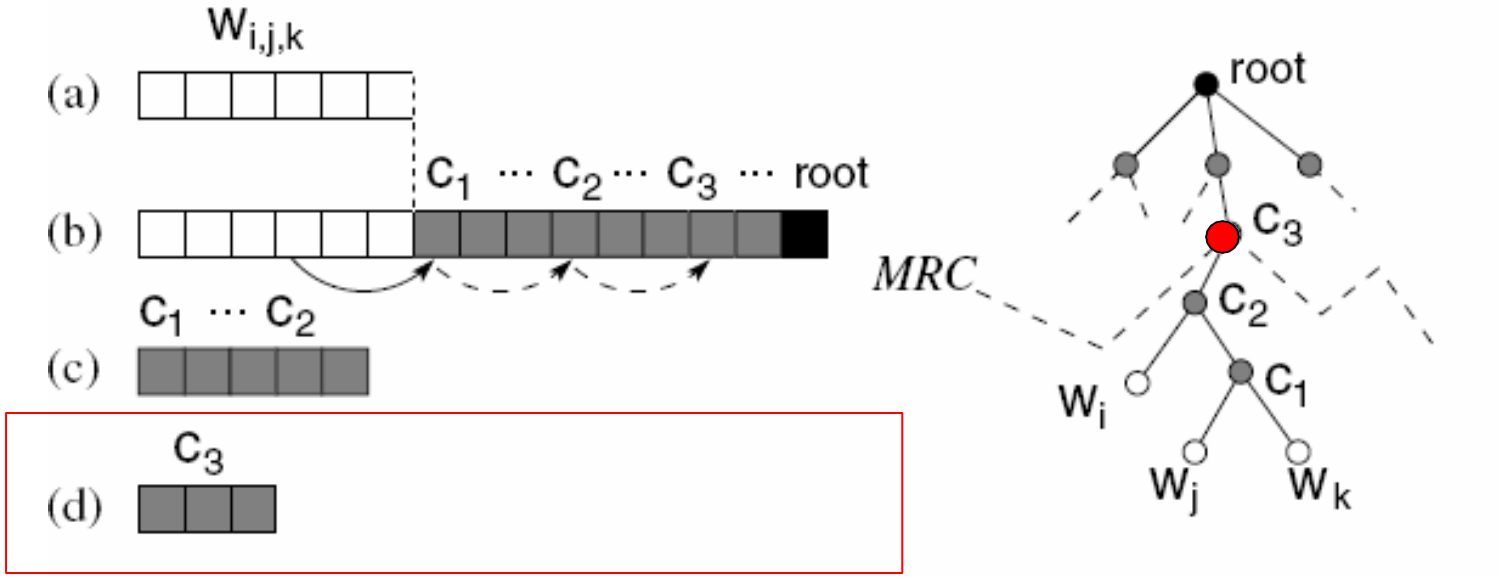
  c) Synset (or hypernyms synsets) indexing, each indexing term is replaced with it's hypernym synset



