Text Analysis and Retrieval

5. Semantics

Assoc. Prof. Jan Šnajder

With contributions from
Prof. Goran Glavaš and prof. Sebastian Padó

University of Zagreb
Faculty of Electrical Engineering and Computing (FER)

Academic Year 2019/2020

v2.2

Semantics

Computational semantics *(Wikipedia)*

Computational semantics is the study of how to automate the process of constructing and reasoning with meaning representations of natural language expressions.

- **Word-level (lexical) semantics**: meaning of words
- **Sentence-level semantics**: representing the meaning of a sentence comprised of the meaning of its parts
- **Discourse-level semantics**: meaning of text that goes beyond a single sentence (coreferences/anaphors, discourse structures)

We'll look into (some of) word- and sentence-level semantics. . .

Outline

Outline

Learning outcomes 1

1. Define and exemplify polysemy and the main lexical relations
2. Describe WordNet and give an example of synsets involving a polysemous word
3. Describe the purpose and the main approaches to word sense disambiguation
4. Describe frame semantics and FrameNet, and give an example of a frame

Lexical semantics

- Lexical semantics is concerned with the meaning of words
  - What the words mean and how they relate to each other
  - Verb and event semantics (semantic roles)
  - Distributional semantics $\Rightarrow$ we'll cover in part 2
  - ...

Word senses and polysemy

- Lemma vs. wordform vs. sense
  - **lemma** = dictionary/citation form of the word (mouse)
  - **wordform** = the specific form of the word (mouses)
  - **sense** = the meaning of the word
- **Polysemy** = the capacity of a word to have multiple senses
  - Many words are **polysemous** (have many senses)
  - Word senses are listed and defined in dictionaries
- **Homonymy** = the relation between different words that share the same surface form
  - e.g., mouse as a noun and mouse as a verb
  - e.g., saw as a verb (inflected) and saw as a noun
- Multiwords are also often considered as words ("words with spaces")
  - phrasal verbs (throw up, give in, take off)
  - light verbs (make decision, give credit)
  - noun compounds (web browser, black eye)
  - ...

Word senses and polysemy

Mouse – two words (homonyms):

The noun "mouse" (4 senses)

- any of numerous small rodents typically resembling diminutive rats having pointed snouts and small ears on elongated bodies with slender usually hairless tails
- shiner, black eye, mouse (a swollen bruise caused by a blow to the eye)
- person who is quiet or timid
- computer mouse (a hand-operated electronic device that controls the coordinates of a cursor on your computer screen as you move it around on a pad

The verb "mouse" (2 senses)

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

Lexicosemantic relations

Relationship between senses is an important aspect of word meaning!

- **Synonymy** – identical or nearly identical senses ⇒ **synonyms**
  - couch/sofa, car/automobile, mouse/black eye
- **Antonymy** – senses with opposite meanings ⇒ **antonyms**
  - long/short, fast/slow, rise/fall
- **Hypernymy** – one sense is more general ⇒ **hypernyms**
  - vehicle/car, fruit/mango, mammal/dog
- **Hyponymy** – one sense is more specific ⇒ **hyponyms**
  - car/vehicle, mango/fruit, dog/mammal

Similarity and relatedness

- Similarity – two words that have similar meanings
  - cats/dogs, man/woman, give/award
- Relatedness – a more loose association between words
  - coffee/cup, money/work, read/glasses
- Similarity = pragmatic relation, Relatedness = syntagmatic relation

- Similarity vs. relatedness:
  - airplane – machine (S)
  - airplane – engine (S)
  - pilot – airplane (R)
  - magic – disappear (R)
  - rich – caviar (R)

WordNet

- Manually constructed lexical database (Fellbaum, 2005)
  - https://wordnet.princeton.edu/
- Covers nouns, verbs, adjectives, and adverbs
- Words are organized into **synsets** – sets of words with the same sense
- For each synset WordNet provides:
  - a list of words that can be used in that sense
  - a gloss – a short description of the sense
  - semantic relations to other synsets (hyponymy, meronymy, . . .)
- Over 100k synsets for English, coverage is still an issue
- Expensive to build, there are smaller wordnets for other languages

WordNet synsets

- Synsets of the word “Tire”:

