Course overview

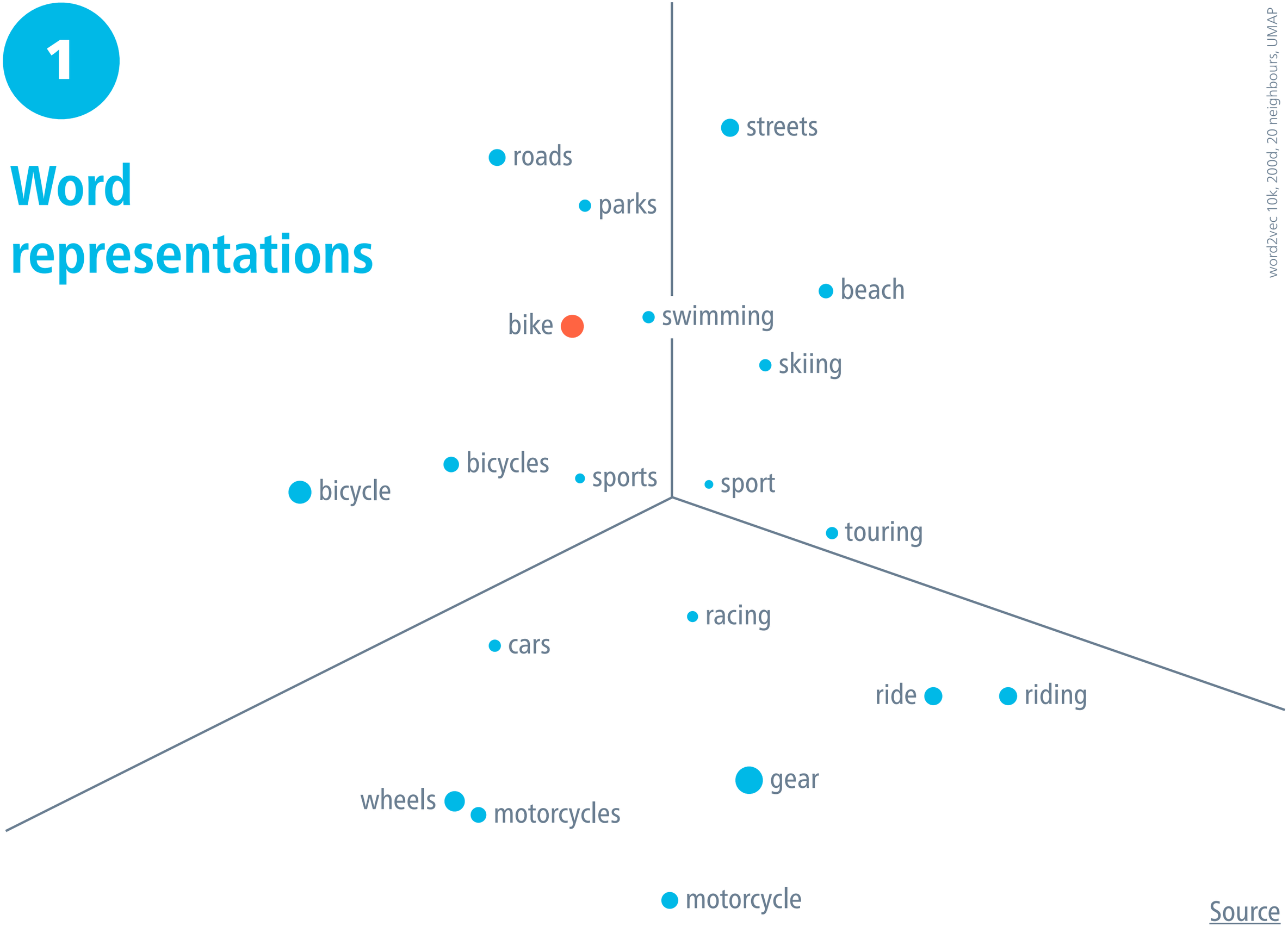
Marco Kuhlmann
Department of Computer and Information Science

LINKÖPING UNIVERSITY

This work is licensed under a
Creative Commons Attribution 4.0 International License.

