

## Computational Linguistics

### Computational Lexical Semantics

Bali Ranaivo-Malançon & Suhaila Saeed

Faculty of Computer Science and Information Technology  
Universiti Malaysia Sarawak

August 2014



This OpenCourseWare@UNIMAS and its related course materials are licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.



# Lexical Semantics

The study of the meaning of words.

- Words name concepts
- Concepts are the cognitive representations of word meanings
- Lexical semantics studies those concepts that have names
  - e.g., the verb WALK names the set {*walk, walked, walking, walks*}
- Central topics
  - Paradigmatic relations of meaning, such as synonymy, antonymy, and hyponymy
  - Syntagmatic relations of meaning including selectional restrictions
  - Taxonomic hierarchies
  - Processes of meaning extension, such as metaphor and metonymy



# Lexicon, Lexeme, Lemma, Sense

- The adjective *lexical* is derived from the noun *lexicon*
- A *lexicon* contains a finite set list of *lexemes*
- Thus, lexemes are the fundamental units of lexical semantics
- **A lexeme is the base form of a word**
  - e.g., the word-forms {*walk, walked, walking, walks*} would be reduced to the lexeme *walk*
- **All the variants of a lexeme form the lemma**
  - e.g., the variants {*walk, walked, walking, walks*} form the lemma WALK
  - In other words, the lemma WALK can be realised by the word-forms {*walk, walked, walking, walks*}



# Lexicon, Lexeme, Lemma, Sense (cont'd)

- A lemma/lexeme has two components
  - ① **Form** - a particular orthographic and phonological form
  - ② **Sense** - a symbolic meaning representation
- "Word sense is a commonly-accepted meaning of a word."  
(Navigli, 2012)
- CONVENTION: LEMMAS are in upper-case, *word-forms* are in lower case italics
- RECALL: Lemmatisation is the process of mapping a word-form to a lemma



# Lexical Ambiguities

- The relationship between words and concepts is not one-to-one
- Syntactic ambiguities are solved by POS tagging techniques
- Lexical semantic ambiguities are solved by word sense disambiguation techniques



# Homonymy vs Homographs vs Homophones

- **Homonyms** are words with identical orthographic and phonological forms, but unrelated meanings
  - *bank* as financial institution
  - *bank* as a sloping land
- **Homographs** are words with identical spelling, but different phonological form and different meanings
  - *bass* [bas] as a fish and *bass* [beys] as a musical term
- **Homophones** are words with the same pronunciation but different spellings
  - *their* and *there*



# Polysemy

- Some words have more than one meaning (fairly similar or completely different)
- The meaning of a word is determined by the examining its context
  - The noun *bass* [beys] has five senses in WordWeb Pro 6.4
    1. The lowest part of the musical range
    2. The lowest part in polyphonic music
    3. An adult male singer with the lowest voice
    4. The lowest adult male singing voice
    5. The member with the lowest range of a family of musical instruments



# Lexical Semantics Concerns

Lexical semantics is concerned with

- the relation of a **lexeme to the world**
- the relation of a **lexeme with other lexemes**





# Synonymy

- Two words that belong to the same language are said synonyms if they have a significant similar semantic content
  - CAR and AUTOMOBILE
- Note that absolute synonymy is rare
  - **Absolute synonymy**: there are no contexts in which substituting one word for the other has any semantic effect



# Antonymy

- Two words that belong to the same language are said antonyms if they have opposite meanings
  - **HOT** in its basic sense is an antonym of **COLD**



# Hyponymy

- Refers to the taxonomic relation where one lexeme (the **hyponym**) is a subclass of another (the **hyperonym** or superordinate)

Hyperonym	Hyponyms
ANIMAL	DOG, CAT, ...
FRUIT	BANANA, MANGO, ...



# Meronymy

- Describes the **part-whole relation**
- **Meronym**: the word referring to the part
- **Holonym**: the word referring to the whole

Holonym	Meronym
HAND	FINGER
FINGER	FINGERNAIL
KNIFE	BLADE
PAGE	BOOK




# Machine Readable Dictionary (MRD)

A computerised dictionary derived from existing a printed dictionary

- The content of a MRD is more structured compared to the content of a printed dictionary
- Some examples of well-known MRDs used by the academic computational linguistics community
  - LDOCE (*Longman Dictionary of Contemporary English*)
  - COBUILD (*Collins Cobuild English Language Dictionary*)
  - OALD (*Oxford Advanced Learner's Dictionary*)



