Computational Modeling in Python: Distributional Semantic Models

Raquel G. Alhama & Andrew Jessop

EMLAR 2021

Plan for today

1. Computational Modelling: what is it, and why should we care?
2. Our case study for today:
   Learning Words from Context
3. Representing word meaning from context:
   Distributional Semantic Models
4. Practical: let's implement the model!

Raquel G. Alhama

Tilburg University

Andrew Jessop

University of Liverpool

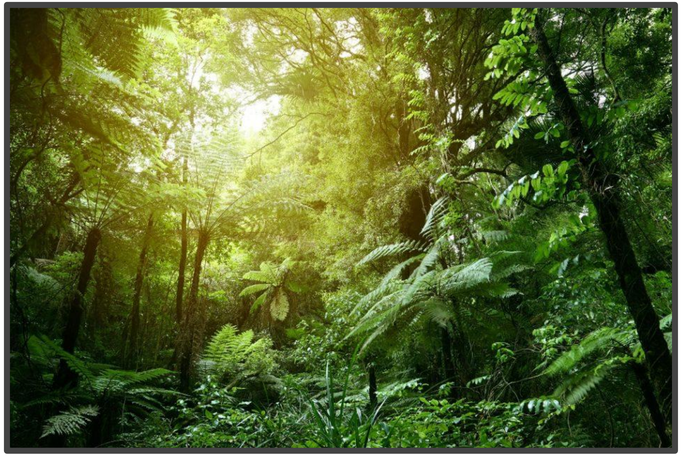
What is a model?

What is a model?

- A model is a **formal** and **simplified** representation of some aspect of reality

What is a model?

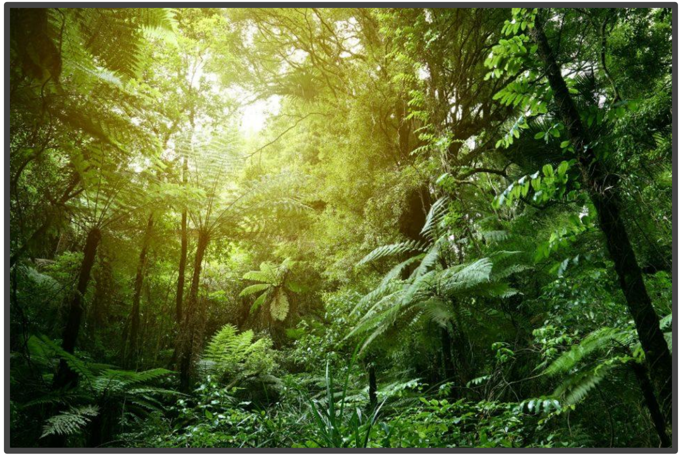
- A model is a **formal** and **simplified** representation of some aspect of reality



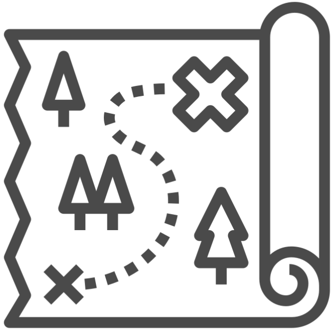
Reality

What is a model?

- A model is a **formal** and **simplified** representation of some aspect of reality



Reality



Simplification

What is a computational model?

What is a computational model?

- A model that can be described in a programming language as a sequence of steps that transform an input x into an output y.

$x \longrightarrow$  $\longrightarrow y$

Why should we use computational models?

- Formalization of a theory
  - Eliminate vagueness of verbal descriptions

Why should we use computational models?

- Formalization of a theory
  - Eliminate vagueness of verbal descriptions
- Prediction of unseen cases

Why should we use computational models?

- Formalization of a theory
  - Eliminate vagueness of verbal descriptions
- Prediction of unseen cases
- Causal explanations and theory building

Why should we use computational models?

- Formalization of a theory
    - Eliminate vagueness of verbal descriptions
- Prediction of unseen cases
- Causal explanations and theory building
- Automatization: we can apply (multiple) models to large amounts of data

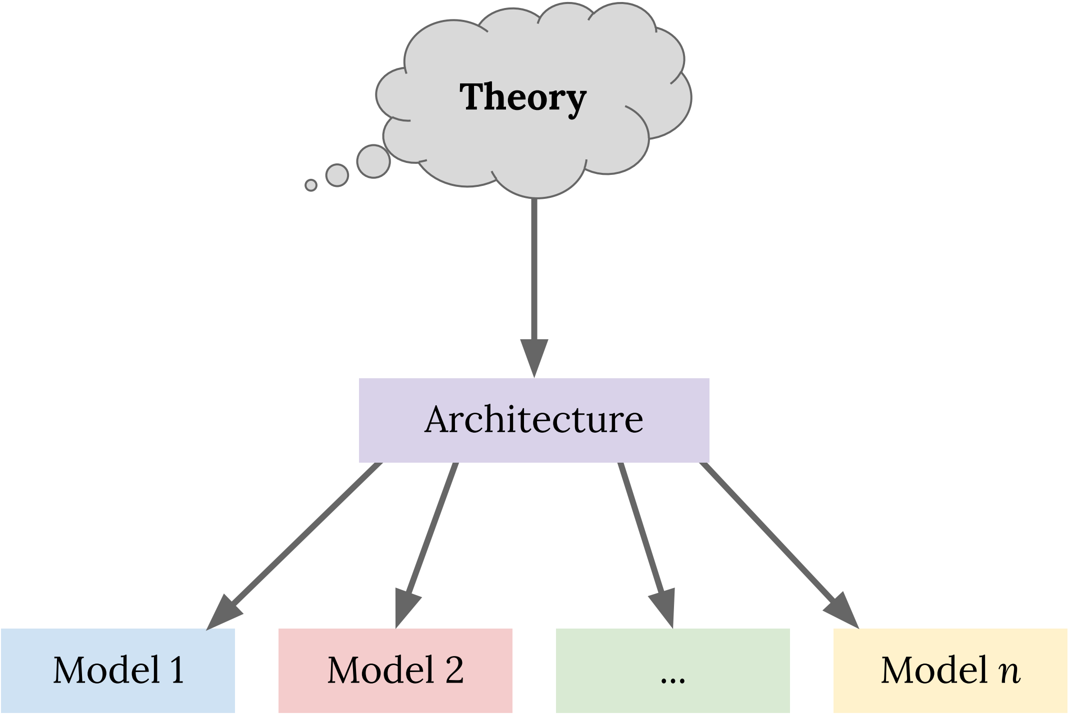
What kinds of computational models exist?

