Methods in Computational Linguistics
Lexical Semantics I: Distributional Hypothesis & Vector Spaces

---

Franziska Weeber

Master of Science *Computational Linguistics*
Institut für Maschinelle Sprachverarbeitung
Universität Stuttgart

November 28, 2025

Today's slides are based on the materials provided by Eva Maria Vecchi.

Overview

1. Word Meaning & Lexical Semantics
2. Semantic Relations
3. Distributional Semantics
4. Building Vector Spaces

Word Meaning & Lexical Semantics

Semantics

- Semantics is the study of meaning communicated through language
- Semantics (and pragmatics) are the glue that connect language to the real world
- Phonology, Morphology, Syntax, etc. are meaningful *only* once Semantics is taken into account, at some level
- We'll start at the word-level: Lexical Semantics

Word Meaning

Word Meaning

I saw my <u>mother</u> just now.

- We know the speaker saw...
  - a female human
  - someone older than speaker
  - someone with a specific relation to speaker
- Word meaning is constructed using
  - Lexical relations e. g. between *woman* and *mother*
  - Links between linguistic and world knowledge

Word Meaning

Semiotic Triangle

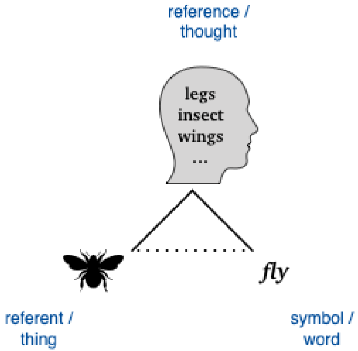


Figure: adapted from Ogden and Richards (1923)

Word Meaning

- We can understand and use expressions that have no real-world referent
  - *unicorns, World War III, love*
- Expressions can share the same referent, but have different meanings.
  1. Then in 1981, Anwar El Sadat was assassinated.
  2. Then in 1981, the President of Egypt was assassinated.
- We can use two expressions without being aware they share the same referent (Frege, 1892)
  3. The morning star is the evening star.
  4. Venus is Venus.
- Because we understand the expression *President of Egypt*, we can use it to refer to a particular individual at any given time
  $\longrightarrow$ **Sense**

Semantic Primatives

- Basic units of semantic representation and cognition, irreducible and non-linguistic
- Should express minimal units for semantic representation
  - word meanings are represented as a collection of semantic primitives
  - capture conceptual/real-world aspects of word meaning
  - formalize what is common to nearly synonymous expressions within a single language (e. g.,give / receive / grammatical alternations) or in different languages
  - account for inferences that a speaker is able to make (e. g.,*John is man* implies *John is male*)

Semantic Primatives

	with a back	raised above ground	for one person	to sit on	with arms	of solid material
chair	+	+	+	+	-	+
armchair	+	+	+	+	+	+
stool	-	+	+	+	-	+
sofa	+	+	-	+	+	+
pouffe	-	+	+	+	-	-

Necessary and sufficient conditions

- x is a lion if and only if L.
- where L is a list of attributes
  - x is an animal,
  - x has four legs,
  - x is a carnivore,
  - x is a feline,
  - x has a mane, ...

Necessary and sufficient conditions

- x is a lion if and only if L.
- where L is a list of attributes
  - x is an animal,
  - x has four legs,
  - x is a carnivore,
  - x is a feline,
  - x has a mane, ...

Word Meaning?

He scored with his left **foot**[1].

They made camp at the **foot**[2] of the mountain.

I ate a **foot**[3]-long hot dog.

Lexical Semantics

The study of...