- tire.n.01 – hoop that covers a wheel
- tire.v.01 – lose interest or become bored with something or somebody
- tire.v.02 – exhaust or get tired through overuse or great strain
- run_down.v.06 – deplete
- bore.v.01 – cause to be bored

WordNet relations

Nouns:

- **Hyperonymy/hyponymy** – IS-A relation (chair – furniture)
- **Meronymy** – a part whole relation (finger – hand)
- **Antonymy** – opposite meaning (wet – dry)
- **Similarity** – similar (but not identical) meaning (warm – hot)

Verbs:

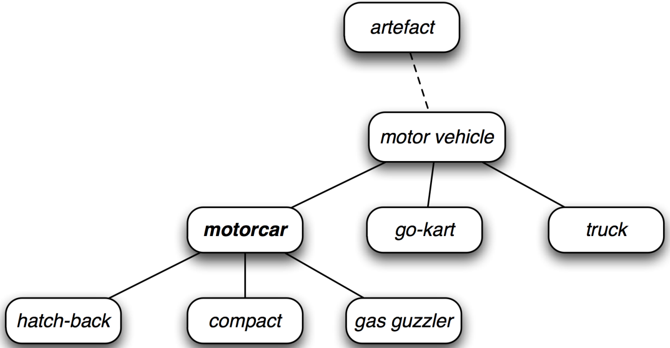
- **Troponymy** – increasingly specific manner of an event (communicate – talk - whisper, move – jog - run)
- **Entailment** – one word entails the other (succeed – try, buy – pay)

WordNet relations

- Assuming we mean a wheel tire (tire.n.01)
- Hypernyms of the word "Tire":
  - hoop.n.02 – a rigid circular band of metal or wood or other material used for holding or fastening or hanging or pulling
- Hyponyms of the word "Tire":
  - wagon_tire.n.01 – a metal hoop forming the tread of a wheel
  - pneumatic_tire.n.01 – a tire made of reinforced rubber and filled with compressed air; used on motor vehicles and bicycles etc
  - car_tire.n.01 – a tire consisting of a rubber ring around the rim of an automobile wheel

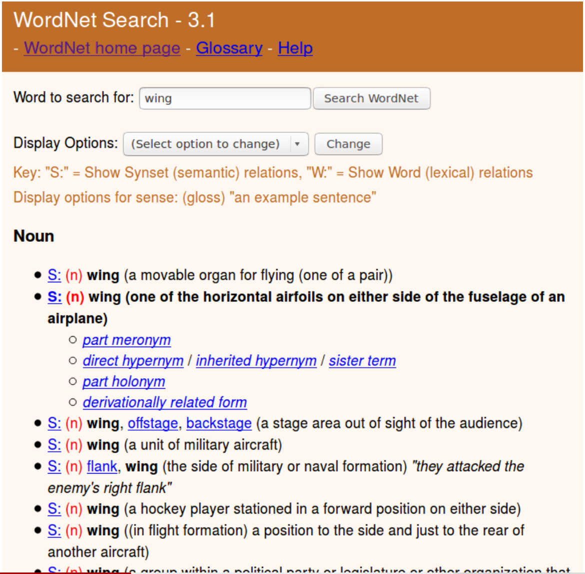
WordNet hierarchy

- Example of the hierarchy from WordNet:



WordNet search

https://wordnet.princeton.edu/



Word sense disambiguation

- **Word-sense disambiguation** (WSD) is the task of identifying which sense (meaning) of a polysemous word is used in a sentence

Distinct senses of the word

The newspaper fired the editor. (The company/organization)
John spilled coffee on the newspaper. (The physical newspapers)

- WSD addresses both polysemy and homonymy
  - Polysemy – a word has multiple, related meanings, e.g. *ring* (wedding ring vs. boxing ring)
  - Homonymy – two unrelated words have the same form, e.g., *saw* (past tense of see vs. a tool)
- The more fine-grained the senses, the more difficult the disambiguation task

WSD approaches

1. **Dictionary-based methods:**
   - Rely on lexical resources, most often WordNet
2. **Supervised WSD**
   - One classifier for each polysemous word
   - Manual sense annotations in text, very time- and resource-consuming
3. **Unsupervised WSD**
   - Clustering different contexts in which the word appears
   - Ideally, each cluster corresponds to one sense
   - Also called word sense induction (WSI)