1

Word representations

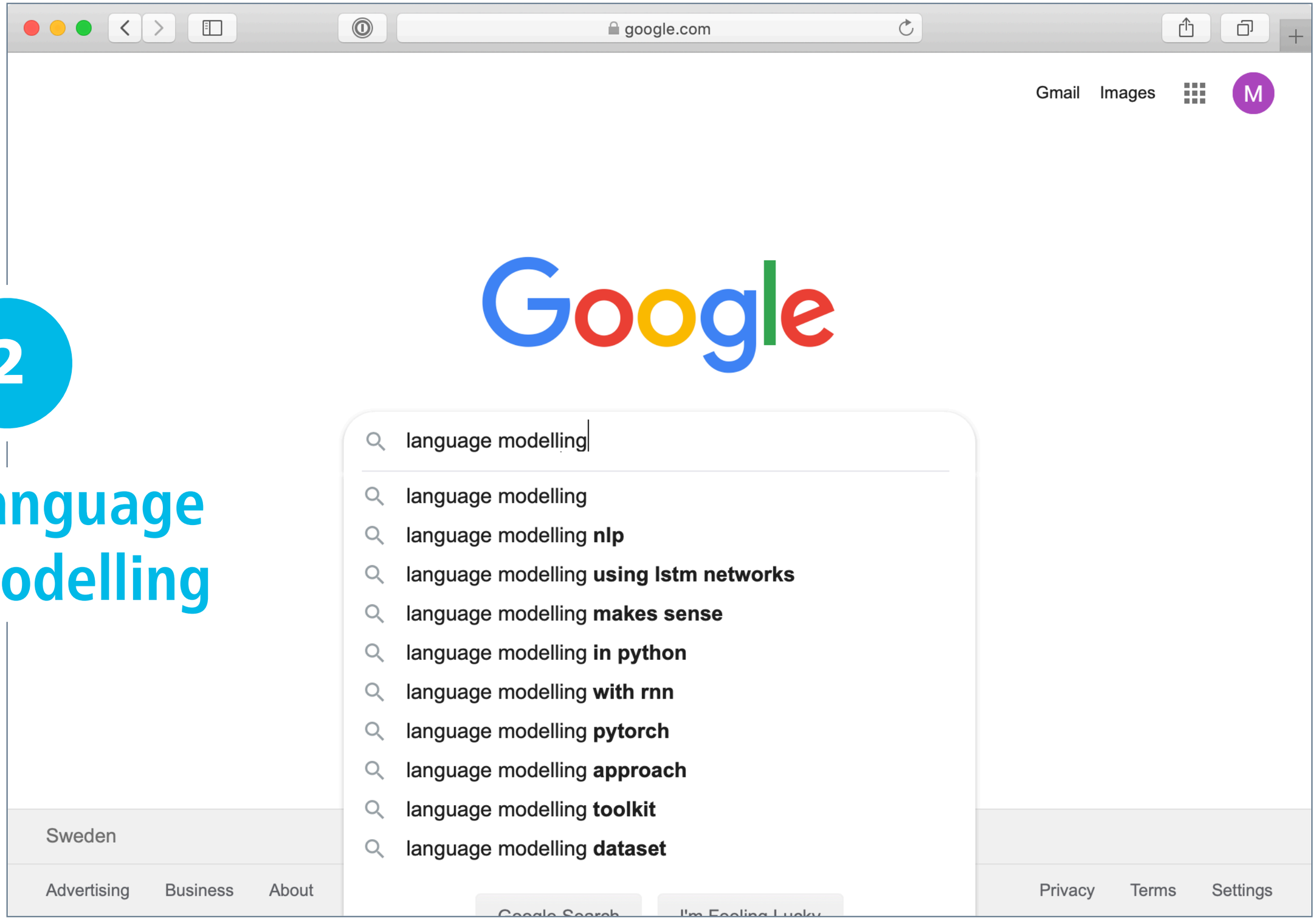


word2vec 10k, 200d, 20 neighbours, UMAP

Source

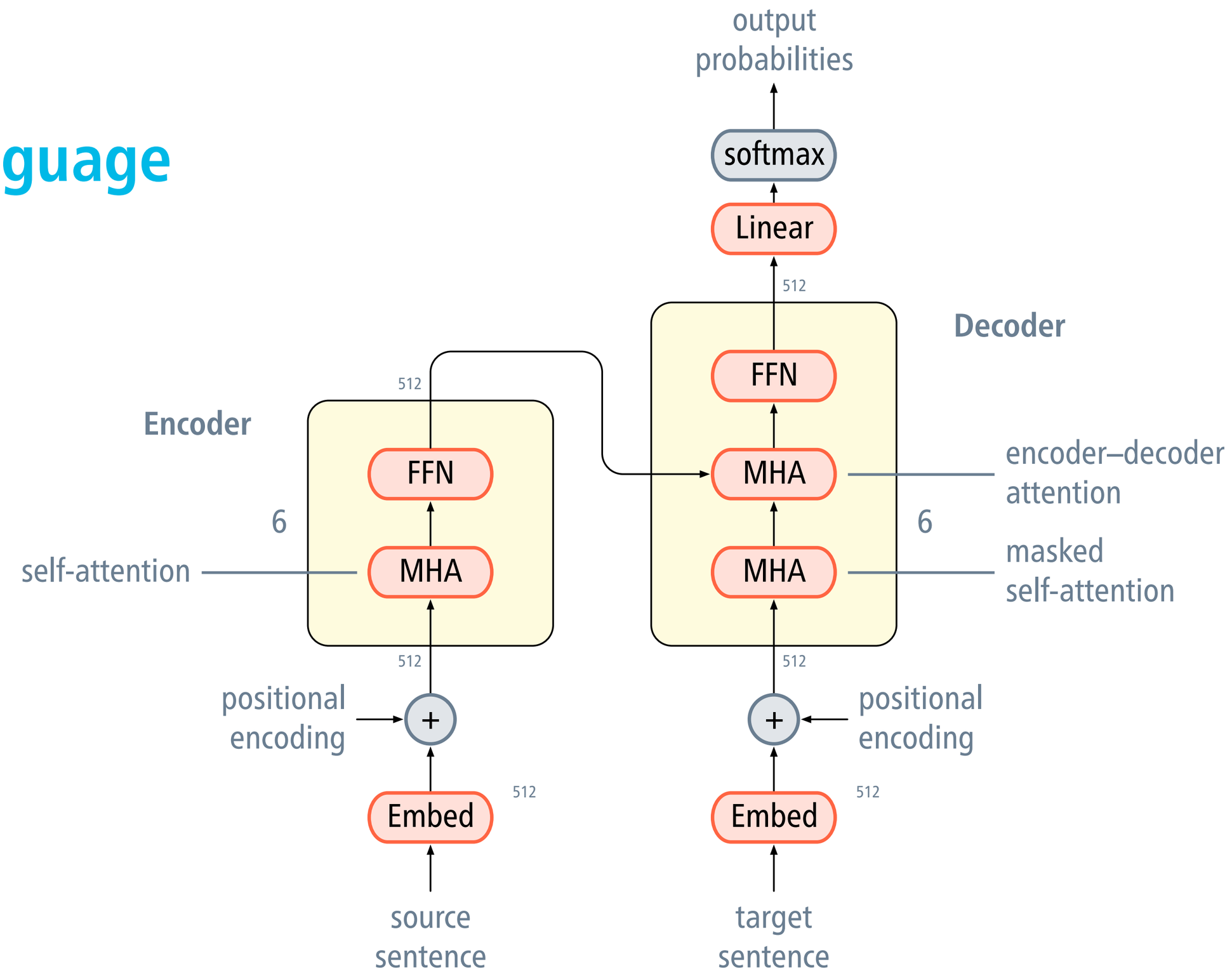
2

Language modelling



3

Large language models

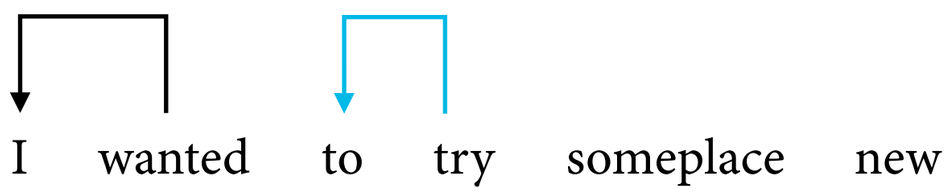


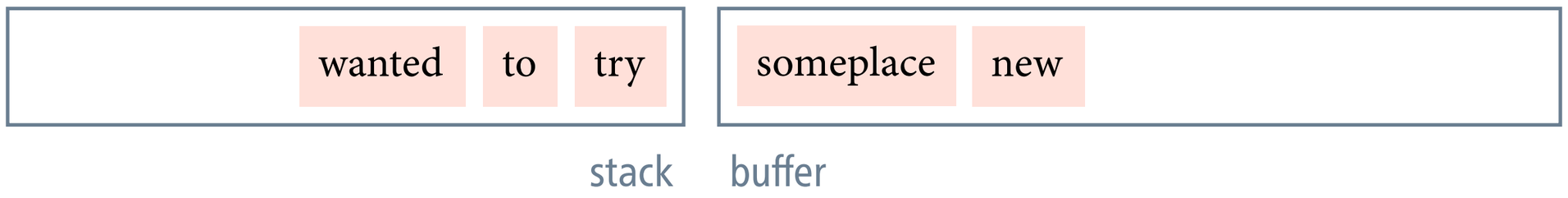
4 Sequence labelling

I	want	to	live	in	peace
PRON	VERB	PART	VERB	ADP	NOUN
NOUN	NOUN	ADP	ADJ	ADV	VERB
		ADV	ADV	ADJ	
				NOUN	

