Natural Language Processing

Lecture 7
Lexical semantics and Latent Semantic Analysis

2021

Word meanings

As we have seen (in Lecture 1), according to the *principle of compositionality*,

The meaning of a complex expression is determined by the meanings of its constituent expressions and the rules used to combine them. [1]

Although the principle is not without its problems,[2] it suggests that to know the meaning of larger textual units (sentences, paragraphs etc.) it is necessary to know the *meaning of words* they are composed of.

[1] Wikipedia: Principle of Compositionality.

[2] See, e.g., Szabó [2020].

Word meanings cont.

Intuitively, several words have more than one meanings, e.g. *mouse* has a different meaning in

A mouse ate the cheese.

and in

Click on the close button with the mouse.

mouse can mean a *type of small rodent* or *an electronic pointing device*. The identification and characterization of word meanings or *word senses* such as these is the task of *lexical semantics*.

Word senses in dictionaries

One way of characterizing word senses is offered by traditional *dictionaries*. E.g., the online version of the *Oxford Advanced Learner's Dictionary* describes these senses as

mouse *noun*

A1

/maʊs/

/maʊs/

1 A1

(plural **mice** /maɪs/ /maɪs/)

a small animal that is covered in fur and has a long thin tail. Mice live in fields, in people's houses or where food is stored.

- *a house mouse*
- *The stores were overrun with rats and mice.*
- *She crept upstairs,* ***quiet as a mouse****.*
- (figurative) *He was a weak little mouse of a man.*

Word senses in dictionaries cont.

and

2 A1

(plural **mice** or **mouses**)

(computing) a small device that is moved by hand across a surface to control the movement of the cursor on a computer screen

- *Use the mouse to drag the icon to a new position.*
- *I prefer a wireless mouse.*
- *The keyboard and mouse are wireless devices.*
- *Click the left **mouse button** twice to highlight the program.*
- *With simple **mouse clicks**, the viewer can navigate the room.*

Word senses in dictionaries cont.

Notable features of these sense descriptions are that

- word senses have precise identifiers: the surface form *mouse*, the POS-tag *noun* and the sense number together unambiguously identify the senses;
- each sense has a *textual definition* which is not formal, but
  - uses a relatively small definitional vocabulary,
  - follows certain conventions, e.g., starts with a more general word plus characteristic property (*small animal*, *small device*);
- there are several *example sentences* illustrating typical patterns in which the sense is used.

Lexical relations

Dictionaries may contain information about *lexical relations* between senses, especially about

- *synonymy*: whether two word senses are (close to) identical;
- *antonimy*: whether two word senses are opposites of each other.

Other important lexical relations include *taxonomical relations*:

- sense s_1 is a *hyponym* of s_2 if it is strictly more specific, e.g. *mouse*$_1$ is a hyponym of *animal*$_1$;
- conversely, sense s_1 is a *hypernym* of s_2 if s_2 is more specific than s_1.

Lexical relations cont.

And, finally, *meronymy*, the *part-whole* relation: e.g., *finger* is a meronym of *hand*.

Collectively, word senses and their lexical relations constitute a *network*, in which

- nodes are sets of synonymous word senses, and
- edges are lexical relations.

Since the hyponymy relation (also called *is_a*) is transitive, it makes sense to have only *direct hyponymy* edges in the network, i.e., the have an $s_1 \xrightarrow{is\_a} s_2$ edge only if there is no node s_3 for which $s_1 \xrightarrow{is\_a} s_3$ and $s_3 \xrightarrow{is\_a} s_2$.