Word Sense Disambiguation – tools

- University of North Texas WSD Tools:
  lit.csci.unt.edu/index.php?P=research/downloads
  Unsupervised graph-based WSD
  SenseLearner: All-Words Word Sense Disambiguation Tool
- KYOTO UKB graph-based WSD
  http://ixa2.si.ehu.es/ukb/
- pyWSD (Simple WSD algorithms in Python)
  https://github.com/alvations/pywsd

Frame semantics

- The meanings of verbs are more difficult to classify/systematize Why?
- Nouns describe **objects**, verbs describe **events**
- Objects can be classified into a hierarchy based on their type
- Verbs: type + temporal dimension + arguments
  1. Requires a larger set of relations
     - Entailment, Event/Subevent, Causation,...
     - No clear single hierarchy!
  2. Requires a level of description for arguments

Frame semantics

- Theory of frame semantics (Fillmore, 1977) explains language via communication processes rather than grammar and orthography
  - Case grammars (Fillmore, 1967) assume that the sentence creation is guided by semantic relations between concepts rather than syntactic relations between phrases
- Predicates define frames and frames define arguments with their semantic roles
- Frames are listed in resources such as FrameNet (Baker et al., 1998)

FrameNet frame

[$_{VICTIM}$ A one-year-old baby] was **snatched** [$_{SOURCE}$ from a shopping centre][$_{TIME}$ last night].
[$_{PERPETRATOR}$The thief] **snatched** [$_{GOODS}$ a 32 million dollar worth Dalí painting][$_{SOURCE}$ from Louvre].

FrameNet

https://framenet.icsi.berkeley.edu/

Hiring

Lexical Unit Index

Definition:

An Employer hires an Employee, promising the Employee a certain Compensation in exchange for the performance of a job. The job may be described either in terms of a Task or a Position. In some cases, the Employee FE will also indicate the Position (see fourth example below).

John was HIRED to clean up the file system.

IBM HIRED Gates as chief janitor.

I was RETAINED at $500 an hour.

The A's SIGNED a new third baseman for $30M.

The same sentence (above) should also have the FE Position on the second layer:

The A's SIGNED a new third baseman for $30M.

FEs:

Core:

Employee [Empee] — The person whom the Employer takes on as an Employee, obligating them to perform some Task in order to receive Compensation.
I was just HIRED yesterday!

Employer [Emper]
Semantic Type: Sentient — The person (or institution) that takes on an Employee, giving them Compensation in return for the performance of an assigned Task.
Last month, IBM HIRED Mike Zisman to head up its storage software group.

Field [Field] — The Field that the Employee is going to work in for their Employer.
It's not easy to get HIRED in academia.

Position [Posit] — The label given to a particular type of employment.
Look, I wasn't HIRED as your waitress!

Task [Task] — The action that the Employee is taken on by the Employer to do.
I was HIRED just to empty the trash cans.

Non-Core:

Compensation [Compense] — The Compensation is the payment that the Employee is set to receive for performing an assigned Task.
They fired our management, HIRED him for 20% more and gave him a free office to set up his own company.

Frame semantics

Today's reading:

https://www.fer.unizg.hr/_download/repository/TAR-2020-reading-04.pdf

Learning outcomes 1 – CHECK!

1. Define and exemplify polysemy and the main lexical relations
2. Describe WordNet and give an example of synsets involving a polysemous word
3. Describe the purpose and the main approaches to word sense disambiguation
4. Describe frame semantics and FrameNet, and give an example of a frame

Discussion points

1. What NLP task could WordNet be useful for?
2. Could WordNet information be used to build features?
3. What NLP tasks could similarity/relatedness be useful for?
4. How could WordNet be used to compute similarity/relatedness between two words?
5. Does each verb uniquely define a FrameNet frame?
6. Are FrameNet frames only linked to verbs? Why?
7. What are the disadvantages of manually labeled resources such as WordNet/FrameNet?
8. Could these resources be built automatically? How?