‘I only want to live in peace, plant potatoes, and dream!’ – Moomin

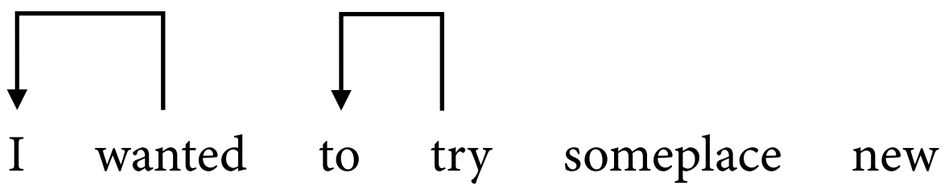
5 Dependency parsing

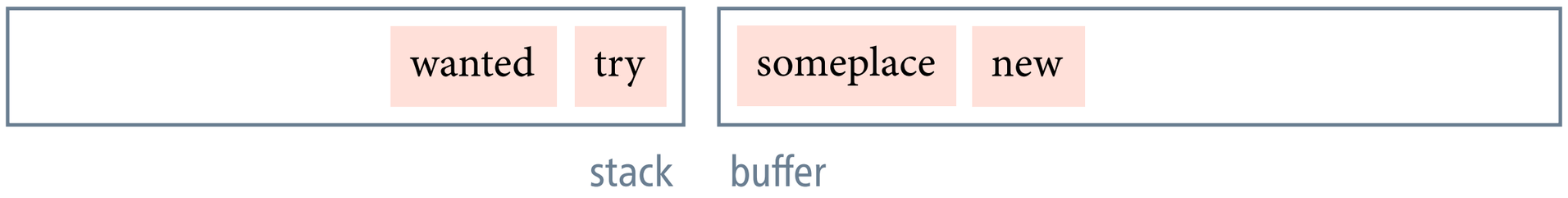




classifier

5 Dependency parsing





classifier

Search and learning

predicted output

input

$$\hat{y} = \underset{y}{\operatorname{argmax}}\ \operatorname{score}(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{\theta})$$

candidate output

model parameters

Eisenstein (2019), § 1.2.2

Search and learning

Eisenstein (2019), § 1.2.2

- **Search module**

  The search module is responsible for finding the candidate output $\boldsymbol{y}$ with the highest score relative to the input $\boldsymbol{x}$.

  requires efficient algorithms

- **Learning module**

  The learning module is responsible for finding the model parameters $\boldsymbol{\theta}$ that maximize the predictive performance.

