The Semantic Level of NLP – Introduction

- **The semantic level of natural language processing deals with the meaning of words and sentences, as well as with the way in which words' meanings *combine* in order to form the meaning of an entire sentence.**

- **At this processing level, computational semantics studies meaning in a context-free manner. Namely it studies the meaning of a sentence *regardless of the context in which that sentence occurs*. (Here the context is given by an entire paragraph, a page etc.)**

- **Note that, when taking the context into account, we move to the pragmatic level of processing natural language.**

- **Word Sense Disambiguation (WSD) signifies determining the meaning of a polysemous word occurring in a specific sentence.**

- **WSD has multiple applications in artificial intelligence and in information systems (man-machine communication, message understanding, machine translation, speech processing, text processing, information retrieval etc.)**

WSD

- **There are three classes of WSD methods and corresponding algorithms: supervised, unsupervised and knowledge-based.**

  - **Hybrid methods and approaches also exist (for instance, performing WSD at the border between unsupervised and knowledge-based techniques).**

- **Determining the meaning of a unique polysemous word in context (the sentence where it occurs) represents *local disambiguation*. We will be using it in order to determine the meaning of an entire text, which represents *global disambiguation*.**

Supervised WSD

- **Supervised WSD is based on machine learning.**

- **The (supervised) problem: given a word and a set of labels (etichete), namely a set of senses that can be associated with that word, one must establish which of these labels (senses) corresponds to that word, in the specific context in which it occurs.**

- **From a technical point of view, this is a classification problem.**

- **The elements which define this classification problem are: *the sense inventory* (inventarul de sensuri), namely the set of *all* senses that can be associated to *each* word that must be disambiguated and *the evidence sources* (sursele de evidenta), namely the information that can be extracted from the context and on which the classification process can rely.**

- **This WSD process uses a *training corpus* (corpus de antrenare) that contains the words which are to be disambiguated already labeled with their correct senses according to the respective context. This annotated corpus is used in order to train a classifier that can label the words occurring within a new and unlabeled text.**

- **<u>The task</u> is that of building a classifier that classifies correctly the new cases, based on the context in which they are used.**

- **Note that, because it requires the existence of an already annotated training corpus, this type of WSD can not be used in practice – at a large scale. Such a training corpus does not always exist and not for all words of a language – because it must be manually built by humans.**

Unsupervised WSD

- **Unsupervised WSD is also corpus-based, but uses an unannotated corpus, which makes it helpful in practice.**

- **Unsupervised WSD does not accomplish straight-forward sense disambiguation, because it does not assign sense tags to words. Rather it discriminates among the different meanings of a word.**

    - **Unsupervised WSD discriminates among word meanings based on information found in unannotated corpora.**

- **Note that unannotated corpora is always available for all languages (see, for instance, the texts on the web).**

{singular – corpus

plural – corpora }

- **Conclusion: These methods are data-driven and language-independent, and rely on the distributional characteristics of unannotated corpora.**

Unsupervised WSD

- **Distributional approaches do not assign meanings to words, but rather allow us to *discriminate* among the meanings of a word by identifying *clusters* of similar contexts, where each cluster shows that word being used in a particular meaning.**

- **Distributional approaches are based on the assumption that words that occur in similar contexts will have similar meanings.**

- **A classical clustering technique that can be used in order to obtain clusters of similar contexts is given by the Naïve Bayes model. (We will see this in a future lecture).**

- **Establishing the clustering technique to be used (classical versus state of the art) is of the essence.**

- **These methods are knowledge-lean (lipsite de cunostinte) because they do not rely on external knowledge sources (such as machine readable dictionaries, concept hierarchies, or sense-tagged text).**

- **Note that unsupervised WSD based on translational equivalence as found in *word-aligned parallel corpora* also exists. We will focus on distributional approaches. Such approaches rely on monolingual corpora.**

Unsupervised WSD

We will focus on unsupervised corpus-based methods of *word sense discrimination* that are knowledge-lean, and do not rely on external knowledge sources such as machine readable dictionaries, concept hierarchies, or sense-tagged text.