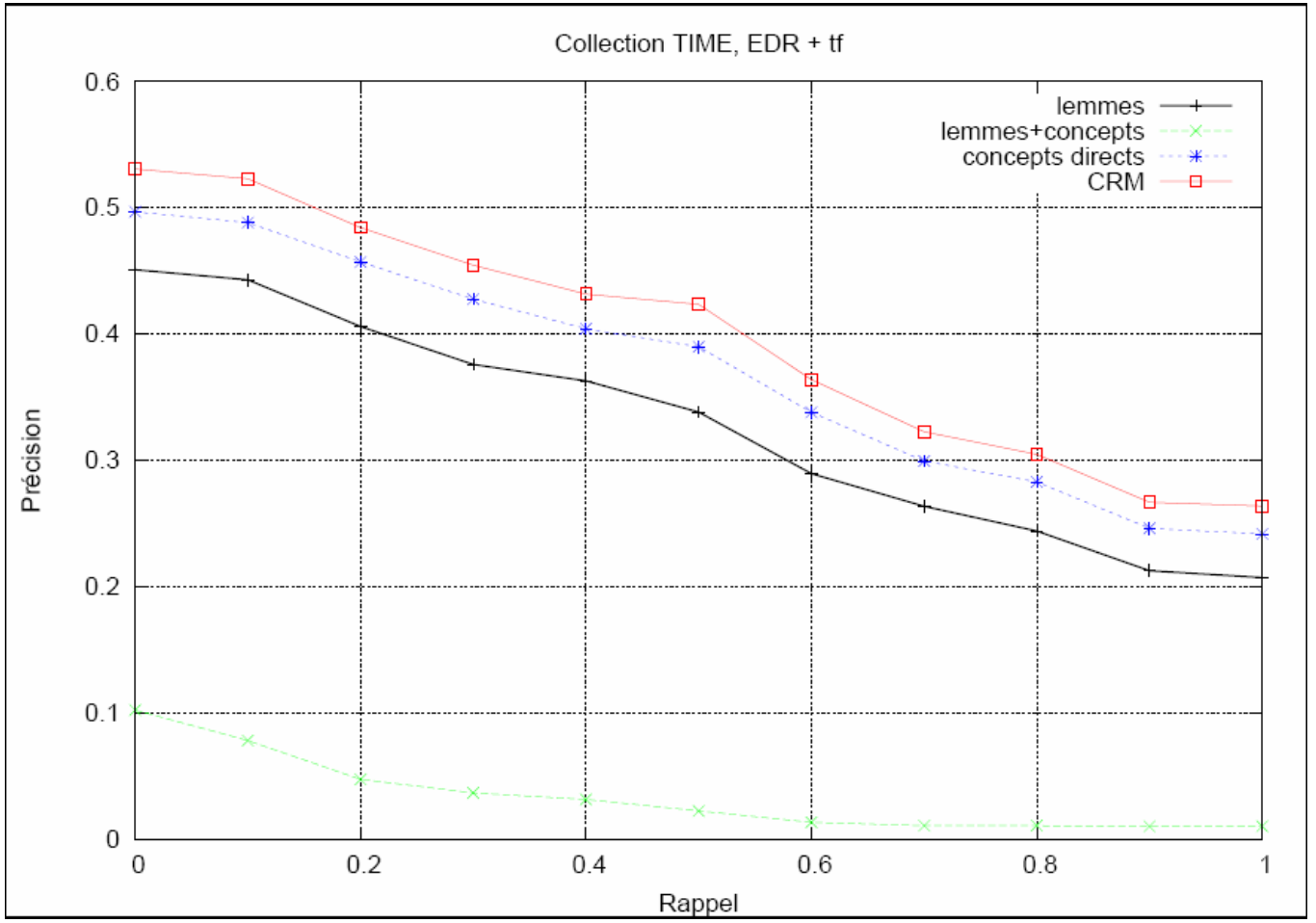
Lexical semantics for Information Retrieval

- Indexing schemes

  d) Minimum Redundancy Cut (MRC) indexing, each indexing term is replaced with it's dominating semantic concept defined by MRC



Lexical semantics for Information Retrieval



Key points

- Difference between compositional and lexical semantics
- Semantic relations and how to use them for defining word senses (Aristotelian Genus-Differentia principle)
- Synsets and how to use them for defining word senses (Wordnet)
- Example of applications for Lexical Semantics

References

- Cruse, D. A. (1986). Lexical Semantics. Cambridge, New York.
- Dan Jurafsky and Jim Martin, Speech and Language Processing, Chapter16, Prentice Hall, 2000.
- Mark Stevenson, Word Sense Disambiguation, CSLI Press, 2003.
- Sanda Harabagiu and Dan Moldovan, Enriching the WordNet Taxonomy with Contextual Knowledge Acquired from Text, in Natural Language Processing and Knowledge Representation: Language for Knowledge and Knowledge for Language, (Eds) S. Shapiro and L. Iwanska, AAAI/MIT Press, 2000, pages 301-334.
- Sanda Harabagiu and Dan Moldovan, A Parallel System for Text Inference Using Marker Propagations, IEEE Transactions in Parallel and Distributed Systems August, 1998, pages 729-747.
- FrameNet web site: http://framenet.icsi.berkeley.edu/