# WordNet

- An open-source electronic lexical database of English
- Developed at Princeton University 
- English nouns, verbs, adjectives, and adverbs are organised into synonym sets called **synsets**
- Each synset consists of a list of synonymous word forms and semantic pointers
- A semantic pointer describes the relationships between a synset and other synsets
- The semantic pointers can be of a number of different types
  - Synonymy; Antonymy; Hypernymy; Hyponymy; Holonymy, Meronymy



# Example of WordNet Entry

The noun book has 11 senses (first 7 from tagged texts)

1. (785) **book** -- (a written work or composition that has been published (printed on pages bound together); "I am reading a good book on economics")
2. (131) **book**, volume -- (physical objects consisting of a number of pages bound together; "he used a large book as a doorstop")
3. (14) ledger, leger, account book, book of account, **book** -- (a record in which commercial accounts are recorded; "they got a subpoena to examine our books")
4. (9) **book** -- (a number of sheets (ticket or stamps etc.) bound together on one edge; "he bought a book of stamps")
5. (8) record, record book, **book** -- (a compilation of the known facts regarding something or someone; "Al Smith used to say, 'Let's look at the record'"; "his name is in all the record books")
6. (7) **book** -- (a major division of a long written composition; "the book of Isaiah")
7. (1) script, **book**, playscript -- (a written version of a play or other dramatic composition; used in preparing for a performance)
8. **book** -- (a collection of playing cards satisfying the rules of a card game)
9. **book**, rule book -- (a collection of rules or prescribed standards on the basis of which decisions are made; "they run things by the book around here")
10. Koran, Quran, al-Qur'an, **Book** -- (the sacred writings of Islam revealed by God to the prophet Muhammad during his life at Mecca and Medina)
11. Bible, Christian Bible, **Book**, Good Book, Holy Scripture, Holy Writ, Scripture, Word of God, Word -- (the sacred writings of the Christian religions; "he went to carry the Word to the heathen")

(SOURCE: WORDNET 2.1)



# Hypernyms of *book*, noun, Sense 2

## Sense 2

**book**, volume -- (physical objects consisting of a number of pages bound together; "he used a large book as a doorstep")

=> product, production -- (an artifact that has been created by someone or some process; "they improve their product every year"; "they export most of their agricultural production")

=> creation -- (an artifact that has been brought into existence by someone)

=> artifact, artefact -- (a man-made object taken as a whole)

=> whole, unit -- (an assemblage of parts that is regarded as a single entity; "how big is that part compared to the whole?"; "the team is a unit")

=> object, physical object -- (a tangible and visible entity; an entity that can cast a shadow; "it was full of rackets, balls and other objects")

=> physical entity -- (an entity that has physical existence)

=> entity -- (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))

(SOURCE: WORDNET 2.1)





# Hyponyms of *book*, noun, Sense 2

## Sense 2

**book**, volume -- (physical objects consisting of a number of pages bound together; "he used a large book as a doorstep")

- => album -- (a book of blank pages with pockets or envelopes; for organizing photographs or stamp collections etc)
- => coffee-table book -- (a elaborate oversize book suitable for displaying on a coffee table)
- => folio -- (a book (or manuscript) consisting of large sheets of paper folded in the middle to make two leaves or four pages; "the first folio of Shakespeare's plays")
- => hardback, hardcover -- (a book with cardboard or cloth or leather covers)
- => journal -- (a record book as a physical object)
- => novel -- (a printed and bound book that is an extended work of fiction; "his bookcases were filled with nothing but novels"; "he burned all the novels")
- => order book -- (a book in which customers' orders are entered; usually makes multiple copies of the order)
- => paperback book, paper-back book, paperback, softback book, softback, soft-cover book, soft-cover -- (a book with paper covers)
- => picture book -- (a book consisting chiefly of pictures)
- => sketchbook, sketch block, sketch pad -- (a book containing sheets of paper on which sketches can be drawn)
- => notebook -- (a book with blank pages for recording notes or memoranda)

(SOURCE: WORDNET 2.1)



# Holonyms of *book*, noun, Sense 6

## Sense 6

**book** -- (a major division of a long written composition; "the book of Isaiah")

PART OF: text, textual matter -- (the words of something written; "there were more than a thousand words of text"; "they handed out the printed text of the mayor's speech"; "he wants to reconstruct the original text")

(SOURCE: WORDNET 2.1)