Outline

Learning outcomes 2

1. State the distributional hypothesis and give an example
2. Explain what a distributional semantic model is, how it's constructed, and what it's used for
3. Differentiate between sparse/dense and count-based/predictive vector representations
4. Define distributional semantic composition and the simplest approach to it

Distributional semantics

- Representation of word meaning based on **distributional hypothesis**:

Distributional hypothesis (Harris, 1954)

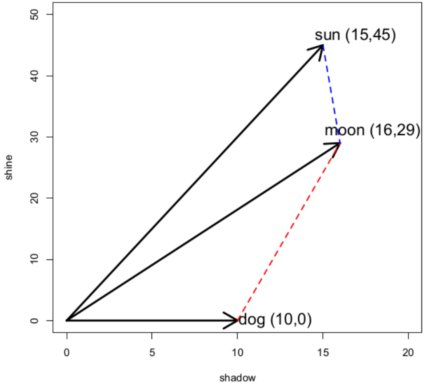
Words that occur in similar contexts tend to have similar meanings

$\Rightarrow$ correlation between context similarity and meaning similarity

- Operationally, this means that, if we can measure the similarity of contexts, we can measure the similarity of meanings
- **Distributional semantics** – represents each word via a distribution of words (or other elements) in its context
- **Distributional semantic model (DSM)** – represents words as vectors of context features obtained from corpus
- Semantic similarity predicted via vector similarity

Distributional semantic models

	planet	night	full	shadow	shine	crescent
moon	10	22	43	16	29	12
sun	14	10	4	15	45	0
dog	0	4	2	10	0	0



Distributional semantics

- Sparse vs. dense
- Count-based vs. predictive

- Count-based vs. predictive. . . which are better?
  - Baroni et al. (2014). **Don't count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors**. ACL 2014
    http://www.aclweb.org/anthology/P14-1023
  - Levy and Goldberg (2014). **Neural word embedding as implicit matrix factorization**. NIPS 2014
    http://papers.nips.cc/paper/5477-neural-word-embedding-as-implicit-matrix-factorization.pdf

Word representations

- Words are discrete objects. We have to figure out a way to represent them in a feature vector for a machine learning model
- Options:
  (1) **One-hot representation**
  (2) **Distributional vectors** (count-based)
  (3) **Word-embeddings** (aka "distributed representations")
    (3a) Dimension-reduced count-based vectors
    (3b) Trained using a neural network
- (1) and (2) give "sparse representations" ⇒ large vectors with many zero elements
- (3) gives "dense representations" ⇒ small vectors filled with real-valued numbers

Pretrained dense representations

- word2vec
  https://code.google.com/archive/p/word2vec/
- Pretrain models for EN:
  https://github.com/mmihaltz/word2vec-GoogleNews-vectors
- 30+ other languages:
  https://github.com/Kyubyong/wordvectors
- Alternatives:
  - GloVe (https://nlp.stanford.edu/projects/glove/)
  - fastText (https://fasttext.cc/) – uses character n-grams
  - ...

word2vec: results

Airplane

word	cosine
plane	0.835
airplanes	0.777
aircraft	0.764
planes	0.734
jet	0.716
airliner	0.707
jetliner	0.706

