Natural Language Processing

Part 4: lexical semantics

Lexical semantics

- A **lexicon** generally has a highly structured form
  - It stores the meanings and uses of each word
  - It encodes the relations between words and meanings
- A **lexeme** is the minimal unit represented in the lexicon
  - It pairs a stem (the orthographic/phonological form chosen to represent words) with a symbolic form for meaning representation (**sense**)
- A **dictionary** is a kind of lexicon where meanings are expressed through definitions and examples

son noun
a boy or man in relation to either or both of his parents.
• a male offspring of an animal.
• a male descendant : *the sons of Adam*.
• (the Son) (in Christian belief) the second person of the Trinity; Christ.
• a man considered in relation to his native country or area : *one of Nevada's most famous sons*.
• a man regarded as the product of a particular person, influence, or environment : *sons of the French Revolution*.
• (also my son) used by an elder person as a form of address for a boy or young man : "*You're on private land, son.*"

Lexicons & dictionaries

- Definitions in dictionaries exploit words and they may be circular (a word definition uses words whose definitions exploit that word)

right adj.
1. of, relating to, situated on, or being the side of the body which is away from the side on which the heart is mostly located
2. located nearer to the **right** hand than to the left
3. done with the **right** hand

.....

  - The paradox is that the dictionary elements are not direct definitions
    - The are description of the lexemes made up of other lexemes assuming that the user has enough information on these other terms!
    - This approach would fail without the assumption that the user has already enough a priori knowledge deriving from the real world
    - However the description provide a great amount of information on the relationships among the words allowing to perform semantic inferences

Relationships & senses

- Several kinds of relationships can be defined between lexemes and senses (some of them are important for automatic processing)
  - **Homonymy**
    It is a relation between words that have the same form (and the same PoS) but unrelated meanings
    - e.g. *bank* (the financial institution, the river bank)
    - It causes ambiguities for the interpretation of a sentence since it defines a set of different lexemes with the same orthographic form ($bank^1$, $bank^2$,..)
    - Related properties are homophony (same pronunciation but different orthography, e.g. *be-bee*) and homography (same orthography but different pronunciation pésca/pèsca)
  - **Polysemy**
    It happens when a lexeme has more related meanings
    - It depends on the word etymology (unrelated meanings usually have a different origin) - e.g. *bank/data bank/blood bank*

Polysemy/Synonymy

For polysemous lexemes we need to manage all the meanings

- We should define a method to determine the meanings (their number and semantics) and if they are really distinct (by experts in lexicography)
- We need to describe the eventual correlations among the meanings
- We need to define how the meanings can be distinguished in order to attach the correct meaning to a word in a given context (word sense disambiguation)

- **Synonymy**
  It is a relationship between tow distinct lexemes with the same meaning (i.e. they can be substituted for one another in a given context without changing its meaning and correctness) – e.g. I received a gift/present
  - The substitutability may not be valid for any context due to small semantic differences (e.g. *price/fare of a service – the bus fare/the ticket price*)
  - In general substitutability depends on the "semantic intersection" of the senses of the two lexemes and, in same cases, also by social factors (*father/dad*)

Hyponymy/Hypernymy

- **Hyponymy** is a relationship between two lexemes (more precisely two senses) such that one denotes a subclass of the other
  - car, vehicle – shark, fish – apple, fruit
  - The relationship is not symmetric
    - The more specialized concept is the hyponym of the more general one
    - The more general concept is the hypernym of the more specialized one
  - Hyponymy (hypernymy) is the basis for the definition of a taxonomy (a tree structure that defines inclusion relationships in an object ontology) even if it is not properly a taxonomy
    - The definition of a formal taxonomy would require a more uniform/rigorous formalism in the interpretation of the inclusion relationship
    - However the relationship defines a inheritance mechanism of the properties from the ancestors of a given a concept in the hierarchy

Wordnet

- It is a lexical database for English (versions for other languages are available) organized as a **semantic network of senses**
  - It represents nouns, verbs, adjectives, and adverbs but it does not include functional terms in the closed classes (prepositions, conjunctions, etc.)
  - The lexemes are grouped into sets of cognitive synonyms (synset), each representing a distinct concept
  - A set of senses (synset) is associated to each lexeme (unique orthographic form)
  - Synsets are linked by conceptual/semantic and lexical relationships
  - Wordnet consists in lexicographic files, an application to load these files into a database and a library of search and browsing functions to visualize and access the database contents

