

PA153 Natural Language Processing

02 – Semantics I

(lexical meaning and its representation)

Karel Pala, Zuzana Nevěřilová

NLP Centre, FI MUNI, Brno

Sep 25, 2019

- 1 Lexical Meaning
- 2 Meaning in Context
- 3 Lexical Meanings in NLP
 - Semantic Classes
- 4 Conclusion, Take Home

Lexical Meaning

(cs: lexikální význam): meaning of a word in *isolation* [Oxford Dictionaries, 2013]

- regardless of meaning of the sentence the word is part of
- regardless of grammar categories

other types of meaning: grammatical meaning, word meaning, sentence meaning

- *buy* – *bought*
- *image* – *picture*
- The old professor *runs* to catch the bus. The cheetah *runs* to catch the prey.

└ Lexical Meaning

└ Lexical Meaning

(cs: lexikální význam): meaning of a word in isolation [Oxford Dictionaries, 2013]

- regardless of meaning of the sentence the word is part of
- regardless of grammar categories

other types of meaning: grammatical meaning, word meaning, sentence meaning

- buy – bought
- image – picture
- The old professor runs to catch the bus. The cheetah runs to catch the prey.

buy and bought – the same lexical meaning, different grammatical meaning

image and picture – different words with the same lexical (and grammatical) meaning

to run – the same meaning, different activity

Lexical Form and Lexical Unit

Lexical Unit (cs: lexikální jednotka, LU) [Ziková, 2003]:

- represented by a **lexical form**
- associated with a particular **lexical meaning**
- has grammatical properties (e.g. transitive verb)
- can have pragmatic properties (e.g. *I* each time references to some other person)

- LU with the same meaning but different form
pause **synonyms** (e.g. beautiful, lovely)
- LU with the same form but different meaning
pause **homonyms** (e.g. bark)

Where is information about lexical meaning

Dictionary/lexicon/lexical database – storage of lexical units

Dictionaries:

- general, Language for General Purpose (LGP) also **defining**/explanatory (with definitions)
- bilingual (contains translations of LUs)
- etymological
- encyclopedic
- reverse
- rhyming
- single-field (contains domain terminology)
- historical
- ...



for NLP, machine readable dictionaries are used




Anatomy of a dictionary entry


1 determine - definition and synonyms ★★ **7**



3 VERB  Pronunciation /dɪ'tɜː(r)mn/  **6**

 Contribute to our Open Dictionary


1 [TRANSITIVE] [OFTEN PASSIVE] to control what something will be
Our prices are determined by the market.
genetically/culturally/biologically determined: *She claims that most human behaviour is socially determined.*

 Synonyms and related words

To limit or control something or someone:
draw a line in the sand, limit, control...

Explore Thesaurus

2 [INTRANSITIVE/TRANSITIVE] to officially decide something
determine whether/why/who: *It is for the court to determine whether she is guilty.*

 Synonyms and related words

To make a decision: *decide, determine, arrive at...*

Explore Thesaurus

- 1: headword
- 2: definition
- 3: grammatical category
- 4-5: pronunciation (sound & IPA)
- 6: inflection
- 7: frequency
- 8: numbered senses

<https://www.macmillandictionary.com/learn/dictionary-entry.html>

└ Anatomy of a dictionary entry



Czech Remark on Czech Dictionaries: SSC SSJC: V SSJČ není žádná odvozená forma, ale v SSČ je. Pro jiná slova je v SSČ mnohem více odvozených forem: květ, květen, květena, květák, květenství, květina, květináč, květinářka, květinářství

Collocation as a Dictionary Entry

A language user has available ... a large number of semi-preconstructed phrases that constitute single choices.

(Sinclair 1991: 110)

New York, ad hoc, foreign language, second language, to save time

special collocation dictionaries: Oxford Collocations Dictionary, Macmillan Collocations Dictionary

In NLP, the term multiword expression (MWE) is used when the notion of MWE is useful:

- phrasemes/idioms (imagine translating them)
- the meaning of the components is different/unclear (MWEs as words with spaces)

Dictionary Definitions

- extensional (denotative) definitions
 - ▶ demonstrative definitions (by pointing)
 - ▶ definition by enumeration (e.g. Baltic states are: Estonia, Latvia, and Lithuania)
 - ▶ definition by subclass (e.g. “flower” means rose, lily, daisy, and the like)
- intensional (connotative) definitions
 - ▶ synonymous definition (e.g. “physician” means “doctor”)
 - ▶ etymological definition (e.g. the word “capital” comes from Latin word “caput” meaning “head”)
 - ▶ operational definition (e.g. “brain activity” happens iff an electroencephalograph shows oscillations)
 - ▶ definition by genus and differentia