Knowledge-based WSD

- **Knowledge-based WSD methods use a preexisting *sense inventory* in order to assign a specific meaning to a specific word in a given context.**

- **The sense inventory can be a classical dictionary in electronic format, a bilingual dictionary, namely machine readable dictionaries, concept hierarchies, semantic networks and lexical databases (such as WordNet) etc.**

- **The choice of sense inventory is of the essence for knowledge-based WSD accuracy. WordNet probably represents the most widely used sense inventory nowadays.**

WSD (in general)

- **WSD systems are based either on preexisting knowledge (thus requiring an external knowledge base) or on machine learning (thus requiring corpora – annotated when referring to supervised WSD, and unannotated when referring to unsupervised WSD).**

- **WSD can be performed at various levels of granularity. For instance, usage of the knowledge base WordNet leads to fine-grained WSD - because WordNet makes fine-grained distinctions between senses and subsenses. For example, most knowledge bases will determine the legal meaning of ("*suit*" - *speta*) without differentiating between *civil suit* (*speta civila*) and *criminal suit* (*speta penala*) as does WordNet.**

WordNet (WN)

- **WordNet (WN) is a lexical database of English created by Prof. George Miller at Princeton University in the nineties and being constantly updated.**

- **WN comprises all content words in English (nouns, adjectives, verbs and adverbs). Nouns and verbs are organized into hierarchies, adjectives and adverbs into clusters.**

- **The building brick of WN is the *synset* or synonym set (a set of synonyms).**

- **Synonymy in WN is not the classical one that linguists refer to. In WN, two words are considered synonyms if they refer to the same *concept*. For instance, the words "chair" and "table" could be considered synonyms because they both refer to the concept of *furniture*.**

- **A polysemous word (a word with multiple senses) refers to a different concept through each of its senses. Therefore, a polysemous word belongs to multiple WN synsets. It belongs to a different sysnset through each of its senses. For example, the polysemous Romanian word**

“masa” could belong to each of the synsets referring to the concepts of *mancare, mobila, multime, masa din fizica etc*.

- **Each WN synset refers to a *concept*.**

- **A WN synset contains all synonyms that lexicalize the concept it refers to, and that have the same part of speech (only nouns, only verbs etc.), together with a string of text named *gloss*, which resembles a classical dictionary definition. The gloss may contain an example of usage.**

- **Note that, in WN, the syntactic category (noun, adjective, verb, adverb) is used as parameter.**

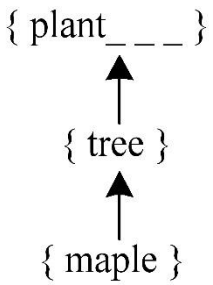
WordNet (WN)

- **WN synsets are linked together through semantic relations (existing between concepts and which are language independent).**

- **NOUN SYNSETS**

The noun synsets are linked through semantic relations such as *hypernymy (hiperonimie)*, the inverse relation *hyponymy (hiponimie)*, *meronymy (meronimie)*, the inverse relation *holonymy (holonimie)* etc. Hypernymy denotes the father concept within the father-son relation. Its inverse, hyponymy denotes the son concept within the same semantic relation. Meronymy is the *part-of* relation, while holonymy is its inverse. For example, take *car* and *tire*, with the tire being *part of* the car (both are nouns). Then *car* is a meronym of *tire* and *tire* is a holonym of *car*.

- **The hypernymy relation organizes the WN noun synsets into hierarchies, as in the following example:**



isa (sau hiperonimie)

(The hypernymy relation is the general *isa* relation from artificial intelligence in the case of the NLP field. The *isa* relation is used in AI to create hierarchies.)

Usage of semantic relations between synsets turns WN into a semantic network.