Wordnet Statistics

PoS	Unique strings	Synset	pairs word-sense
Noun	117,798	82,115	146,312
Verb	11,529	13,767	25,047
Adjective	21,479	18,156	30,002
Adverb	4,481	3,621	5,580
Total	155,287	117,659	206,941

- Nouns have an average of 1.24 senses, verbs 2.17, adjectives 1.40, adverbs 1.25
- The actual total number of distinct stings is 147,278 (the same string can belong to more than one PoS class)

Synset

- A synset is a set of synonyms that define a concept or word meaning
  - About half of the synsets (~54%) contains only one term, about one third (~29%) 2 terms, about 10% 3 terms
  - An annotation (gloss) explaining the meaning is associated to each synset (especially to those containing a single term)
    - A synset contains ~1.75 terms in average)

2 senses of teacher

Sense 1

synset **teacher#1, instructor#1** -- (a person whose occupation is teaching)
=> educator#1, pedagogue#1, pedagog#1 -- (someone who educates young people)

Sense 2

synset **teacher#2** -- (a personified abstraction that teaches; "books were his teachers"; "experience is a demanding teacher")
=> abstraction#1, abstract#1 -- (a concept or idea not associated with any specific instance; "he loved her only in the abstract--not in person")

Synset – verb example

5 senses of **derive**

Sense 1
deduce#1, infer#1, deduct#3, derive#1 -- (reason by deduction; establish by deduction)
=> reason#1, reason out#1, conclude#1 -- (decide by reasoning; draw or come to a conclusion; "We reasoned that it was cheaper to rent than to buy a house")

Sense 2
derive#2, gain#1 -- (obtain; "derive pleasure from one's garden")
=> obtain#1 -- (come into possession of; "How did you obtain the visa?")

Sense 3
derive#3 -- (come from; "The present name derives from an older form")
=> evolve#2 -- (undergo development or evolution; "Modern man evolved a long time ago")

Sense 4
derive#4, educe#2 -- (develop or evolve from a latent or potential state)
=> make#3, create#1 -- (make or cause to be or to become; "make a mess in one's office"; "create a furor")

Sense 5
derive#5, come#18, descend#2 -- (come from; be connected by a relationship of blood, for example; "She was descended from an old Italian noble family"; "he comes from humble origins")