- 1000s of different architectures available
- There are some common ways of classifying models (e.g., symbolic vs. connectionist)
- Often there are multiple ways to model the same behaviour or cognitive processes
- But this is not necessarily a problem
- Different models might serve different goals, while still being derived from the same theory/hypotheses

What kinds of computational models exist?

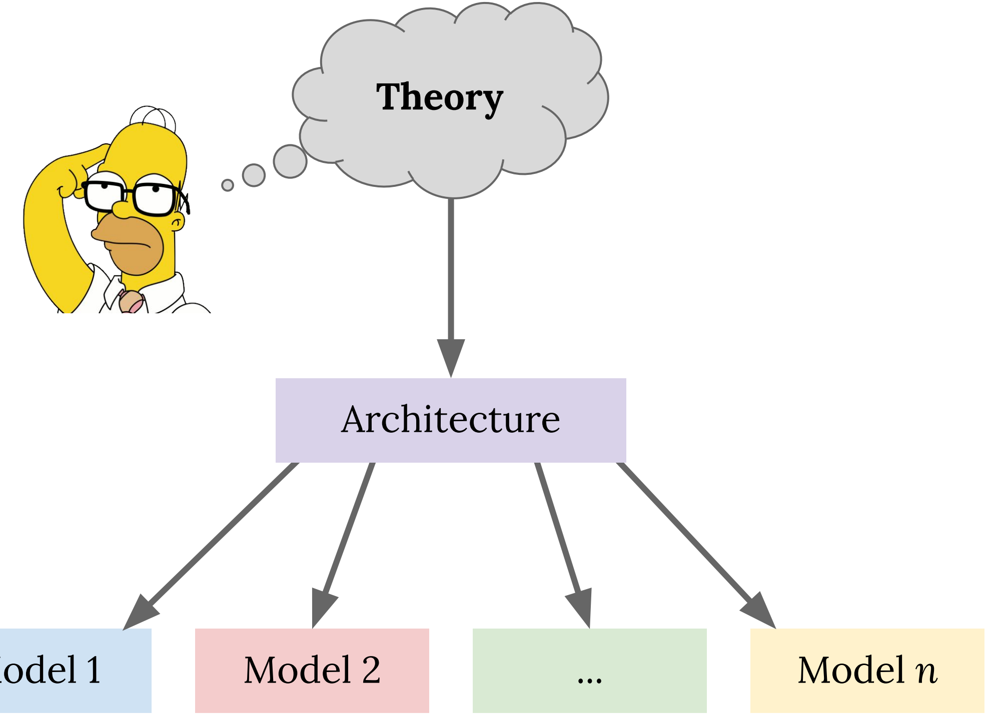
- 1000s of different architectures available
- Most models are designed to answer specific questions
  - How do children segment words? How do people represent semantics?
- There are some common ways of classifying models (e.g., symbolic vs. connectionist)

Lane and Gobet (2012)



Lane and Gobet (2012)

Researcher has a **verbal theory** of how the cognitive system works



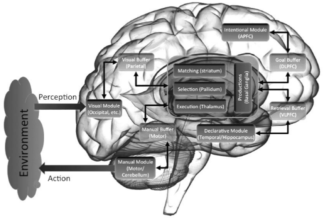
Lane and Gobet (2012)

Researcher has a **verbal theory** of how the cognitive system works

Theory

Architecture

Model 1 | Model 2 | ... | Model *n*



They **design** a **computational system** based on this theory

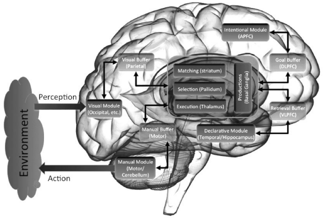
Lane and Gobet (2012)

Researcher has a **verbal theory** of how the cognitive system works

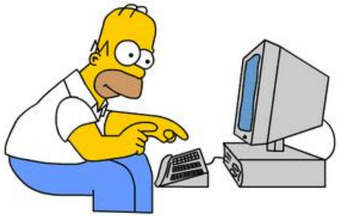
Theory

Architecture

They design a **computational system** based on this theory



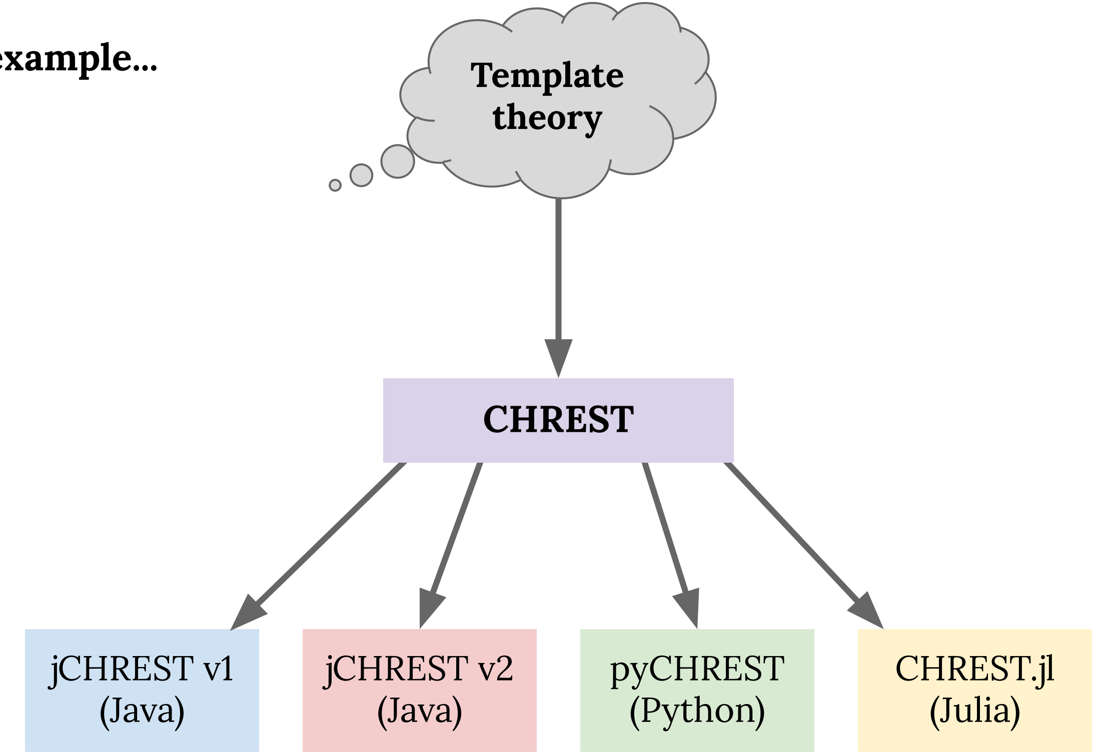
They build a **concrete implementation** of the architecture using code