Definition by Genus and Differentia

a triangle: A plane figure that has 3 straight bounding sides.

a quadrilateral: A plane figure that has 4 straight bounding sides.

hyperonymy

Example: check a Wikipedia definition



└ Definition by Genus and Differentia

Definition by Genus and Differentia

a triangle: A *plane figure* that has 3 straight bounding sides.a quadrilateral: A *plane figure* that has 4 straight bounding sides.

hyperonymy

Example: check a Wikipedia definition

dictionary entries assume at least some knowledge of the language (e.g. English GPL is around 2000 words)

for NLP dictionaries for humans are not fully suitable

Meaning in Context

Lexical meaning is not always enough (in fact, it is not enough **most of the time**)

⇒ know the context

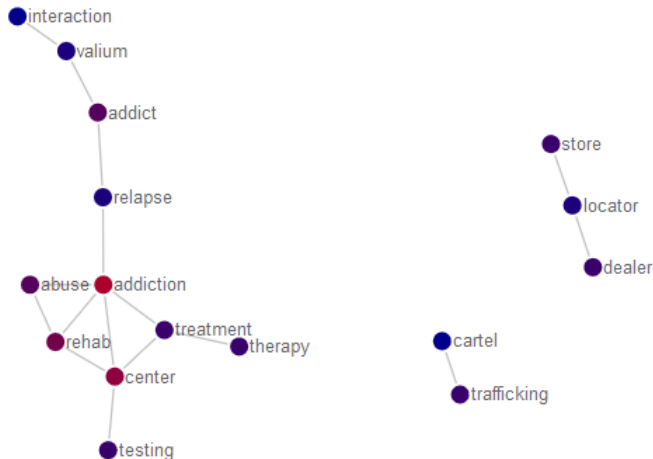
Word Sense Disambiguation (cs: lexikální desambiguace)

function: $(w, c) \rightarrow s$

- $w \in \mathcal{W}$ – set of words
- $c \in \mathcal{C}$ – set of contexts
- $s \in \mathcal{S}$ – set of meanings

Word Sense Disambiguation

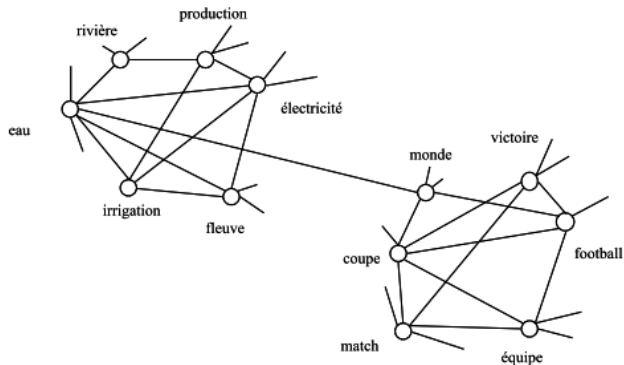
All algorithms rely on a lexical database with discrete meanings . . .
Different dictionaries have different granularity of meanings.
Meanings are not fully discrete.



Word Sense Disambiguation is dead

long live the ...

Word Sense *Discrimination*



[Véronis, 2004]



Componential analysis (komponentová analýza)

= description of meaning using a (small) set of semantic features that are either present, or not present, or irrelevant.

- man = +HUMAN +ADULT +MALE
- woman = +HUMAN +ADULT -MALE
- boy = +HUMAN -ADULT +MALE
- toddler = +HUMAN -ADULT ±MALE

[Katz and Fodor, 1963] a [Bierwisch, 1971]

Componential analysis I

Later related with **semantic primes** and natural semantic metalanguage:

Substantives	I, YOU, SOMEONE, PEOPLE, SOMETHING/THING, BO
Relational Substantives	KIND, PART
Determiners	THIS, THE SAME, OTHER ELSE ANOTHER
Quantifiers	ONE, TWO, SOME, ALL, MUCH/MANY, LITTLE/FEW
Evaluators	GOOD, BAD
Descriptors	BIG, SMALL
Mental predicates	THINK, KNOW, WANT, DON'T WANT, FEEL, SEE, HEA
Speech	SAY, WORDS, TRUE
Actions, Events, Movement	DO, HAPPEN, MOVE
Existence, Possession	BE (SOMEWHERE), THERE IS, BE (SOMEONE/SOMET
Life and Death	LIVE, DIE
Time	WHEN/TIME, NOW, BEFORE, AFTER, A LONG TIME FOR SOME TIME, MOMENT
Space	WHERE/PLACE, HERE, ABOVE, BELOW, FAR, NEA TOUCH (CONTACT)
Logical Concepts	NOT, MAYBE, CAN, BECAUSE, IF
Intensifier, Augmentor	VERY, MORE
Similarity	LIKE/AS/WAY