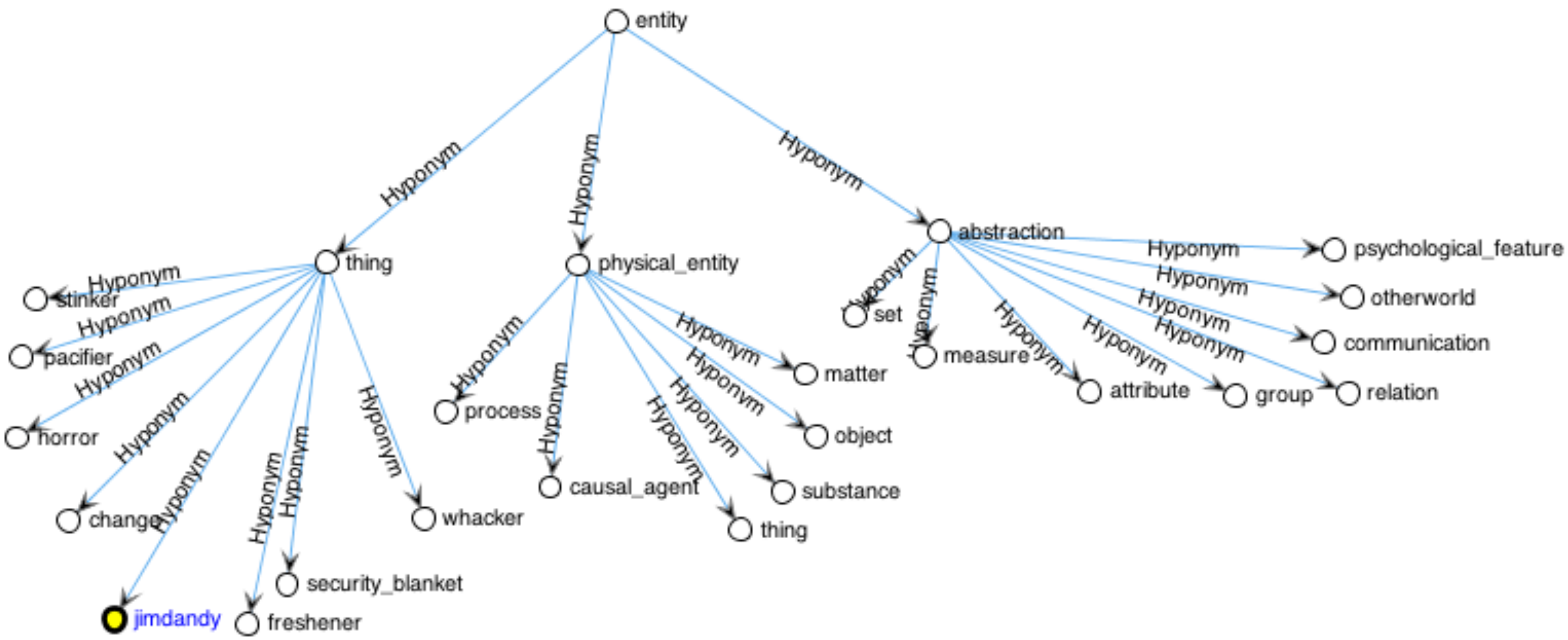
Names

- Names are organized in a hierarchy of specializations (hyponyms) and generalizations (hypernyms)
  - In the 3.0 version there is a unique root category {entity} referred to as **unique beginner** whereas in the previous versions there are mode unique beginners (25 in version 1.7.1)

vase#1 -- (an open jar of glass or porcelain used as an ornament or to hold flowers)
=> **jar#1** -- (a vessel (usually cylindrical) with a wide mouth and without handles)
=> **vessel#3** -- (an object used as a container (especially for liquids))
=> **container#1** -- (any object that can be used to hold things (especially a large metal boxlike object of standardized dimensions that can be loaded from one form of transport to another))
=> **instrumentality#3, instrumentation#1** -- (an artifact (or system of artifacts) that is instrumental in accomplishing some end)
=> **artifact#1, artefact#1** -- (a man-made object taken as a whole)
=> **whole#2, unit#6** -- (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#1, physical object#1** -- (a tangible and visible entity; an entity that can cast a shadow; "it was full of rackets, balls and other objects")
=> **physical entity#1** -- (an entity that has physical existence)
=> **entity#1** -- (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))

Name hierarchy

- The first three levels of the name hierarchy starting from the unique beginner {entity} (root) in Wordnet 3.0