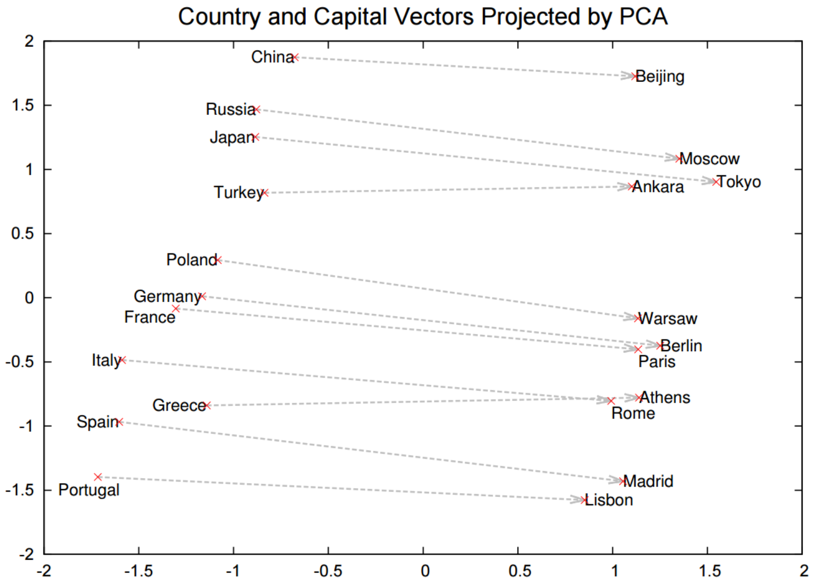
Cat

word	cosine
cats	0.810
dog	0.761
kitten	0.746
feline	0.732
puppy	0.707
pup	0.693
pet	0.689

Dog

word	cosine
dogs	0.868
puppy	0.811
pit_bull	0.780
pooch	0.763
cat	0.761
pup	0.741
canines	0.722

Word2Vec: results



Semantic composition

- **Semantic composition** = composing the meaning of phrases, sentences or longer text fragments from the meaning of individual words
  - "red" + "apple" = "red apple"
- Language is mostly compositional (we can express an infinitude of meanings using a finite set of words)
- How to automate semantic composition to represent the meaning of text fragments?
- We need this for text classification, comparing text similarity, etc. ... basically, all tasks that work with text fragments!

Distributed vectors and semantic composition

- The simplest approach: **add up the vectors**!
  - a continuous version of BOW, hence the name CBOW
  - use as features for (shallow) ML models (SVM, LR, etc.)
  - widespread practice and works remarkably well!
- Models/vectors specifically tuned for semantic composition:
  - doc2vec in gensim

    https://radimrehurek.com/gensim/models/doc2vec.html
  - "Universal Paraphrastic Sentence Embeddings"

    https://github.com/jwieting/iclr2016
  - skip-thought vectors

    http://papers.nips.cc/paper/5950-skip-thought-vectors.pdf
- Use CNN/RNN/LSTSM and let the model learn composition implicitly

Learning outcomes 2 – CHECK!

1. State the distributional hypothesis and give an example
2. Explain what a distributional semantic model is, how it's constructed, and what it's used for
3. Differentiate between sparse/dense and count-based/predictive vector representations
4. Define distributional semantic composition and the simplest approach to it

Discussion points

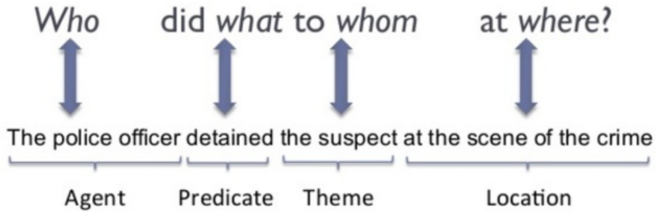
1. Does it make sense to lemmatize before building a DSM? Why?
2. What is the advantage of a distributional vector vs. one-hot encoding?
3. What is the advantage of dense representations over sparse representations?
4. What are the advantages of DSMs over WordNet?
5. What are the disadvantages of DSMs over WordNet?
6. Will semantic composition always work? (Consider: “cloud nine”, ”burn the bridges”, etc.)
7. How to deal with unknown (out-of-vocabulary) words?
8. Could we somehow incorporate syntactic information into a DSM? What for and how?

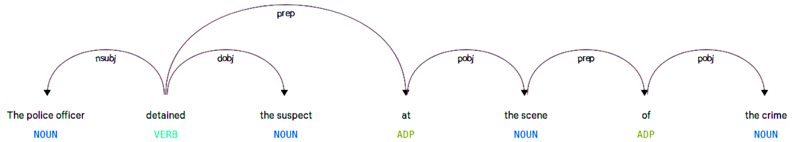
Outline

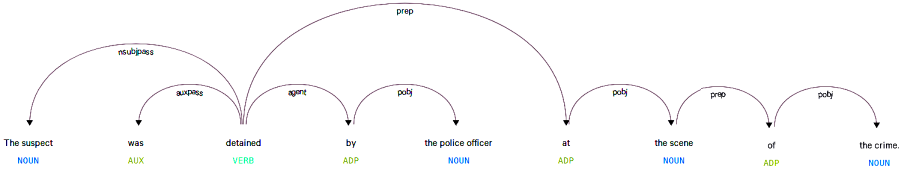
Learning outcomes 3

1. Define semantic parsing and explain how it differs form syntactic parsing
2. Differentiate between deep and shallow semantic parsing
3. Define semantic role labeling and how it can be framed as a machine learning task
4. Differentiate between FrameNet and PropBank semantic roles