Model 1 | Model 2 | ... | Model n

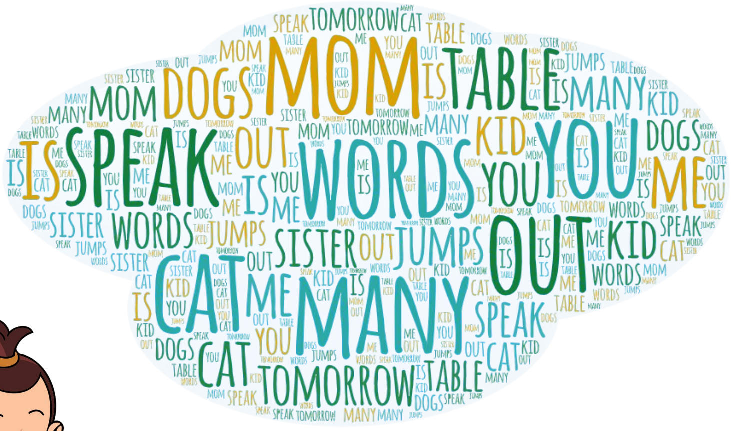
Lane and Gobet (2012)

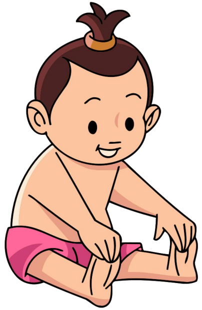
For example...