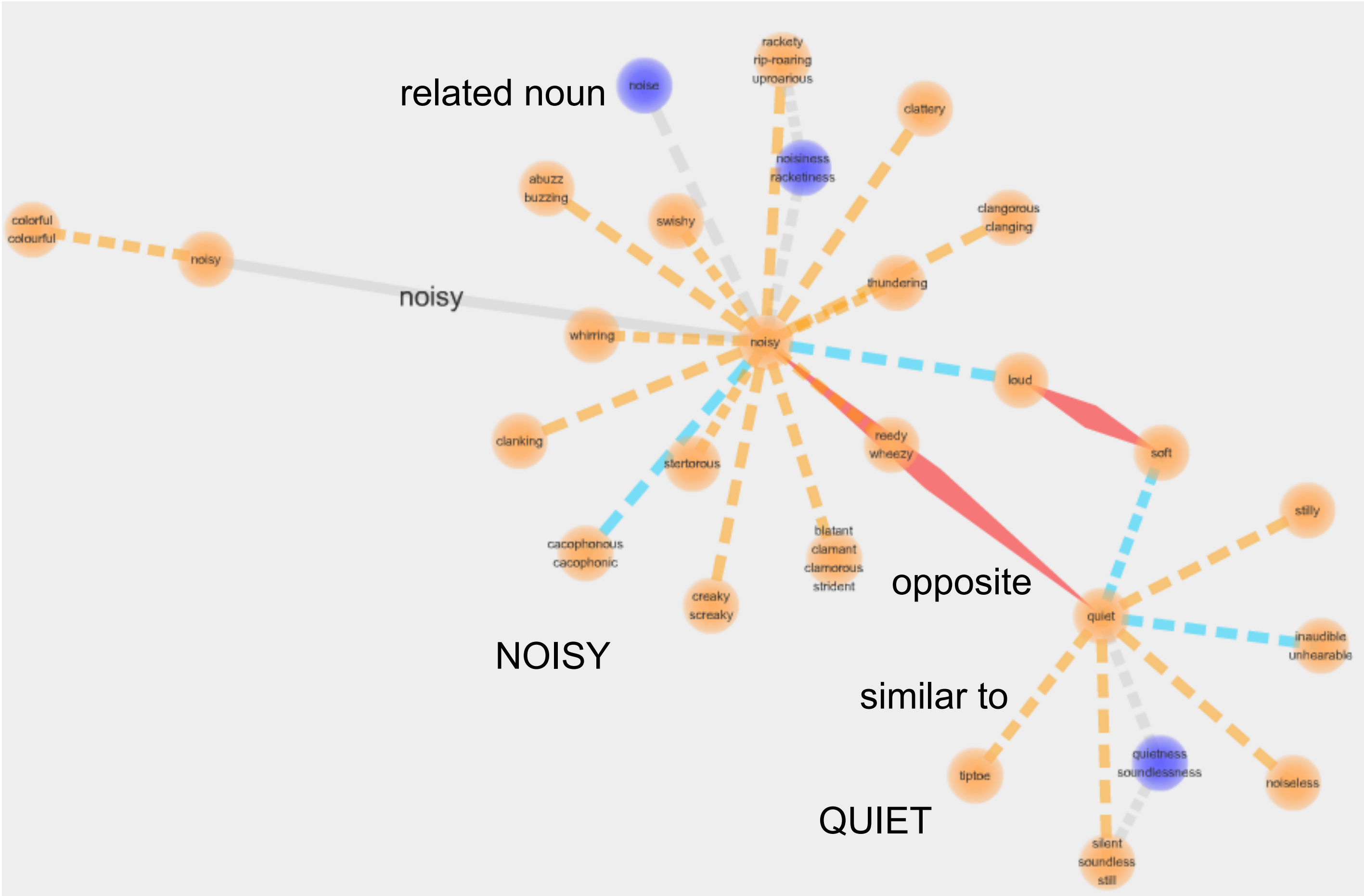
Wordnet relationships– names & verbs

- For nouns the following relationships are provided among synsets
  - Hyperonymy - X is a kind of Y (car → vehicle)
  - Hyponymy – Y is a kind of X (vehicle → car)
  - Coordinate terms– Y is a coordinate term of X is X and Y share a common hyperonym (car and motorcycle)
  - Holonymy – X is part of Y (wheel → bicycle)
  - Meronymy – Y is part of X (bicycle → wheel)
- For verbs the following relationships are provided among synsets
  - Hyperonymy – the activity of X is a kind of Y (to see → to perceive)
  - Troponymy – the activity Y executes X in some sense (to eat → to devour)
  - Entailment – Y is required to perform X (to snore→to sleep)
  - Coordinate terms– Terms share a common hyperonym (to hear-to see as cases of to perceive)

Wordnet relations– adjectives & adverbs

- Words can be linked to other words by lexical relationships such as antonomy (words that have opposite meanings)
  - good ⇔ bad, day ⇔ night, exit ⇔ entrance
- For adjectives the following relationships are defined
  - Related nouns– (noisy → noise)
  - Similar to– (noisy → clanking)
    - The **descriptive adjectives** are organized into groups containing a main synset (head) and satellite synsets. Each group is organized around a pair (sometimes a triple) of antonyms corresponding to the main terms. The satellite synsets are those linked by the "*Similar to*" relationship.
    - Relational adjectives are used to categorize the noun and they have neither a group structure nor an antonym (e.g. musical, nervous)
- For the adverbs the following relationships are defined
  - Base adjective– (slowly → slow)