Syntactic vs. semantic parsing





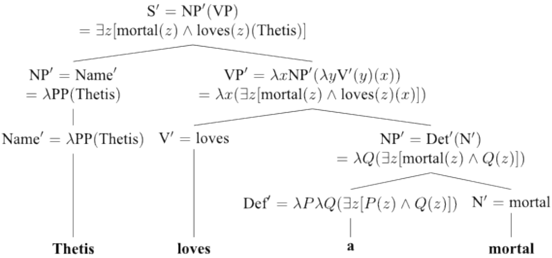


Semantic parsing

- **Semantic parsing** – mapping natural-language sentences (or, generally, text) to a formal representation of meaning
  - **Deep semantic parsing** – mapping to a **complete** meaning representations (including negation, quantifiers, determiners, etc.), which has a rich ontology of types and supports automated reasoning
  - **Shallow semantic parsing** aka **Semantic role labeling (SRL)** – identifying the main semantic roles (constructing the predicate-argument structure)
- Both deep and shallow approaches rely on some formal theory that defines the meaning representations
  - Abend, O., & Rappoport, A. (2017). **The State of the Art in Semantic Representation**. ACL 2017
    http://www.cs.huji.ac.il/~oabend/papers/sem_rep_survey.pdf

Deep parsing ⇒ Formal semantics

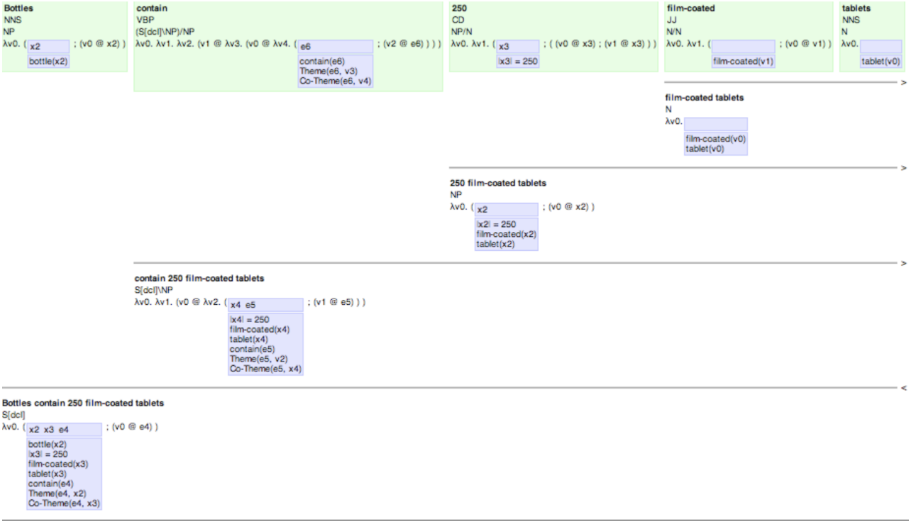
- **Formal semantics:** Traditional approach to natural language semantics, focused at sentence-level and discourse-level semantics
- **Montague grammar:** based on predicate logic and lambda calculus, constructs predicate formulas based on parse trees:



- Example: Groningen Meaning Bank – a semantically annotated corpus (http://gmb.let.rug.nl)
- Semantically brilliant, but of limited use for practical TAR

Groningen Meaning Bank

https://gmb.let.rug.nl/



Shallow parsing

Today's reading:

https://www.fer.unizg.hr/_download/repository/TAR-2020-reading-04.pdf

Learning outcomes 3 – CHECK

1. Define semantic parsing and explain how it differs form syntactic parsing
2. Differentiate between deep and shallow semantic parsing
3. Define semantic role labeling and how it can be framed as a machine learning task
4. Differentiate between FrameNet and PropBank semantic roles