Today: Learning Words From Linguistic Context

Word Learning





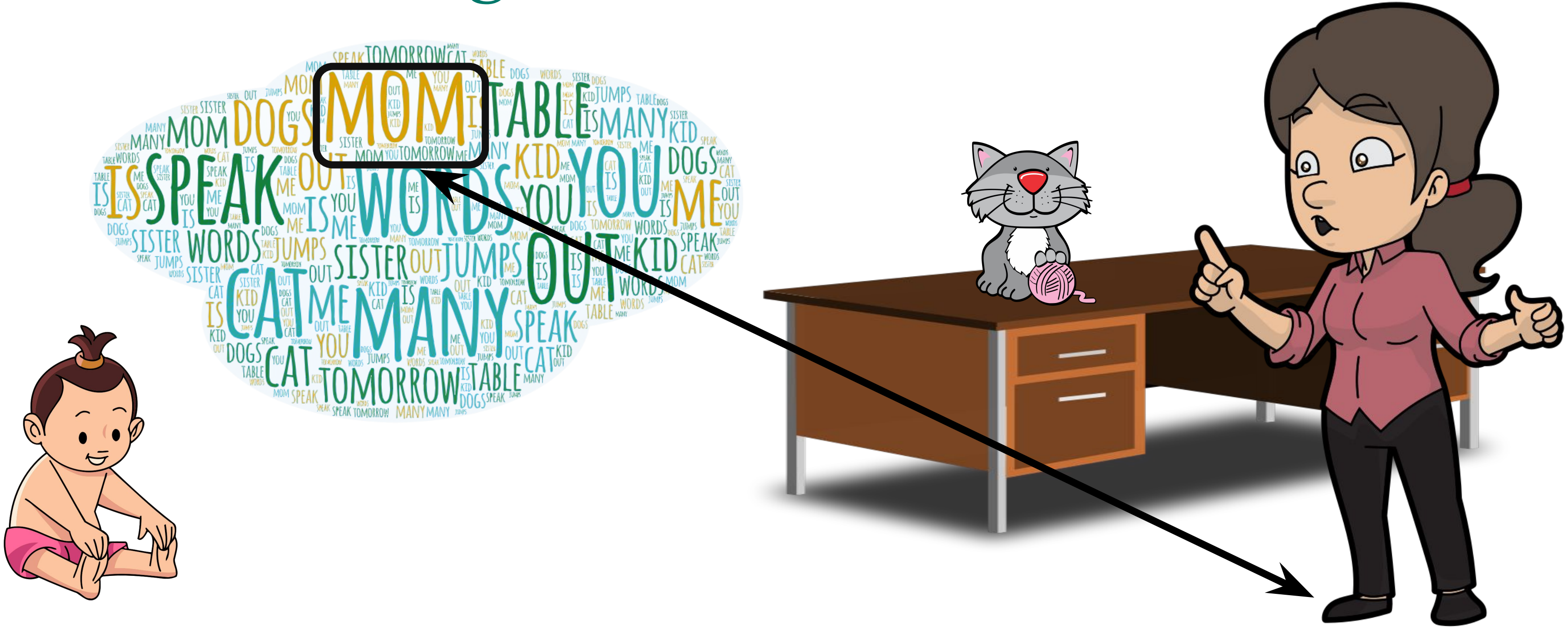
words

meaning

Raquel G. Alhama & Andrew Jessop

EMLAR 2021

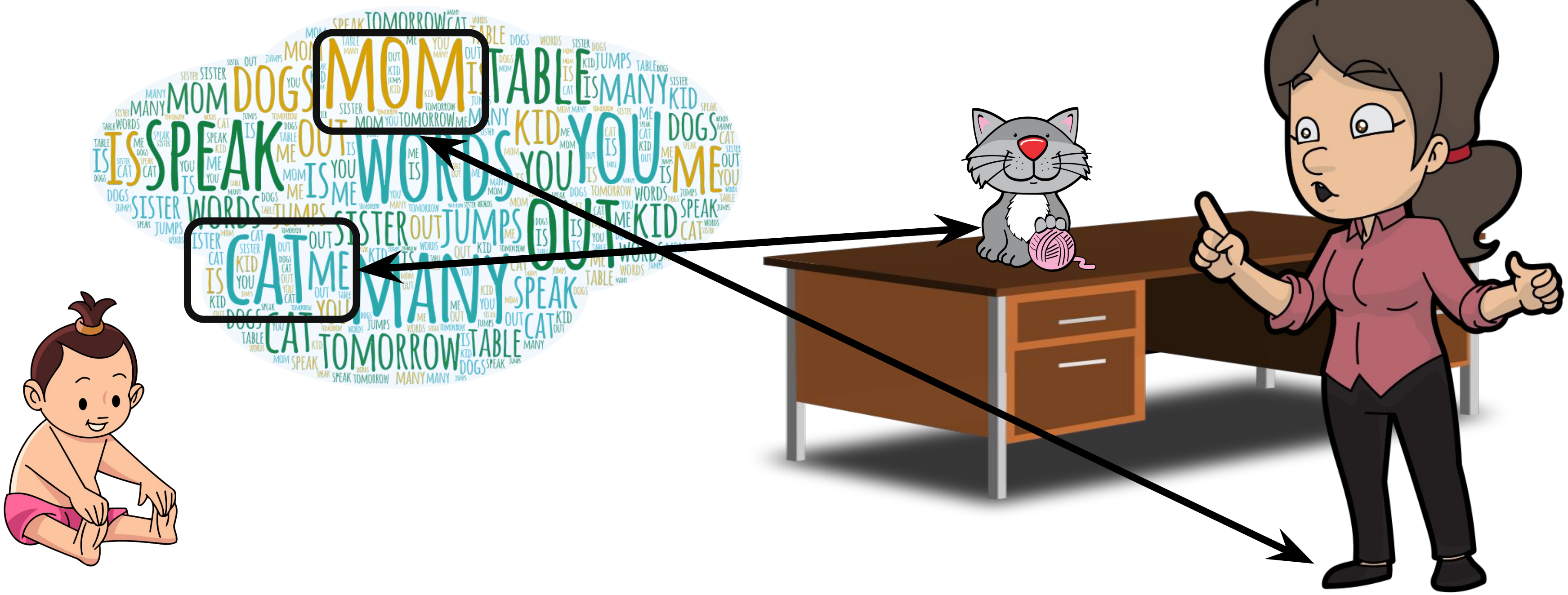
Word Learning



words

meaning

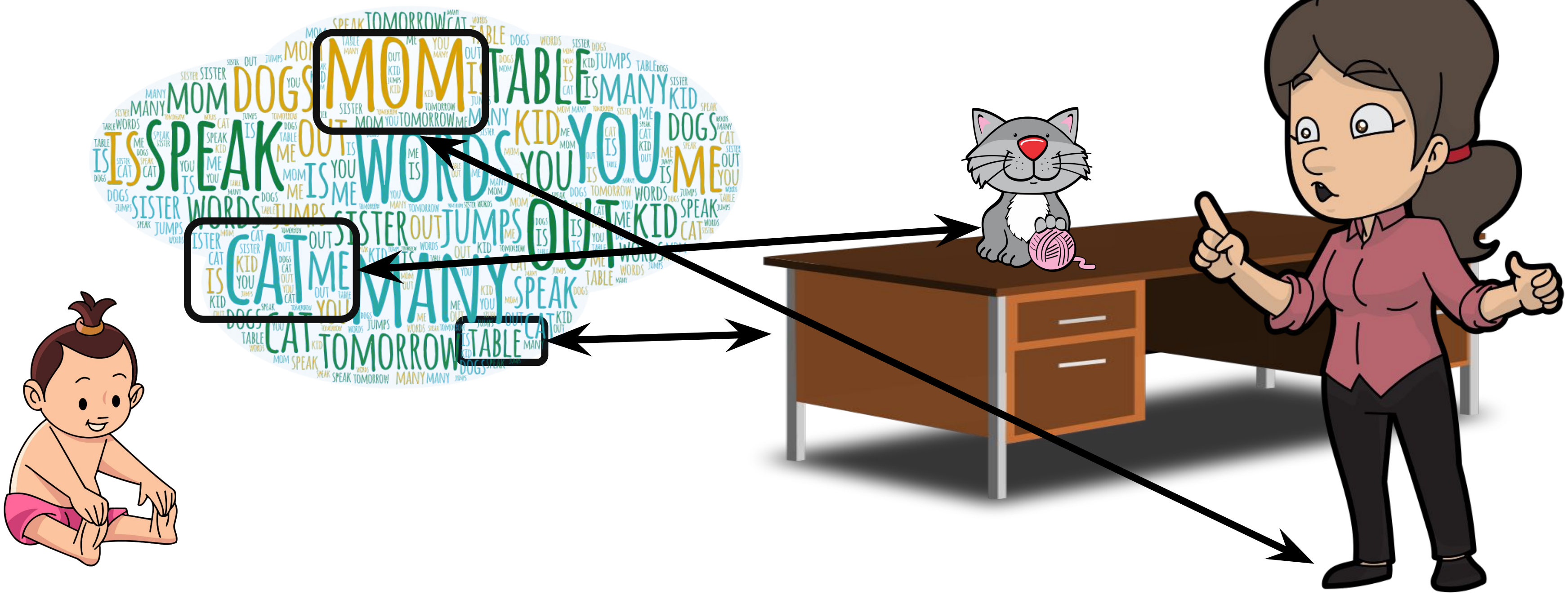
Word Learning



words

meaning

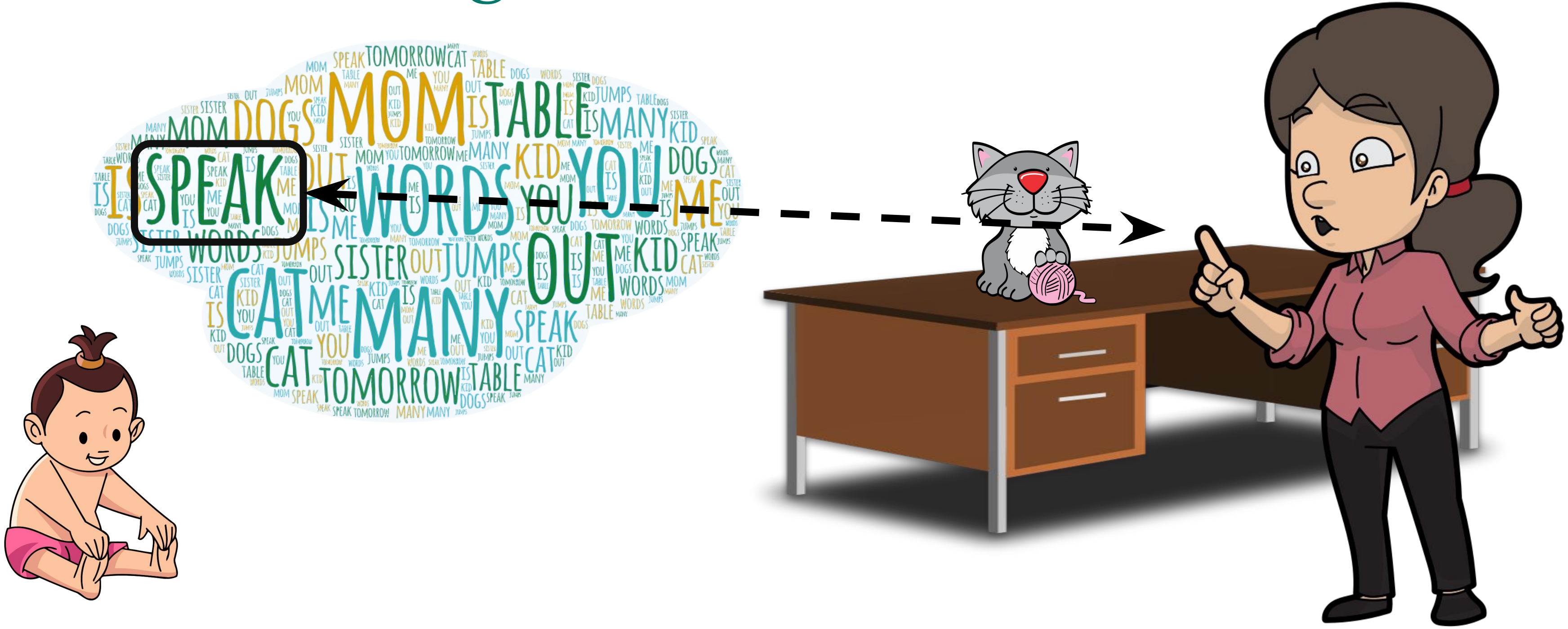
Word Learning



words

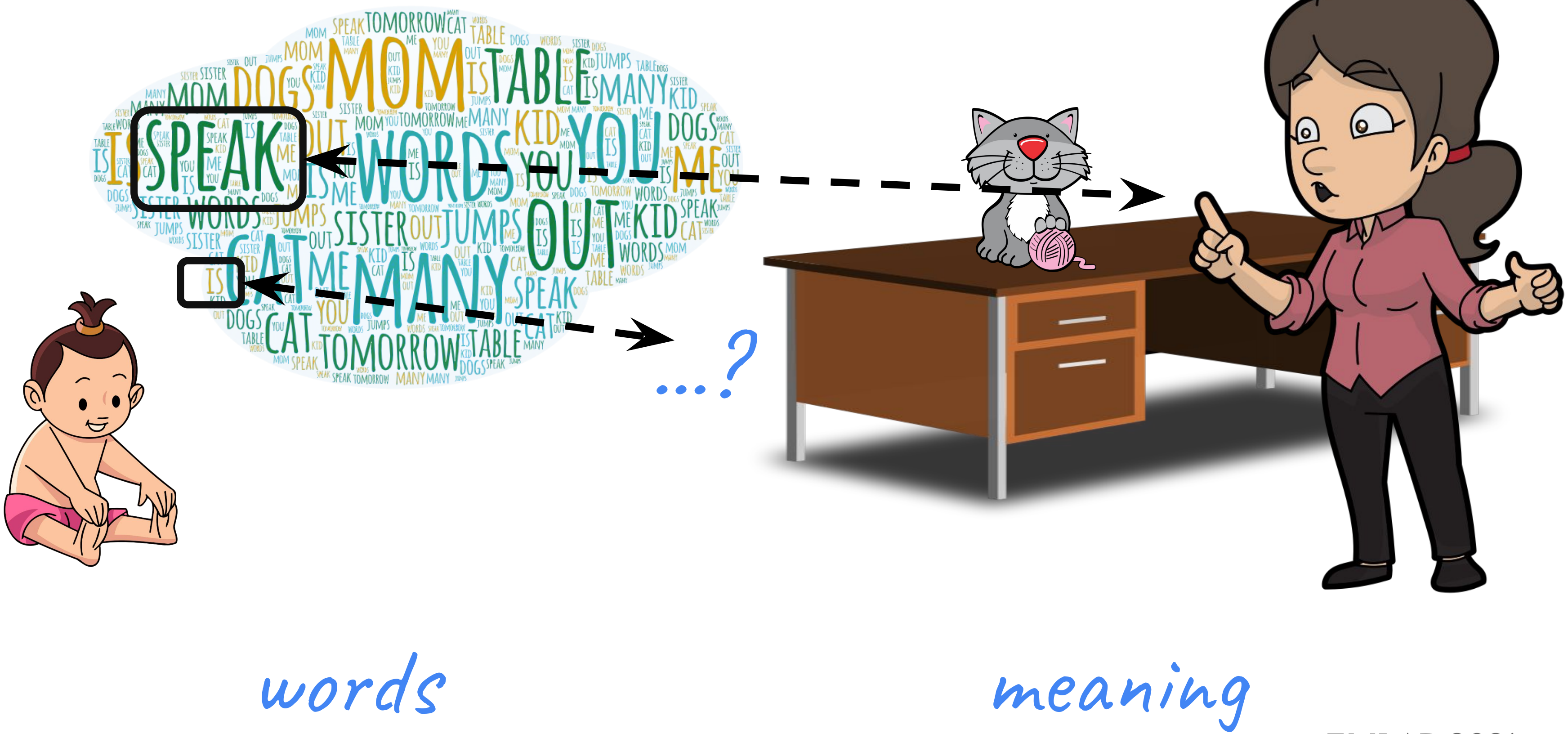