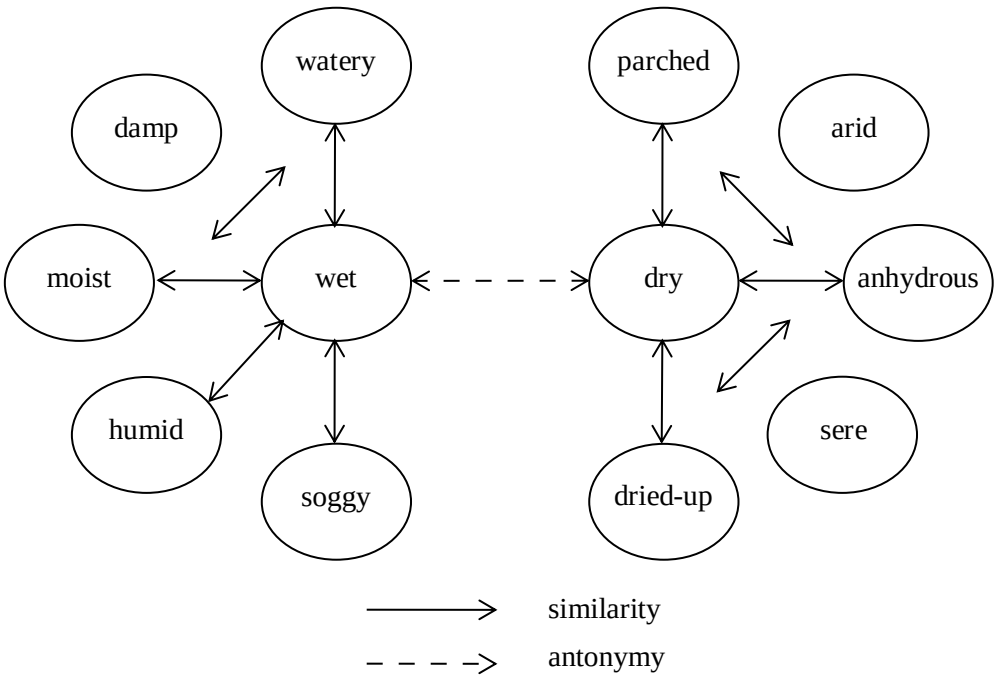
Through the hypernymy relation concepts inherit all properties of their superconcepts (parent concepts). In the above example, the maple has all the properties of a tree. Moreover, it has all the properties of a plant, in general. Going through the hierarchy, we will know that a maple has all the properties of a plant. The inheritance of properties turns WN into a knowledge base as well.

WN is a lexical database of English, a semantic network and a knowledge base. The fact that WN is a knowledge base makes it useful for a great variety of AI applications.

ADJECTIVE SYNSETS

- **Other, different semantic relations hold between adjective sysnsets. The main semantic relation is considered to be *antonymy*.**

- **Using *the antonymy relation* adjective synsets are organized into *clusters*. The following is an example of cluster formed around the antonyms *wet / dry*:**



This structure holds for the majority of descriptive adjectives (those that assign to a noun a value of an attribute).

- **The *similarity relation* holds between adjectives only. It brings in *indirect antonyms*. In the example, *wet* and *dry* are *direct antonyms*. *Wet* and *arid* are *indirect antonyms*. *Moist* does not have a direct antonym, but its indirect antonym can be found following the path *moist->wet->dry*.**

- **Other semantic relations important for adjectives are *also-see, pertaining-to* and *attribute*.**

- **The antonymy relation was proven to bring in important information for WSD.**

VERB SYNSETS AND ADVERB SYNSETS

- **Verb synsets are also organized into hierarchies, the same as noun synsets, only according to the *entailment* relation (to entail = a atrage dupa sine)**

- **Another important semantic relation for verbs is the *causal* relation.**

- **Adverb synsets are organized into clusters just like adjective synsets.**

SEE THE WN IMPLEMENTATION IN NLTK AT THE LAB!

SIMILARITY AND RELATEDNESS

(Similaritate si inrudire)

- **Measures of semantic similarity or relatedness are used in various NLP applications such as: WSD, text summarization, information retrieval and information extraction etc.**

- **Semantic relatedness (inrudire) is a more general concept than semantic similarity (similaritate). Semantic relatedness has as inverse the semantic distance.**

- **Semantic similarity is a particular case of relatedness. Similarity is more specialized than relatedness. For instance, *doctors* and *hospitals* can be related, but are not similar. *Car* and *tire* are not similar, but are certainly related.**

- **Various measures for semantic similarity and/or relatedness exist in the literature.**

- **It has turned out that, for the task of WSD, relatedness is more important and relevant than similarity. (Example: *car* and *tire* are related but not similar. However, the occurrence of *car* in the same sentence as**

tire **will help disambiguate** ***tire*** **with its noun sense – as part of a car – versus the verb sense of being tired).**

- ✓ **The knowledge-based WSD algorithm that we will study during the next lecture will strongly rely on relatedness between concepts.**