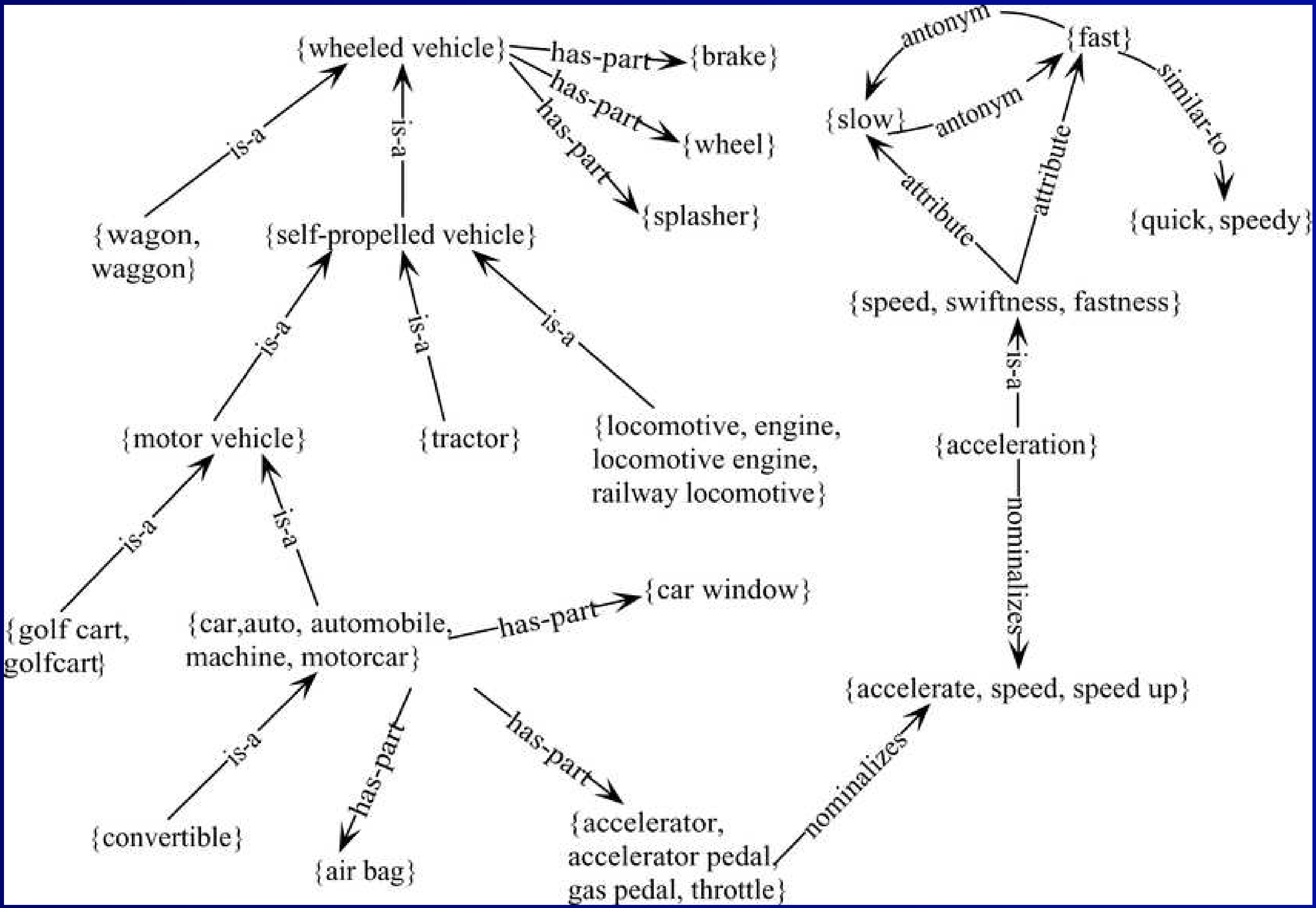
WordNet

To be usable for NLP purposes, lexical semantic information has to be accessible as a computational resource with a well defined query API, and, starting from the mid. 1980s a number of projects developed such resources.

The most important has been the *WordNet* English lexical database, which contains a large number of synonym sets with definitions, examples and lexical relations. After its success, WordNets were developed for a large number of other languages, now more than 200 WordNets are available.

WordNet cont.

A part of the English WordNet network:



(Figure from Navigli [2009].)

Knowledge bases as lexical resources

In addition to dedicated lexical databases, *knowledge bases* can also serve as useful lexical semantic resources, since they contain information about *entities* and *concepts*, which can be linked to words in a vocabulary. Important examples include

- *Wikis*, most importantly the English Wikipedia, here various types of links and references between the entries provide relational information;
- *formal ontologies*: these describe relationships between concepts in a formal logical language.

Word sense disambiguation

To use the information about word senses provided by these lexical resources, NLP applications must be able to determine in which sense words are used in the input, i.e., perform *word sense disambiguation (WSD)*. The details of the WSD task depend on which lexical resource it is based on and how the resource is used. Given a resource containing word senses,

- *supervised WSD* uses machine learning methods on training data which is annotated with the correct word senses; while
- *knowledge-based WSD* exploits the information in the lexical resource, e.g. the lexical relations and definitions in WordNet.

Vector-based lexical semantics

The lexical semantic approach we have seen so far has certain features that make it difficult to achieve large coverage and adapt to new languages or domains:

- the lexical databases were manually assembled by highly qualified experts;
- the development of high-performance WSD modules typically requires a large amount of expert-annotated training data.

These problems led to research into alternatives that assign useful word meaning representation in an *unsupervised* fashion, simply learning them from text corpora.

Vector-based lexical semantics cont.

Although there have been attempts to learn *semantic networks* from text corpora, the first successful unsupervised lexical semantic methods have been learning *word vectors* from text corpora, i.e., embedding functions of the form

$$E : V \rightarrow \mathbb{R}^d$$

which assign d-dimensional ($d \in \mathbb{N}$) vectors to each word in the V vocabulary. Of course, not any such function will do: the obvious requirement is that the learned vectors have to convey useful information about the *meaning* of the words they are assigned to.

Vector-based lexical semantics cont.

One way of ensuring the connection is to utilize the *distributional hypothesis*:

- "You shall know a word by the company it keeps." [3]
- "Linguistic items with similar distributions have similar meanings." [4]

This suggests that if the word vectors reflects the *distribution* of the words they are assigned to, then they will also reflect the words' meanings.

---

[3] J.R. Firth, *Papers in Linguistics 1934–1951 (1957).*

[4] Wikipedia: Distributional semantics.

Co-occurrence matrices

The most direct way of getting word vectors that reflect the words' distribution in a corpus is to consider *co-ocurrence* matrixes. If there are D documents in the corpus and V is the corpus vocabulary then

- *term-document* matrices are $|V| \times D$ dimensional matrices in which each row is a word vector whose i-th element is the occurrence count of the word in the i-th document, while
- *term-term* matrices are $|V| \times |V|$ dimensional matrices in which each row is a word vector whose i-th element is the co-occurrence count of the word with the i-th *other word*.

Latent Semantic Analysis

An important problem of using these vectors directly is their huge dimensionality and sparsity. To solve this problem, *Latent Semantic Analysis* methods apply dimension reducing matrix factorization methods, typically *truncated SVD* to find a *low-rank approximation* of the original C co-occurrence matrix. With SVD the factorization is

$$C \approx USV^{\intercal}$$

with U, V orthonormal and S diagonal. In case of truncated SVD, the rows of the U matrix can be used as low-dimensional, approximate representations of the co-occurrence based original word vectors.

References

Zoltán Gendler Szabó. Compositionality. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford, fall 2020 edition, 2020. URL `https://stanford.io/3nLHOUJ`.

Roberto Navigli. Word sense disambiguation: A survey. *ACM computing surveys (CSUR)*, 41(2):1–69, 2009. URL `https://bit.ly/2LwiQAP`.