meaning

Word Learning



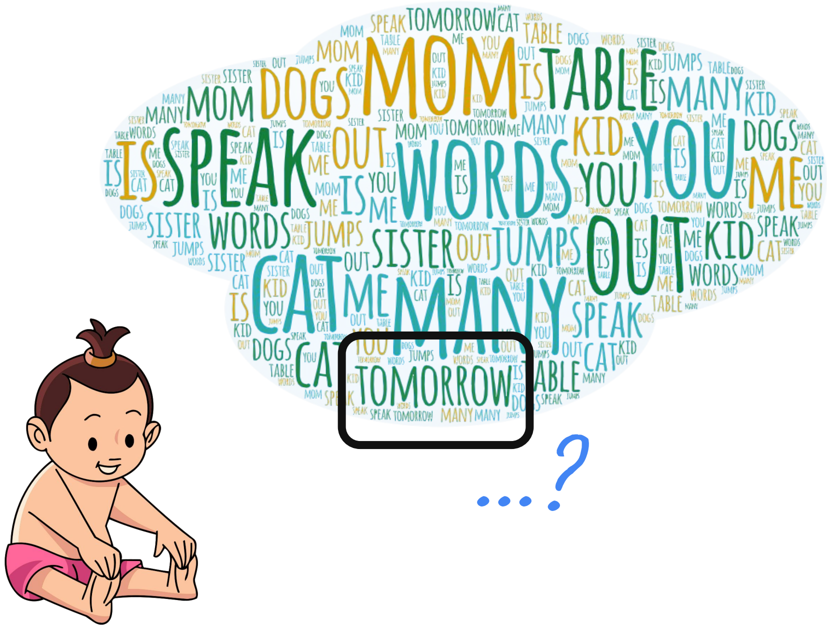
words meaning

Word Learning

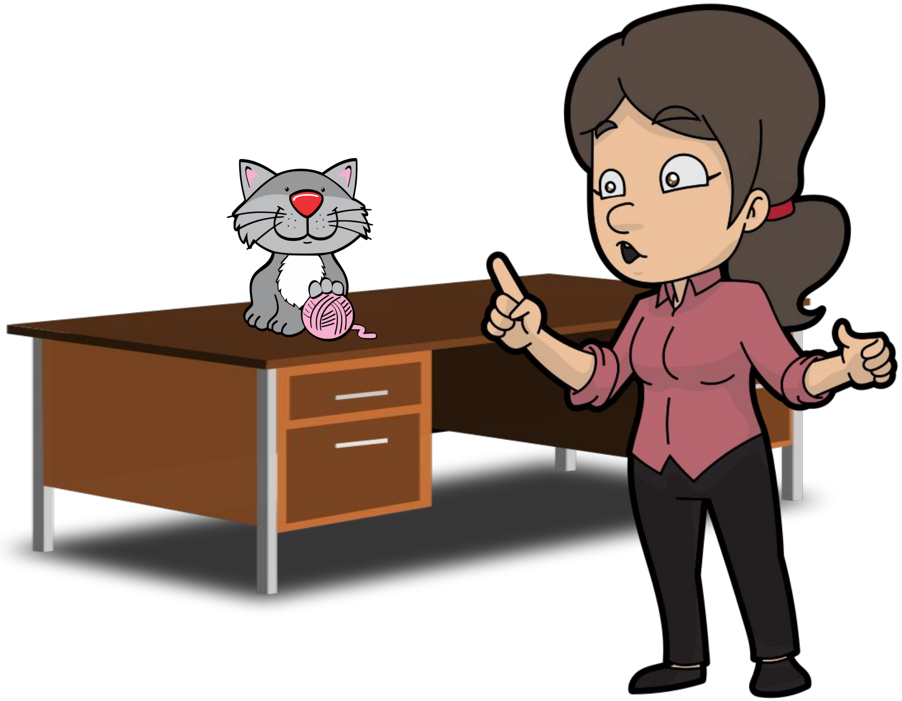


Raquel G. Alhama & Andrew Jessop

Word Learning



words



meaning

Word Learning

- Mapping words to meaning is a difficult task
- How do children constrain the hypothesis space?

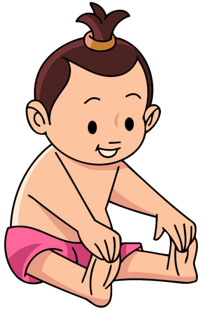
Word Learning

- Mapping words to meaning is a difficult task
- How do children constrain the hypothesis space?
  - Biases to select referent
    - mutual exclusivity bias, whole-object assumption, ...

Word Learning

- Mapping words to meaning is a difficult task
- How do children constrain the hypothesis space?
  - Biases to select referent
    - mutual exclusivity bias, whole-object assumption, ...
  - Cross-situational learning