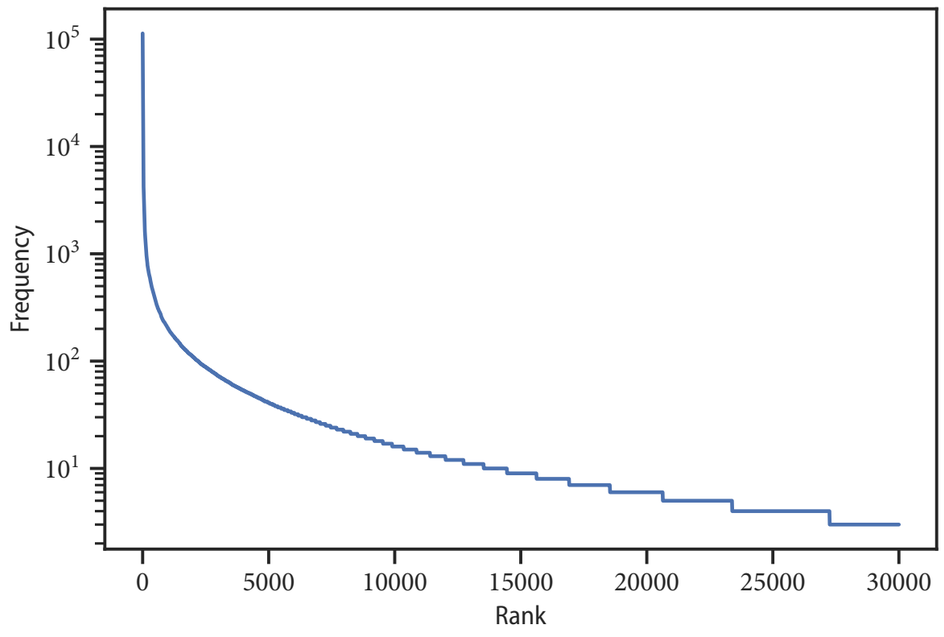
  for example, using supervised machine learning

Language is special

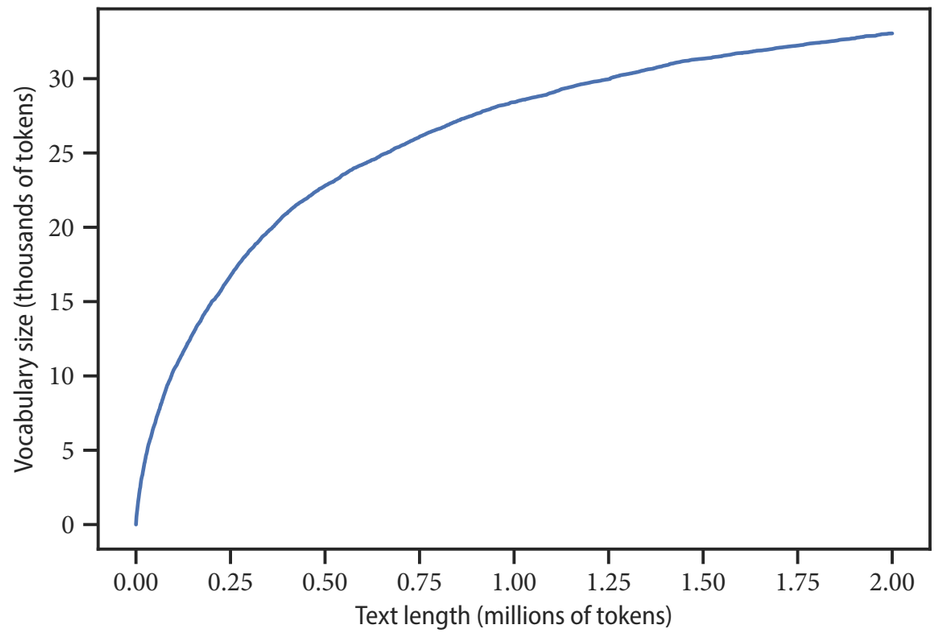
Eisenstein (2019), § 1.1

- Unlike images or audio, text data is fundamentally discrete, with meaning created by combinatorial arrangement.
- Even though text appears as a sequence, machine learning methods must account for its implicit hierarchical structure.
- The distribution of linguistic elements follows a power law – algorithms must be robust to unobserved events.

Zipf’s law and Heaps’ law



Zipf’s law



Heaps’ law