# Mining Unstructured Data
# 4. Lexical semantics

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Facultat d'Informàtica de Barcelona

# Outline

# Semantics

Semantics deals with the meaning:

- Lexical semantics: deals with the meaning of individual words
- Compositional semantics: deals with the construction of meaning usually in high concordance with syntax

This session focuses on lexical semantics

Dependency and Constituent parsing are part of compositional semantics, they deal mainly with syntax but can be useful for semantic parsing

# Outline

# Motivation of lexical semantics

Some examples of usefulness:

- Discovery of semantic patterns
  - Ex: USA bombed Hiroshima
    They began to bombard the defenses
    → A sense_12533 B
- Determine discourse relations
  - Ex: [Anna will show up later.] [She has missed the train.] → explanation
  - Ex: [Mathew is good cooking.] [Albert fails making every dish] → contrast
- Twitter sentiment analysis
  - Ex: @vooda1: CNN Declines to Air White House Press Conference Live YES! THANK YOU @CNN FOR NOT LEGITIMI...
    positive
  - Ex: @Slate: Donald Trump's administration: "Government by the worst men."
    negative

# Outline

# Resources of lexical semantics



- Knowledge-based resources: represented as graphs

  Ex: WordNet (English lexical ontology)
  SentiWordNet (sentiment polarity into WordNet)
  BabelNet (Wikipedia+WordNet)
  VerbNet (syntactic/semantic verbal behaviour)
  FrameNet (conceptual behaviour –fine-grained event representation–)
  ConceptNet (common sense knowledge)

# Resources of lexical semantics



- Knowledge-based resources: represented as graphs

  Ex: WordNet (English lexical ontology)
  SentiWordNet (sentiment polarity into WordNet)
  BabelNet (Wikipedia+WordNet)
  VerbNet (syntactic/semantic verbal behaviour)
  FrameNet (conceptual behaviour –fine-grained event representation–)
  ConceptNet (common sense knowledge)

- Corpus-based resources: contextual usage of words

  Ex: Latent Semantic Analysis (LSA)
  Word embeddings (word2vect, glove, fasttext, ...)
  Contextual word embeddings as compositional semantics (BERT, RoBERTA, GPT3, ...)

# Resources of lexical semantics



| | |
|---|---|
| WordNet | `https://wordnet.princeton.edu/` |
| SentiWordNet | `https://github.com/aesuli/SentiWordNet` |
| BabelNet | `https://babelnet.org/` |
| VerbNet | `https://verbs.colorado.edu/verbnet/` |
| FrameNet | `https://framenet.icsi.berkeley.edu/fndrupal/` |
| LSA | accessible from |
| Word embeddings | `https://radimrehurek.com/gensim/` |

# Outline

# Outline

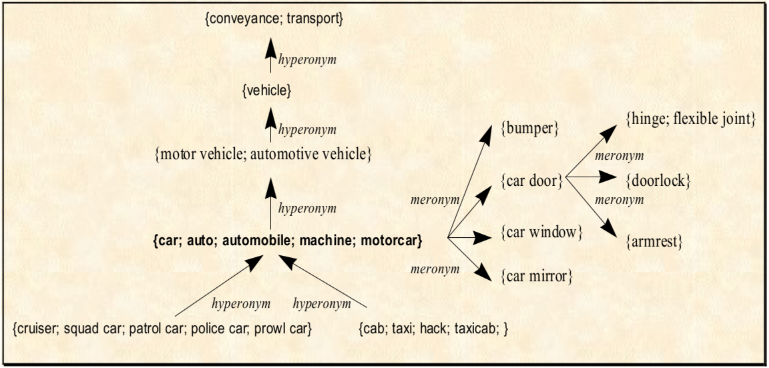
# WordNet

- Free large lexical database of English
- Contains only nouns, verbs, adjectives and adverbs
- Words are grouped into synonyms sets (*synsets*)
- each *synset* has an associated gloss and some examples
- *synsets* are interlinked by means of lexical relations
  https://en-word.net/lemma/demo

# Lexical relations



## Example of Lexical Relation Net



# Lexical relations

- Synonymy: same meaning. Ex: age - historic_period
- Antonymy: opposite meaning. Ex: dark - light
- Homophome: same sound. Ex: son - sun
- Homograph: same written form. Ex: lead (noun - verb)
- Polysemy: different related meaning. Ex: newspaper (paper - firm)
- Homonymy: different unrelated meaning. Ex: position (place - status)
- Hypernymy: parent. Ex: cat - feline
- Hyponymy: child. Ex: feline - cat
- Holonym: group, whole. Ex: student - class
- Meronym: member, part. Ex: class - student
- Metonymy: substitution of entity. Ex: We ordered many delicious dishes at the restaurant.