Semantic Classes

= group words that share a semantic feature

van – truck – motor vehicle – self-propelled vehicle – wheeled vehicle –
vehicle – transport – instrumentation – artifact – whole – object – physical
entity – entity

taxonomy, hierarchy, tree structure

The Tree of Porphyry

Supreme genus:

Substance

Differentiae:

material

immaterial

Subordinate genera:

Body

Spirit

Differentiae:

animate

inanimate

Subordinate genera:

Living

Mineral

Differentiae:

sensitive

insensitive

Proximate genera:

Animal

Plant

Differentiae:

rational

irrational

Species:

Human

Beast

Individuals:

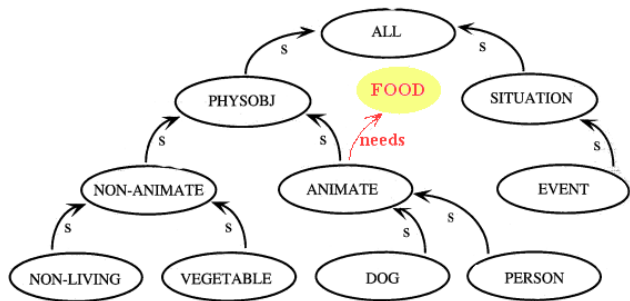
Socrates

Plato

Aristotle

etc.

Semantic Networks, Inference



Semantic Networks

WordNet (Princeton WordNet, PWN) – lexical network

- originally a psychology project (G. A. Miller, od r. 1985)
- usable by humans and computers (NLP) [Fellbaum, 1998]
- basic unit: synonymical set (synset), cs: synonymická řada
- relations between synsets:
 - ▶ hyperonymy/hyponymy: truck – van
 - ▶ holonymy/meronymy (part of, member of): car – brake
 - ▶ troponymy: whisper – speak
 - ▶ near-antonym: day – night
 - ▶ derivation: wide – width
- POS: nouns, adjectives, verbs, adverbs
- each POS organized in a different way

WordNet

size: PWN (117K synsets)

follow-up projects: EuroWordNet (en, nl, it, es, de, fr, cs, et)

- ILI - InterLingual Index
- Top Ontology (63 kategorií)
- Base Concepts

BalkaNet: bg, cs, ro, gr, sr, tr

Global WordNet Association (GWA)

Czech W.: 28K synsets

Ontologies

Lexical network – lexical knowledge

Ontology – knowledge

Ontology = explicit specification of shared conceptualization

- domain ontologies
- general o. SUMO/MILO (Suggested Upper Merged Ontology, Mid-Level Ontology)
- common sense o. ConceptNet

Conclusion, Take Home

lexical meaning description:

- human friendly: dictionaries
- NLP friendly: semantic primes, semantic networks, ontologies

sense disambiguation

- human friendly: numbered senses in dictionaries
- NLP friendly: measures, vectors . . .

References I



Bierwisch, M. (1971).

On classifying semantic features.

In M. Bierwisch, K. E. H., editor, *Progress in Linguistics*, pages 27–50.
Mouton.



Fellbaum, C. (1998).

WordNet: An Electronic Lexical Database (Language, Speech, and Communication).

The MIT Press.

Published: Hardcover.



Katz, J. and Fodor, J. (1963).

The structure of a semantic theory.

Language, (39):170–210.

References II



Oxford Dictionaries (2013).

lexical meaning. Oxford Dictionaries.

online.

<http://oxforddictionaries.com/definition/english/lexical-meaning> (accessed October 03, 2013).



Véronis, J. (2004).

Hyperlex: Lexical cartography for information retrieval.

In *Computer Speech and Language: Special Issue on Word Sense Disambiguation*, page 23.



Ziková, M. (2003).

Současný český jazyk: Tvoření slov.

online.

http://www.phil.muni.cz/cest/lide/zikova/CJA009_1.rtf
(accessed October 03, 2013).