## Meronyms of *book*, noun, Sense 2

### Sense 2

**book**, volume -- (physical objects consisting of a number of pages bound together; "he used a large book as a doorstop")

HAS PART: binding, book binding, cover, back -- (the protective covering on the front, back, and spine of a book; "the book had a leather binding")

HAS PART: fore edge, foreedge -- (the part of a book that faces inward when the book is shelved; the part opposite the spine)

HAS PART: spine, backbone -- (the part of a book's cover that encloses the inner side of the book's pages and that faces outward when the book is shelved; "the title and author were printed on the spine of the book")

(SOURCE: WORDNET 2.1)



# WordNet in Other Languages

Some examples,

- **EuroWordNet** containing European languages
- **MultiWordNet** - mainly European languages
- **Asian WordNet** containing Thai ,Korean, Japanese, Indonesian, Myanmar, Vietnamese, Mongolian, Bengali



# Wiktionary: Wiki + Dictionary

A human-readable dictionary/lexicon and publicly available on the Web.

- Active online community collaborate to edit and maintain a wiktionary (wiki principles)
- The wiktionary project is run by the *Wikimedia Foundation* (the same organisation handling Wikipedia)
- OBJECTIVE: to create multilingual and free dictionaries for all known natural languages ▶ Wiktionary
  - 158 languages on August 2014
- The content of a wiktionary is semi-structured ⇒ Requires algorithms to extract the information to be used by NLP applications



# Wiktionary: Wiki + Dictionary (cont'd)

- A screenshot of the list of information for the entry *book*

The screenshot shows the Wiktionary interface for the word "book". On the left is a sidebar with navigation links like "Main Page", "Community portal", and "Tools". The main content area has tabs for "Entry", "Discussion", and "Citations". The word "book" is displayed in a large font. Below it is a "Contents" table of contents with sections for English, Limburgish, and Middle English. The English section is expanded to show sub-sections for pronunciation, etymology, and statistics. At the bottom, there is a "Pronunciation" section with audio players for US and UK English and a list of rhymes.

a multilingual free encyclopedia

**Wiktionary**  
[ˈwɪkʃənəri] n.,  
a wiki-based Open  
Content dictionary

View · Pure | Search

Main Page  
Community portal  
Preferences  
Requested entries  
Recent changes  
Random entry  
Help  
Donations  
Contact us

Tools  
What links here  
Related changes  
Upload file  
Special pages  
Printable version  
Permanent link  
Page information  
Cite this page  
Add definition

In other projects  
Wikipedia

Visibility  
Show other boxes  
Show derived terms  
Show translations  
Show quotations

In other languages   
Afrikaans  
العربية  
Azərbaycanca  
Български

Entry Discussion Citations

## book

**Contents** [hide]

- 1 English
  - 1.1 Pronunciation
  - 1.2 Etymology 1
    - 1.2.1 Noun
      - 1.2.1.1 Synonyms
      - 1.2.1.2 Derived terms
      - 1.2.1.3 Translations
      - 1.2.1.4 See also
    - 1.2.2 Verb
      - 1.2.2.1 Synonyms
      - 1.2.2.2 Derived terms
      - 1.2.2.3 Translations
  - 1.3 Etymology 2
    - 1.3.1 Verb
  - 1.4 Statistics
  - 1.5 Anagrams
  - 1.6 References
- 2 Limburgish
  - 2.1 Noun
- 3 Middle English
  - 3.1 Etymology
  - 3.2 Noun

**English** [edit]

**Pronunciation** [edit]

- enPR: bōk, IPA<sup>(kə)</sup>: /bʊk/
- Audio (US) 0:00 MENU *plural* Audio (US) 0:00 MENU
- Audio (UK) 0:00 MENU
- Rhymes: -ʊk

(SOURCE: [HTTP://EN.WIKTIONARY.ORG/WIKI/BOOK](http://en.wiktionary.org/wiki/book))



# Thesaurus

- A list of words grouped together according to similarity of meaning (synonymy, sometimes antonymy)
- The entry TEACHER in *Roget's Thesaurus*

The screenshot shows the website thesaurus.com with the search term 'teacher'. The page displays the following information:

- Search Bar:** thesaurus.com | teacher
- Navigation:** see definition of teacher, Like, +1, All, show all
- Advertisement:** Keller Graduate School @ www.keller.edu/ Further Your Career & Earn Your MBA At Keller@. Request More Info Now!
- Filters:** Relevance (A-Z), Complexity, Length
- Synonyms for teacher:** Common, Informal, III, III
- Synonyms List:**

