Syntax and Semantics

Semantic Analysis Phase

Semantics

What is Semantics?

The study of meaning: Relation between symbols and their denotata.

John told Mary that the train moved out of the station at 3 o'clock.

Basic Notation:

1. Distributional Semantics
2. Lexical Semantics

Shifting from arrangement to meaning domain. That is Syntax to Semantics

Computational Semantics

Way to find out how computers can produce semantic representation.

“Study of, how to automate the process of constructing and reasoning with meaning representation of natural language expression.”

Automate the process of building representation.

Computational Semantics

Computational Semantics

The study of how to automate the process of constructing and reasoning with meaning representations of natural language expressions.

Methods in Computational Semantics generally fall in two categories:

- **Formal Semantics:** Construction of precise mathematical models of the relations between expressions in a natural language and the world.
  John chases a bat $\rightarrow \exists x[bat(x) \wedge chase(john, x)]$
- **Distributional Semantics:** The study of statistical patterns of human word usage to extract semantics.

Meaning of an individual word then combined words meaning

Distributional Hypothesis

Distributional Hypothesis: Basic Intuition

"The meaning of a word is its use in language." (Wittgenstein, 1953)

"You know a word by the company it keeps." (Firth, 1957)

→ Word meaning (whatever it might be) is reflected in linguistic distributions.

"Words that occur in the same contexts tend to have similar meanings." (Zellig Harris, 1968)

→ Semantically similar words tend to have similar distributional patterns.

Distributional Semantics

A Linguistic Perspective

"If linguistics is to deal with meaning, it can only do so through distributional analysis." (Zellig Harris)

"If we consider words or morphemes A and B to be more different in meaning than A and C, then we will often find that the distributions of A and B are more different than the distributions of A and C. In other words, difference in meaning correlates with difference of distribution." (Zellig Harris, "Distributional Structure")

Differential and not *referential*

Distributional Semantics

A Cognitive Perspective

Contextual representation

A word's contextual representation is an abstract cognitive structure that accumulates from encounters with the word in various linguistic contexts.

We learn new words based on contextual cues

He filled the **wampimuk** with the substance, passed it around and we all drunk some.

We found a little **wampimuk** sleeping behind the tree.

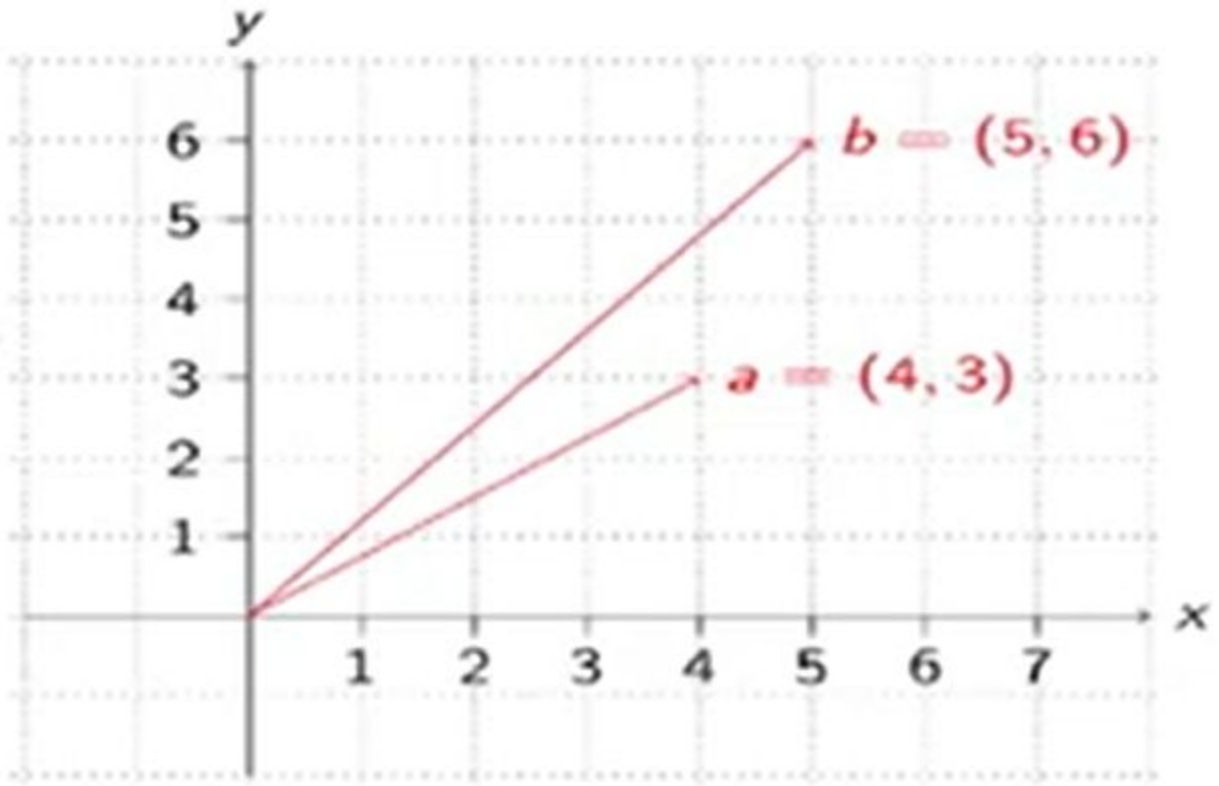
Distributional Semantic Models (DSM)