Example – descriptive adjective



Word sense disambiguation

- Word sense disambiguation (WSD) is the task of selecting the correct sense for a word in a given sentence
  - This problem has to be faced for words having more meanings
  - It requires a dictionary listing all the possible senses for each word
  - It can be faced for each single word or jointly for all the words in the sentence (all the meaning combinations should be considered)

> I ate a cold **dish**
> I washed a dirty **dish**
> The *served* a cold **dish**

  - Several approached to WSD have been proposed
    - (Machine Readable) Dictionary and knowledge-based, Machine Learning Supervised methods, Semi-supervised and Unsupervised methods

Supervised learning

- WSD can be approached as a classification task
  - The correct sense is the class to be predicted
  - The word is represented by a set (vector) of features to be processed as the classifier input
    - Usually the feature include a representation of the word to be disambiguated (target) and of its context (a given number of words at the left and the right of the target word)
    - The word itself, the word stem, the word PoS can be exploited as features
  - The classifier can be learnt from examples given a labeled dataset
  - Different models can be exploited to implement the classifier (Naïve Bayes, neural networks, decision trees…)
  - The limitation of the learning based approach is scalability when a large number of labeled examples is required

Naïve Bayes

- The bayesian approach aims at maximizing of the probability of sense *s* given the feature vector f_w describing the target word

$$\hat{s} = \underset{s \in S}{\mathrm{argmax}}\, p(s|f_w) = \underset{s \in S}{\mathrm{argmax}} \frac{p(f_w|s)p(s)}{p(f_w)}$$

  - With the simplifying assumption that the feature vector entries (words in context) are independent of each other p(f_w|*s*) can be written as

$$p(f_w|s) = \prod_{j=1}^{n} p(f_j|s)$$

    - the probabilities p(f_j|*s*) model the statistics for distribution of feature j (e.g. a given word) in the context of word w when having the sense s
    - p(s) is the a priori probability of each sense of the word

Dictionary-based methods

- A dictionary can provide useful information about the contexts related to the word senses (glosses)
  - A simple approach is the Lesk algorithm (1986)
    - The algorithm computes the intersection among the glosses associated to the different meanings of the words in the sentence
    - The combination yielding the maximum overall intersection is selected (the complexity is combinatorial in the number of senses)

pine cone

PINE
1. kinds of evergreen tree with needle-shaped leaves
2. waste away through sorrow or illness

CONE
1. solid body which narrows to a point
2. something of this shape whether solid or hollow
3. fruit of certain evergreen trees

$pine_1 \cap cone_1 = 0$

$pine_1 \cap cone_2 = 0$

$pine_1 \cap cone_3 = 2$

$pine_2 \cap cone_1 = 0$

$pine_2 \cap cone_2 = 0$

$pine_2 \cap cone_3 = 0$

Limitations of the Lesk algorithm

- The Lesk algorithm yields a 50-70% accuracy
  - The main limitation is its dependence on the quality of the glosses/ examples provided for each sense in the dictionary since they are usually short and do not carry enough information to train a classifier
    - The words in the context and their definition should share a significant intersection (they should share the maximum number of terms)
  - The coverage can be improved by adding the words related to the target but not already contained in the glosses
    - for example the words of the definitions containing the target word only when the actual sense of the target word is clear in that context
  - In the computation of the intersection/similarity among the context more flexible measure can be exploited
    - Correlation with TermFrequency-InverseDocumentFrequency weights in order to reduce the importance of most common words

Word similarity

- Synonymy is a kind of strict similarity
  - It defines a complete equivalence (substitutability) in given contexts
- A word similarity can be defined a kind of "semantic distance"
  - Definition on a given Thesaurus (e.g. Wordnet)
    - Words (senses) are linked by different relationships in a thesaurus
    - A semantic distance can be defined by the length of the minimum path leading from the fist word to the other using the links of a give relationship

$$\text{sim}(w_1, w_2) = -\log \text{pathlenght}(s(w_1), s(w_2))$$

  - Definition on the statistical distribution
    - Probability of finding the two words in similar contexts
    - Words are represented by N-dimensional vector and the similarity/distance is computed in this space