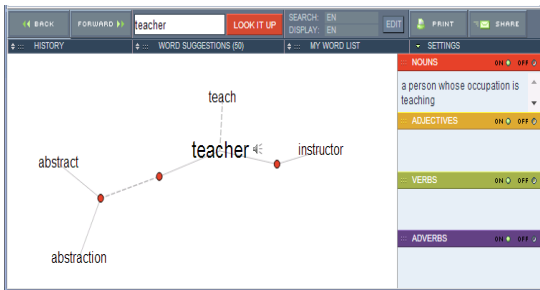
assistant	professor	disciplinarian	teach
coach	scholar	guide	trainer
educator	schoolteacher	mentor	abecedary
faculty member	supervisor	pedagogue	
instructor	tutor	preceptor	
lecturer	adviser	pundit	
- Antonyms for teacher:** pupil, student

(SOURCE: THESAURUS.COM, HTTP://THESAURUS.COM/BROWSE/TEACHER)



# Thesaurus (cont'd)

- The entry TEACHER in a visual thesaurus: *Thinkmap Visual Thesaurus* ▶ Visual Thesaurus



(SOURCE: VISUAL THESAURUS, [HTTP://WWW.VISUALTHESAURUS.COM/TRIALOVER/](http://www.visualthesaurus.com/trialover/))





# Ontology

A formal system that organised the knowledge of a domain into a rigorous and exhaustive way.

The domain is represented with its relevant entities and their related linguistic realisations.

- WordNet is a prototype of terminological ontology
- "In practical terms, developing an ontology includes:
  - Defining classes in the ontology
  - Arranging the classes in a taxonomic hierarchy
  - Defining slots and describing allowed values for these slots
  - Filling in the values for slots for instances" (Noy & McGuinness, 2001)
- An ontology together with a set of individual instances of classes constitutes a **knowledge base**



# Word Sense Disambiguation (WSD)

The task of identifying the correct meaning (or sense) of an ambiguous word in its current context.

- Schütze (1998) has divided the WSD problem into two sub-problems
  - ① **Sense discrimination**
    - "divides the occurrences of a word into a number of classes by determining for any two occurrences whether they belong to the same sense or not"
  - ② **Sense labelling**
    - "assigns a sense to each class, and, in combination with sense discrimination, to each occurrence of the ambiguous word"



# WSD Methods

- Knowledge-based disambiguation
  - Resources: lexical resources such as dictionaries and thesauri
  - Scope: all open class words
- Supervised disambiguation
  - Resources: large sense-annotated corpora
  - Scope: one target word
- Unsupervised disambiguation
  - Resources: unlabelled corpora (no reference to any sense inventory)
  - Typical algorithms: clustering word sense occurrences in the unlabelled corpus
  - Scope: a set of words



# The Lesk Algorithm

- A knowledge-based disambiguation algorithm
- Basic idea: Choose the sense whose dictionary gloss or definition shares the most words with the target word's neighbourhood

## Algorithm

- 1 For each sense  $i$  of  $w_1$
- 2 For each sense  $j$  of  $w_2$
- 3 Compute  $\text{Overlap}(i, j)$ , the number of words in common between the definitions of sense  $i$  and sense  $j$
- 4 Find  $i$  and  $j$  for which  $\text{Overlap}(i, j)$  is maximised
- 5 Assign sense  $i$  to  $w_1$  and sense  $j$  to  $w_2$



## The Lesk Algorithm (cont'd)

- We want to disambiguate the word *bank* in the following sentence:

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

- We get word senses from WordNet [WordNet Online](#)
- Sense bank<sup>1</sup> has two (non-stop) words overlapping with the context in the sentence: *deposits* and *mortgage*, while sense bank<sup>2</sup> has zero, so sense bank<sup>1</sup> is chosen



# References

- Navigli, R. (2012). A Quick Tour of Word Sense Disambiguation, Induction and Related Approaches. M. In Proceedings of SOFSEM 2012, pp. 115-129.
- Noy, N. F. and McGuinness, D. L. (2001). Ontology Development 101: A Guide to Creating Your First Ontology. Stanford University Systems Laboratory Technical Report KSL-01-05, March 2001.
- Schütze, H. (1998). Automatic Word Sense Discrimination. Association for Computational Linguistics. Journal of Computational Linguistics, Vol. 24, pp. 97-123.