- what individual lexical items mean,
- how we can represent their meaning,
- and how to combine the meaning of individual items to obtain an interpretation for a phrase/utterance

Lexical Semantics in Computational Linguistics

- Recognize word senses in text (manually and automatically)
- Define similarities between words
- Determine how strongly a verb "goes with" its subject (selectional preferences)
- Recognize and interpret figurative uses of words
- Describe relations between words (or better, between word senses)

Information Retrieval

- Goal to find relevant documents, even if differently phrased
- QUERY: "*female astronauts*"
- DOCUMENT: "*In the history of the Soviet space program, there were only three female cosmonauts: Valentina Tereshkova, Svetlana Savitskaya, and Elena Kondakova*"
- System must recognize that astronaut and cosmonaut have similar meanings (in a given context!).

Machine Translation

Bar-Hillel (1960)

The box is in the pen.

- World knowledge necessary to disambiguate *polysemous* words
- Correct translation depends on selecting the correct sense of *pen*

Semantic Relations

Semantical Relations

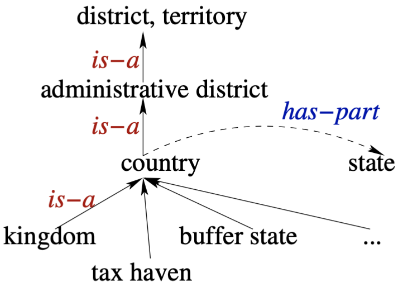
Jurafsky and Martin (2025)

- Words have different senses
- Words with the same sense can be aggregated to a synset
- Characterize each sense by its semantic relations

Relation	**Description**	**Example**
synonymy	same meaning	*sofa - couch*
antonymy	opposites	*dark - light*
hyponym	subclass	*car* is a hyponym of *vehicle*
hypernym	superclass	*vehicle* is a hypernym of *car*
structured polysemy	multiple senses of a word are semantically related	*bank* as organization and building

Semantic Ontologies

WordNet



- Semantic Ontologies: structured dictionaries that define word senses and relation to other word senses
- WordNet: large lexical resource that organizes words and synsets according to their semantic relations

Limitations of Relational Models

- Relational models such as WordNet are glorified thesauri
- Require many years of development and depend on skilled lexicographers
- Inconsistencies throughout the resource
- Ontology is only as good as ontologist(s) – it is not only data

Distributional Semantics

Meaning and Distribution

We found a little, hairy wampimuk sleeping behind the tree (Lazaridou et al., 2014)

- "Die Bedeutung eines Wortes liegt in seinem Gebrauch." (Wittgenstein, 1953)
  - meaning = use = distribution in language
- "You shall know a word by the company it keeps." (Firth, 1957)
  - distribution = collocations = habitual word combinations
- **Distributional hypothesis:** difference of meaning correlates with difference of distribution (Harris, 1954)
  - semantic distance
  - Assumption: semantically similar words occur in the context of the same words $\longrightarrow$ "similar" as approximation of "synonymous"
- "What people know when they say that they know a word is not how to recite its dictionary definition – they know how to use it [...] in everyday discourse." (Miller, 1986)

Ex: What does "bardiwac" mean?

- *He handed her a glass of bardiwacs.*
- *Beef dishes are made to complement the bardiwacs.*
- *Nigel staggered to his feet, face flushed from too much bardiwac.*
- *Malbec, one of the lesser-known bardiwac grapes, responds well to Australia's sunshine.*
- *I dined off bread and cheese and this excellent bardiwac.*
- *The drinks were delicious: blood-red bardiwac as well as light, sweet Rhenish.*

$\longrightarrow$ Bardiwac is a red wine

Distributional semantics

Landauer and Dumais (1997); Turney and Pantel (2010), ...

```
he curtains open and the nnnn shining in on the barely
ars and the cold , close nnnn " . And neither of the w
rough the night with the nnnn shining so brightly , it
made in the light of the nnnn . It all boils down , wr
 surely under a crescent nnnn , thrilled by ice-white
sun , the seasons of the nnnn ? Home , alone , Jay pla
m is dazzling snow , the nnnn has risen full and cold
un and the temple of the nnnn , driving out of the hug
 in the dark and now the nnnn rises , full and amber a
bird on the shape of the nnnn over the trees in front
 But I could n't see the nnnn or the stars , only the
rning , with a sliver of nnnn hanging among the stars
 they love the sun , the nnnn and the stars . None of
the light of an enormous nnnn . The plash of flowing w
man 's first step on the nnnn ; various exhibits , aer
 the inevitable piece of nnnn rock . Housing The Airsh
oud obscured part of the nnnn . The Allied guns behind
```