- Computational models that build contextual semantic repesentations from corpus data
- DSMs are models for semantic representations
  - The semantic content is represented by a vector
  - Vectors are obtained through the statistical analysis of the linguistic contexts of a word
- Alternative names
  - corpus-based semantics
  - statistical semantics
  - geometrical models of meaning
  - vector semantics
  - word space models

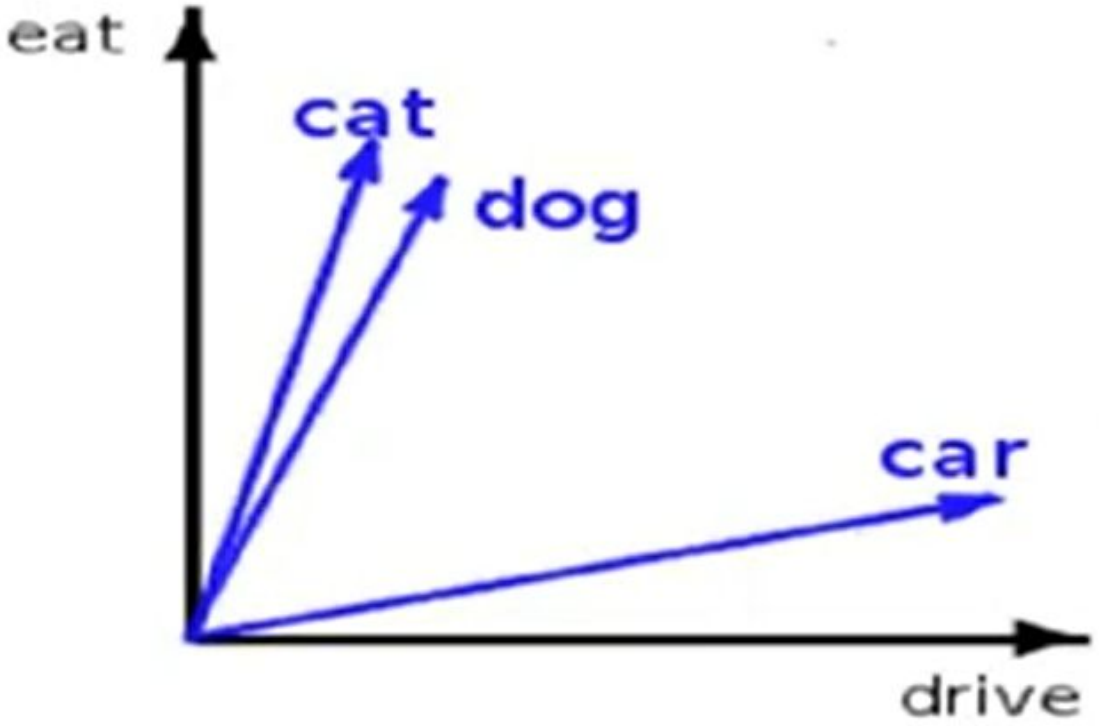
Distributional Semantics: General Intuition

- **Distributions** are vectors in a multidimensional semantic space, that is, objects with a magnitude and a direction.
- The **semantic space** has dimensions which correspond to possible contexts, as gathered from a given corpus.

Vector Space



Vector Space



In practice, many more dimensions are used.

cat = [...dog 0.8, eat 0.7, joke 0.01, mansion 0.2,...]

Word Space

Small Dataset

An automobile is a wheeled motor vehicle used for transporting passengers .
A car is a form of transport , usually with four wheels and the capacity to carry around five passengers .
Transport for the London games is limited , with spectators strongly advised to avoid the use of cars .
The London 2012 soccer tournament began yesterday , with plenty of goals in the opening matches .
Giggs scored the first goal of the football tournament at Wembley , North London .
Bellamy was largely a passenger in the football match , playing no part in either goal .

Target words: ⟨automobile, car, soccer, football⟩
Term vocabulary: ⟨wheel, transport, passenger, tournament, London, goal, match⟩

Constructing Word Space

Informal algorithm for constructing word spaces

- Pick the words you are interested in: **target words**
- Define a **context window**, number of words surrounding target word
  - The context can in general be defined in terms of documents, paragraphs or sentences.
- Count number of times the target word co-occurs with the context words: **co-occurrence matrix**
- Build vectors out of (a function of) these co-occurrence counts

Constructing Word Space: Distributional Vector

distributional matrix = targets X contexts

	wheel	transport	passenger	tournament	London	goal	match
automobile	1	1	1	0	0	0	0
car	1	2	1	0	1	0	0
soccer	0	0	0	1	1	1	1
football	0	0	1	1	1	2	1


football (0,2)
soccer (0,1)
automobile (1,0)
car (2,0)
goal
transport

Computing Similarity

	wheel	transport	passenger	tournament	London	goal	match
automobile	1	1	1	0	0	0	0
car	1	2	1	0	1	0	0
soccer	0	0	0	1	1	1	1
football	0	0	1	1	1	2	1

Using simple vector product

automobile . car = 4
automobile . soccer = 0
automobile . football = 1
car . soccer = 1
car . football = 2
soccer . football = 5

Thank you.