# Outline

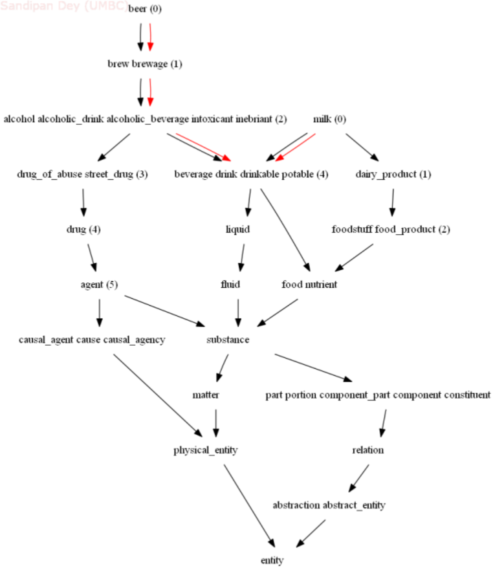
# Similarities in WordNet

- Shortest Path Length: $Sim(s_1, s_2) = \frac{1}{SPL(s_1,s_2)}$
  where $SPL(s_1, s_2)$ = Shortest Path Length from $s_1$ to $s_2$ as vertex-countings conto i vertici non gli archi
- Leacock & Chodorow: $Sim(s_1, s_2) = -log_2 \frac{SPL(s_1,s_2)}{2 \cdot MaxDepth}$
  where $depth(s)$ = SPL(TopSynset,s) numero di bit necessari
  $MaxDepth = \max_{s \in WN} depth(s)$
- Wu & Palmer: deepest the common ancestor they have more similarity
  $Sim(s_1, s_2) = \frac{2 \cdot depth(LCS(s_1,s_2))}{depth_{LCS(s_1,s_2)}(s_1) + depth_{LCS(s_1,s_2)}(s_2)}$
  where $LCS(s_1, s_2)$ = Lowest Common Subsumer of $s_1$ and $s_2$
  $depth_{s'}(s)$ = SPL(TopSynset,s) ~~throw~~ s' through
- Lin: $Sim(s_1, s_2) = \frac{2 \cdot IC(LCS(s_1,s_2))}{IC(s_1) + IC(s_2)}$
  where $IC(s) = -log_2 P(s)$ = information content of s (from frequencies in a corpus)

# Example / exercise








$spl(beer, milk) = 5$
$Sim_{spl}(beer, milk) = 0.2$

$Sim_{wp}(beer, milk) = 0.75$

$Sim_{spl}(drug, milk)?$
$Sim_{wp}(drug, milk)?$

# Outline

# Definition

Extension of wordnet that adds for each synset 3 measures:

- positive_score
- negative_score
- objective_score = 1 - positive_score - negative_score

| Wordnet | | SentiWordnet | | |
|---|---|---|---|---|
| Antonym Synsets | Gloss | obj | pos | neg |
| bad.a.01 | having undesirable or negative qualities | 0.375 | 0.0 | 0.625 |
| good.a.01 | having desirable or positive qualities... | 0.25 | 0.75 | 0.0 |
| bad.n.01 | that which is below standard or expectations as of ethics or decency | 0.125 | 0.0 | 0.875 |
| good.n.03 | that which is pleasing, valuable, useful | 0.375 | 0.625 | 0.0 |

# Outline

# Outline

# Sentiment analysis



Different subtasks:

- Opinion detection: given a piece of text (document or sentence), is it an objective text or a subjective one?
- Polarity classification: given a subjective piece of text, is it a positive opinion or a negative one?
- Opinion extraction: given a subjective piece of text, recognise the focuses of the opinion (templates <entity, aspect, polarity>).

# Outline

# Unsupervised sentiment analysis

Possible simple solution with lexical information:

$$h(D) = \sum_{w \in \hat{D}} word\_score(w) \qquad word\_score(w) = 1/|S(w)| * \sum_{s \in S(w)} score(s))$$

$\hat{D}$ is usually the set of adjectives, or nouns and adjectives, or nouns, verbs, adjectives and adverbs. $S(w)$ is the set of synsets for word $w$.

- Opinion detection:

$$score(s) = 1 - obj_s \quad or \quad score(s) = obj_s$$

- Polarity classification:

$$score(s) = pos_s - neg_s$$

Pros:

- no need for training corpora

Cons:

- low results
- need for POS tagger

# Supervised sentiment analysis

Possible simple solution with lexical information:

Bag of words with Naïve Bayes

$$h(D) = h(w_1, \dots, w_n) = \operatorname*{argmax}_y P(y) \prod_{i=1}^{n} P(w_i|y)$$

where $y$ is the category (positive/negative, subjective/objective), and $w_1, \dots, w_n$ is the bag of words related to $D$

- Given a training corpus $C = \{d_i\}$ partitioned into subsets $Y_1$ and $Y_2$
  - $P(y) \approx P_{MLE}(y) = \frac{|Y_i|}{|C|}$
  - $P(w_i|y) \approx P_{MLE}(w_i|Y_j) = \frac{c(w_i, Y_j)}{\sum_{w_i \in Y_j} c(w_i, Y_j)}$

Pros:

- higher results
- no need for POS tagger

Cons:

- need for training corpora

# Hybrid approach for sentiment analysis



Possible solution with lexical information:

- Combine two supervised methods with SentiWordnet method
- I.e., consensuate the output of the three methods, using *voting*, for instance:

  if at least 2 of the methods answer $y$ then output $y$
  else output the answer of the method with better accuracy in the training corpus

The combination improves the results of the isolated methods

# Annex



- Base on the Bayes' theorem:

$$P(y|x_1, \dots, x_n) = \frac{P(y)P(x_1, \dots, x_n|y)}{P(x_1, \dots, x_n)}$$

- Naïve assumption of independence between features:

$$P(y|x_1, \dots, x_n) \approx P(y) \prod_{i=1}^{n} P(x_i|y)$$

- *Maximum likelihood estimation* of $P(y)$ and $P(x_i|y)$ as training model
- Test prediction as:

$$h(x_1, \dots, x_n) = \operatorname*{argmax}_y P(y) \prod_{i=1}^{n} P(x_i|y)$$

- Need a smoothing technique to avoid zero counts: in NLTK never seen features are discarded