Distributional semantics

Landauer and Dumais (1997); Turney and Pantel (2010), ...

he curtains open and the moon shining in on the barely
ars and the cold , close moon " . And neither of the w
rough the night with the moon shining so brightly , it
made in the light of the moon . It all boils down , wr
surely under a crescent moon , thrilled by ice-white
sun , the seasons of the moon ? Home , alone , Jay pla
m is dazzling snow , the moon has risen full and cold
un and the temple of the moon , driving out of the hug
in the dark and now the moon rises , full and amber a
bird on the shape of the moon over the trees in front
But I could n't see the moon or the stars , only the
rning , with a sliver of moon hanging among the stars
they love the sun , the moon and the stars . None of
the light of an enormous moon . The plash of flowing w
man 's first step on the moon ; various exhibits , aer
the inevitable piece of moon rock . Housing The Airsh
oud obscured part of the moon . The Allied guns behind

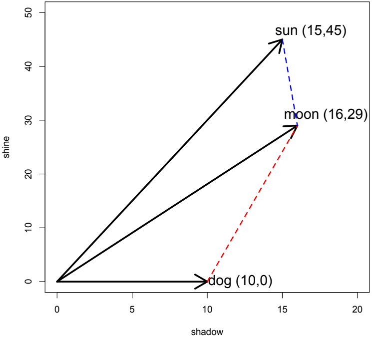
Distributional semantics

The geometry of meaning

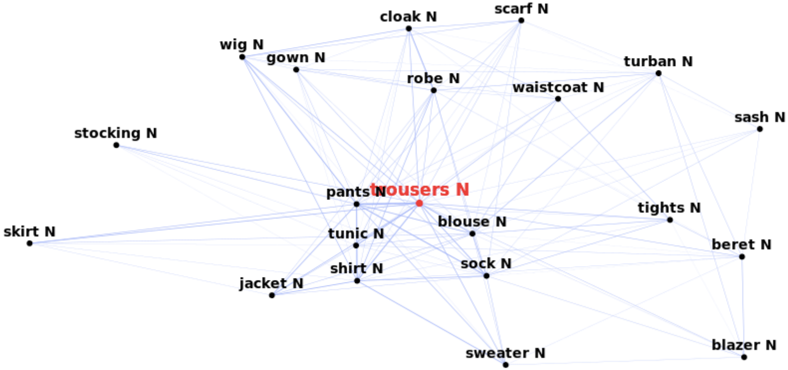
Distributional Semantic Model (DSM): a scaled and/or transformed co-occurrence matrix $\mathbf{M}$, such that each row $\mathbf{x}$ represents the distribution of a target term across contexts.

- e. g., within a document, within a window of [content] words before and after, etc.