But… words occur in continuous speech

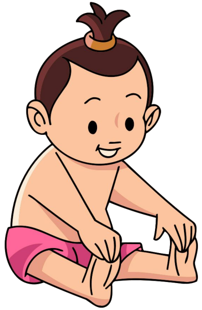


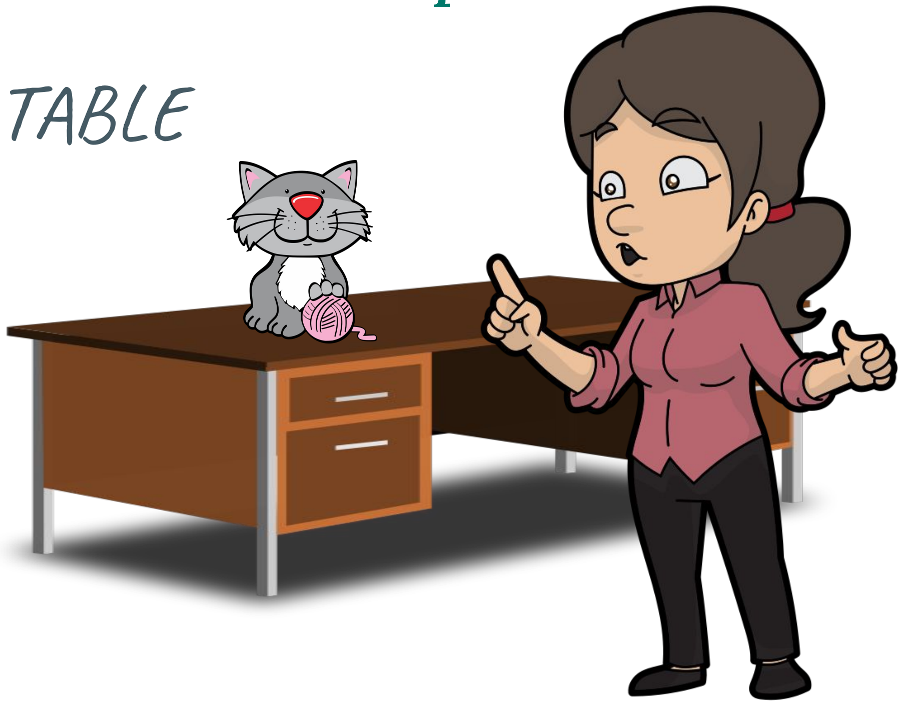
words

meaning

But... words occur in continuous speech

THE $\boxed{\text{CAT}}$ IS ON THE TABLE





words

meaning

Word Learning

- Mapping words to meaning is a difficult task
- How do children constrain the hypothesis space?
  - Biases to select referent
    - mutual exclusivity bias, whole-object assumption, ...
  - Cross-situational learning
  - Distributional information from the linguistic input
    - CONTEXT!

Word Learning from distributional cues

- Children can track distributional information at a very young age (Saffran et al. 1996, Aslin et al. 1998, ...)

Word Learning from distributional cues

- Children can track distributional information at a very young age (Saffran et al. 1996, Aslin et al. 1998, ...)
- How can distributional information shape meaning representations?

Word Learning from distributional cues

- Children can track distributional information at a very young age (Saffran et al. 1996, Aslin et al. 1998, ...)
- How can distributional information shape meaning representations?
  - Let's look at Distributional Semantic Models

Distributional Semantic Models

Raquel G. Alhama & Andrew Jessop

EMLAR 2021

Meaning as use

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

- “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” (Miller, 1986)

The Distributional Hypothesis

- “Difference of meaning correlates with difference of distribution” (Harris, 1954)

- “You shall know a word by the company it keeps” (Firth, 1957)

Distributional Semantic Models

- Distributional Semantic Models (DSMs) derive semantic representations of words based on the distributional hypothesis

- These models are also known as Vector Space Models

Distributional Semantic Models

- The representations have the form of vectors (embeddings)

  e.g.

249	0	365	446	0	0	90	0	2136

- These representations are distributed rather than symbolic
  - similar words should have similar representations!

Distributional Semantic Models

- The number of words that we consider to be part of the context of another word is a parameter called "window size"
- We will now see an example of how we derive word representations in DSMs, for a model with window size = 2

Distributional Semantic Models

THE CAR TRAVELLED AT HIGH SPEED

Distributional Semantic Models

THE CAR TRAVELLED AT HIGH SPEED

	the	car	travelled	at	high	speed	...
the	0	0	0	0	0	0	0
car	0	0	0	0	0	0	0
travelled	0	0	0	0	0	0	0
at	0	0	0	0	0	0	0
high	0	0	0	0	0	0	0
speed	0	0	0	0	0	0	0
...	0	0	0	0	0	0	0

Distributional Semantic Models

THE CAR TRAVELLED AT HIGH SPEED

	the	car	travelled	at	high	speed	...
the	0	1	1	0	0	0	0
car	0	0	0	0	0	0	0
travelled	0	0	0	0	0	0	0
at	0	0	0	0	0	0	0
high	0	0	0	0	0	0	0
speed	0	0	0	0	0	0	0
...	0	0	0	0	0	0	0

Distributional Semantic Models

THE CAR TRAVELLED AT HIGH SPEED

	the	car	travelled	at	high	speed	...
the	0	1	1	0	0	0	0
car	1	0	1	1	0	0	0
travelled	0	0	0	0	0	0	0
at	0	0	0	0	0	0	0
high	0	0	0	0	0	0	0
speed	0	0	0	0	0	0	0
...	0	0	0	0	0	0	0

Distributional Semantic Models

THE CAR TRAVELLED AT HIGH SPEED

	the	car	travelled	at	high	speed	...
the	0	1	1	0	0	0	0
car	1	0	1	1	0	0	0
travelled	1	1	0	1	1	0	0
at	0	0	0	0	0	0	0
high	0	0	0	0	0	0	0
speed	0	0	0	0	0	0	0
...	0	0	0	0	0	0	0

Distributional Semantic Models

THE CAR TRAVELLED AT HIGH SPEED

	the	car	travelled	at	high	speed	...
the	0	1	1	0	0	0	0
car	1	0	1	1	0	0	0
travelled	1	1	0	1	1	0	0
at	0	1	1	0	1	1	0
high	0	0	0	0	0	0	0
speed	0	0	0	0	0	0	0
...	0	0	0	0	0	0	0

Distributional Semantic Models

THE CAR TRAVELLED AT HIGH SPEED

	the	car	travelled	at	high	speed	…
the	0	1	1	0	0	0	0
car	1	0	1	1	0	0	0
travelled	1	1	0	1	1	0	0
at	0	1	1	0	1	1	0
high	0	0	1	1	0	1	0
speed	0	0	0	0	0	0	0
…	0	0	0	0	0	0	0

Distributional Semantic Models

THE CAR TRAVELLED AT HIGH SPEED

	the	car	travelled	at	high	speed	...
the	0	1	1	0	0	0	0
car	1	0	1	1	0	0	0
travelled	1	1	0	1	1	0	0
at	0	1	1	0	1	1	0
high	0	0	1	1	0	1	0
speed	0	0	0	1	1	0	0
...	0	0	0	0	0	0	0