	shadow	shine	planet	night
moon	16	29	10	22
sun	15	45	14	10
dog	10	0	0	4

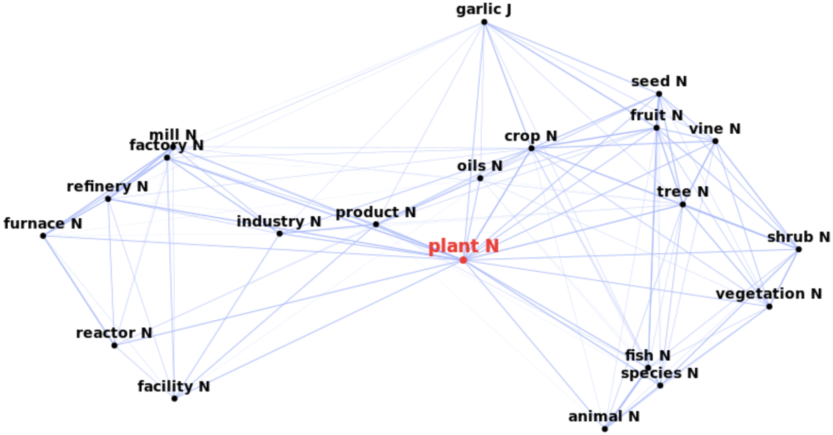


Nearest Neighbors of *trousers*



*Based on DSM built on EN Wikipedia, (filtered) dependency contexts

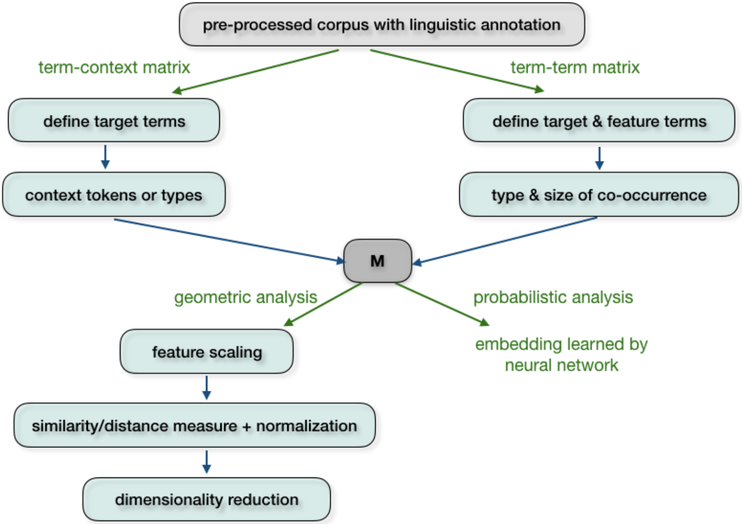
Nearest Neighbors of *plant*



*Based on DSM built on EN Wikipedia, (filtered) dependency contexts

Building Vector Spaces

Building a distributional model



Linguistic Pre-processing & Term Definition

Start with a corpus with (relevant) linguistic annotations

- Tokenization
- POS-tagging (*light*_N vs. *light*_J vs. *light*_V)
- Stemming/lemmatization
  - *go*, *goes*, *went*, *gone*, *going* $\rightarrow$ *go*
- Dependency parsing or shallow syntactic chunking

- Effect of linguistic pre-processing
  - Nearest neighbors of *walk* (in British National Corpus, DSM defined by head of the subject of *walk*)
    - Word forms: stroll, walking, walked, go, path, drive, ride, wander, sprinted, sauntered
    - Lemmatized forms: hurry, stroll, stride, trudge, amble, wander, walk-NN, walking, retrace, scuttle

Context Type

1. Context term appears in same fixed window
2. Context term is a member in the same linguistic unit as target
   - e. g. paragraph, sentence, turn in conversation
3. Context term is linked to target by a syntactic dependency (e. g. subject, modifier)

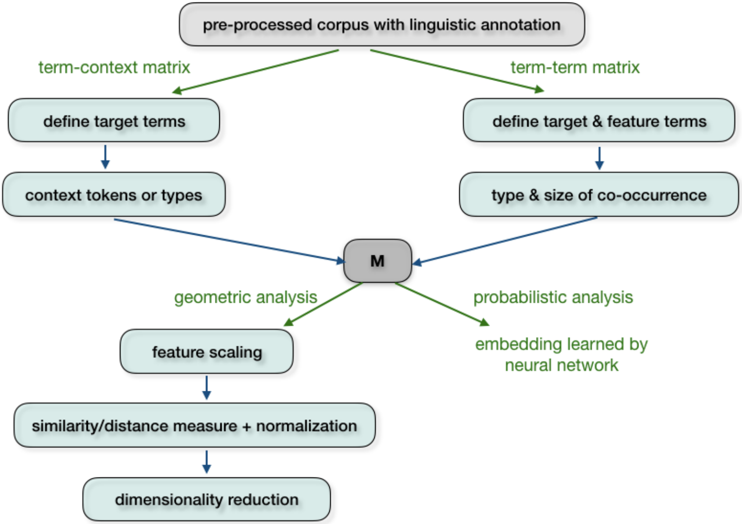
- Context type (e. g. window size) can have impact on how terms are related to those in its nearest neighborhood
  - smaller window sizes → pragmatically related (e. g. car, van, vehicle, truck)
  - larger window sizes → syntagmatically related (e. g. car, drive, park, windscreen)

Similarity vs. Relatedness

It is generally accepted that there are (at least) two dimensions of word associations:

- Semantic Similarity: two words sharing a high number of salient features (attributes) $\longrightarrow$ paradigmatic relatedness
  - (near) synonymy (*car-automobile*)
  - hyperonymy (*car-vehicle*)
  - co-hyponymy (*car-van-lorry-bike*)
- Semantic Relatedness: two words semantically associated without being necessarily similar $\longrightarrow$ syntagmatic relatedness
  - function (*car-drive*)
  - meronymy (*car-tire*)
  - location (*car-road*)
  - attribute (*car-fast*)
  - other (*car-petrol*)

Building a distributional model



Feature Scaling

Feature scaling is used to "discount" less important features:

- Logarithmic scaling: $O' = log(O+1)$ (cf. Weber-Fechner law for human perception)
- Relevance weighting, e. g. tf.idf (information retrieval)
  - $tf.idf = tf \cdot log(D/df)$
  - tf = co-occurrence frequency O
  - df = document frequency of feature (or nonzero count)
  - D = total number of documents (or row count of $\mathbf{M}$
- Statistical association measures (Evert 2004, 2008) take frequency of target term and feature into account
  - i. e., how surprised are we to see context term associated with target word?
  - measures differ in how they balance O and E

Simple association measures

For details, see the corpus analysis lecture!

- Pointwise Mutual Information (PMI): $PMI(x, y) = \log \frac{f_{obs}}{f_{exp}}$
  Disadvantage: PMI overrates combinations involving rare terms

- t-score: $assoc_{t-test}(x, y) = \frac{f_{obs} - f_{exp}}{\sqrt{f_{obs}}}$

- Log-Likelihood: $G^2 = \pm 2 \cdot \left(f_{obs} \cdot log_2 \frac{f_{obs}}{f_{exp}} - (f_{obs} - f_{exp}) \right)$

- Odds Ratio: $\mathrm{OR} = \frac{O_{11} \cdot O_{22}}{O_{12} \cdot O_{21}}$

Vector Similarity

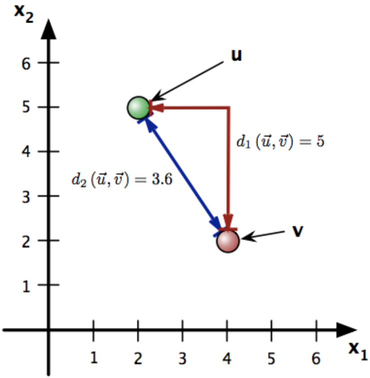
- Minkowski p-distance / geometric distance: length of the connecting line of two points
  - takes both length of vectors (= frequency of words) and patterns of occurrence into account

- Cosine similarity / angular distance: measure the angle between vectors
  - (correctly) ignores length of vectors (= frequency of words)
  - small angle between vectors = similar proportion of context words
  - Cosine of an angle is easy to compute
    - $\cos \longrightarrow 1$: angle is 0° (very similar)
    - $\cos \longrightarrow 0$: angle is 90° (very dissimilar)

- Similarity rankings with euclidean distance and cosine similarity are equivalent: if vectors have been normalized ($||\mathbf{u}||_2 = ||\mathbf{v}||_2 = 1$), both lead to the same neighborhood ranking