Distributional Semantic Models

	the	car	travelled	at	high	speed	...
the	0	1	1	0	0	0	0
car	1	0	1	1	0	0	0
travelled	1	1	0	1	1	0	0
at	0	1	1	0	1	1	0
high	0	0	1	1	0	1	0
speed	0	0	0	1	1	0	0
...	0	0	0	0	0	0	0

Distributional Semantic Models

THE TRUCK TRAVELLED FAST

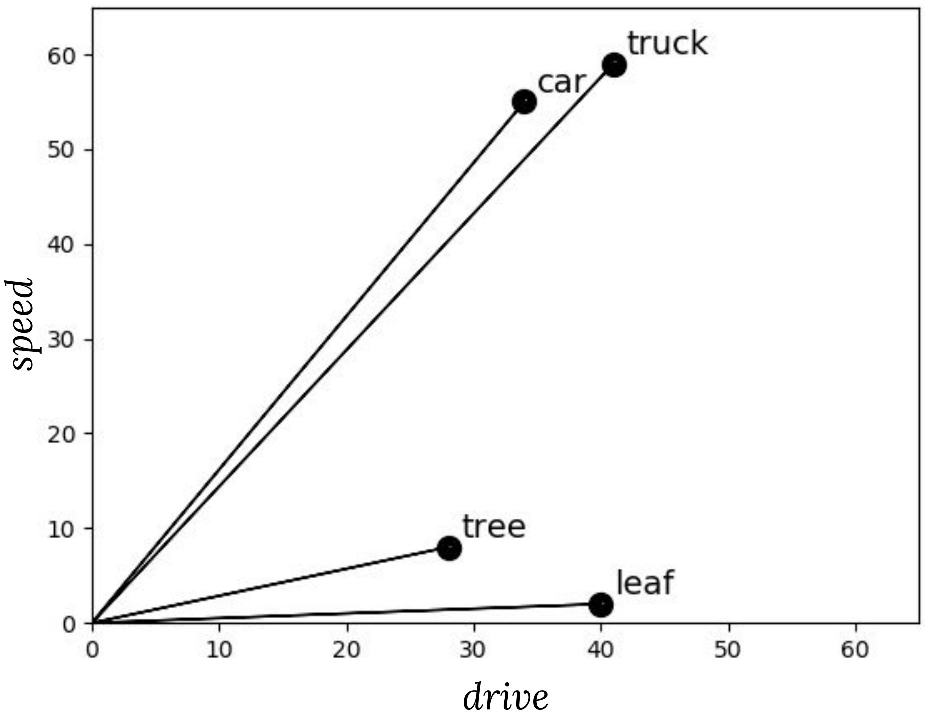
	the	car	travelled	at	high	speed	truck	fast
the	0	1	2	0	0	0	1	0
car	1	0	1	1	0	0	0	0
travelled	2	1	0	1	1	0	1	1
at	0	1	1	0	1	1	0	0
high	0	0	1	1	0	1	0	0
speed	0	0	0	1	1	0	0	0
truck	1	0	1	0	0	0	0	1
fast	0	0	1	0	0	0	1	0

Distributional Semantic Models

	the	car	travelled	at	high	speed	truck	fast
the	0	1	2	0	0	0	1	0
car	1	0	1	1	0	0	0	0
travelled	2	1	0	1	1	0	1	1
at	0	1	1	0	1	1	0	0
high	0	0	1	1	0	1	0	0
speed	0	0	0	1	1	0	0	0
truck	1	0	1	0	0	0	0	1
fast	0	0	1	0	0	0	1	0

- This is a term-term co-occurrence matrix (for a toy example)
- The vectors reflect the use of words in real linguistic productions.
- Similar vectors for semantically related words!

Distributional Semantic Models

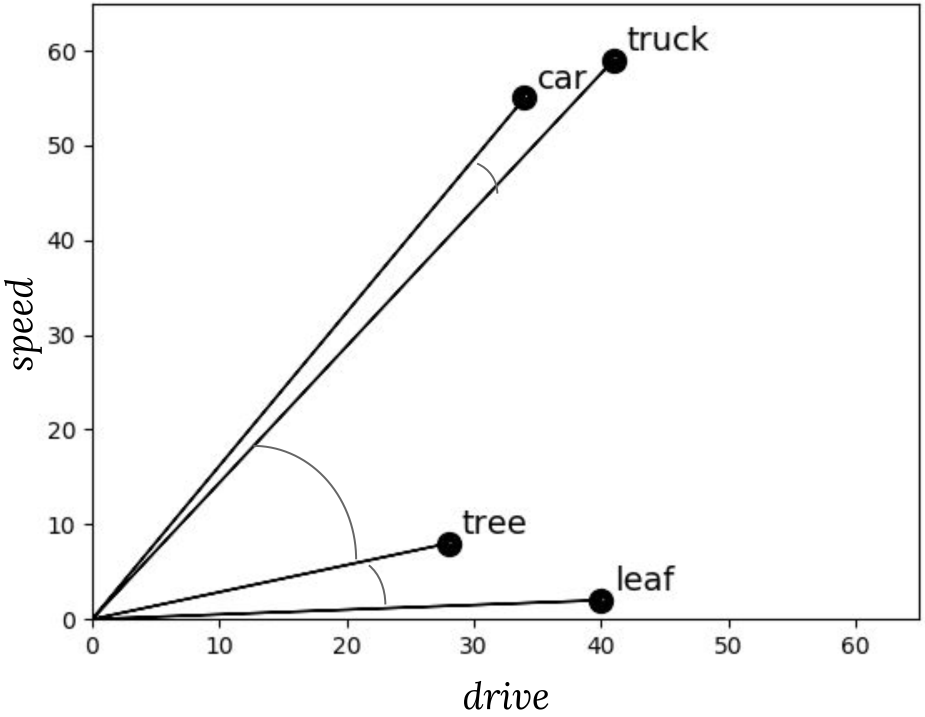


- We can interpret each position in an n-dimensional vector as a coordinate of a point in n-dimensional vector space

  (graph reduced to 2 dimensions)

- Thus, these word representations have geometric properties (e.g. distance between words!)

Distributional Semantic Models



- Cosine similarity: cosine of the angle of the vectors

$$\cos(\Theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|}$$

The cosine is between [-1, 1]. A measure of 0 indicates orthogonal vectors (completely unrelated).

Let's implement this!