Geometric Distance

- Distance between vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n \to$ (dis)similarity
  - $\mathbf{u} = (u_1, \ldots, u_n)$
  - $\mathbf{v} = (v_1, \ldots, v_n)$
- Euclidean distance $d_2(\mathbf{u}, \mathbf{v})$
- "City block" Manhattan distance $d_1(\mathbf{u}, \mathbf{v})$
- Both are special case of the Minkowski p-distance $d_p(\mathbf{u}, \mathbf{v})$ (for $p \in [1, \infty]$)

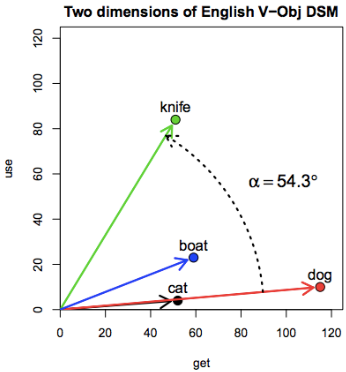


Similarity Measures

- Angular distance α between vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ is given by

$$\cos \alpha = \frac{\mathbf{u}^\top \mathbf{v}}{||\mathbf{u}||_2 \cdot ||\mathbf{v}_2||}$$

- Cosine measure of similarity: $\cos \alpha$
  - $\cos \alpha = 1 \rightarrow$ collinear
  - $\cos \alpha = 0 \rightarrow$ orthogonal



Latent Semantic Dimensions

- Collocations/cooccurrences are shallow!
  - Vectors in standard vector spaces are very sparse
  - But, there are semantic "undercurrents" beneath the surface.
- The latent property of a target is implied through associations with symbolic dimensions which will be correlated
- Identify the latent dimensions and project the data onto these new dimensions $\rightarrow$ dimensionality reduction

Dimensionality Reduction

- Remove unimportant features beforehand (e. g., by word frequency or POS tag)
- Latent Semantic Analysis (LSA)
  - Automatic mathematical/statistical technique for compressing the sparse *word x context* matrix into a dense low dimensional matrix.
  - Based on singular value decomposition
- Principal Component Analysis (PCA)
  - Identify new important and uncorrelated features (principal components) that capture the highest variance in the data
  - Based on covariance matrix, eigenvectors and eigenvalues
  - Non-principal components can then be removed without losing much information

References

Yehoshua Bar-Hillel. 1960. The present status of automatic translation of languages. *Advances in computers* 1:91–163.

JR Firth. 1957. A synopsis of linguistic theory 1930-1955. *Studies in Linguistic Analysis, Special Volume/Blackwell* .

Gottlob Frege. 1892. Über begriff und gegenstand. *"Zeitschrift für Philosophie und philosophische Kritik"* 100:25–50.

ZS Harris. 1954. Distributional structure.

Daniel Jurafsky and James H. Martin. 2025. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, with Language Models*. 3rd edition. Online manuscript released August 24, 2025. https://web.stanford.edu/ jurafsky/slp3/.

Thomas K Landauer and Susan T Dumais. 1997. A solution to plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological review* 104(2):211.

Angeliki Lazaridou, Elia Bruni, and Marco Baroni. 2014. Is this a wampimuk? cross-modal mapping between distributional semantics and the visual world. In Kristina Toutanova and Hua Wu, editors, *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, Baltimore, Maryland, pages 1403–1414. https://doi.org/10.3115/v1/P14-1132.

George A Miller. 1986. Dictionaries in the mind. *Language and cognitive processes* 1(3):171–185.

C. K. Ogden and I.A. Richards. 1923. *The Meaning of Meaning*. Harcourt, Brace Co. / Kegan Paul, Trench, Trubner Co., New York / London.

Peter D Turney and Patrick Pantel. 2010. From frequency to meaning: Vector space models of semantics. *Journal of artificial intelligence research* 37:141–188.

Ludwig Wittgenstein. 1953. *Philosophische Untersuchungen*. Suhrkamp Verlag, Frankfurt am Main.