

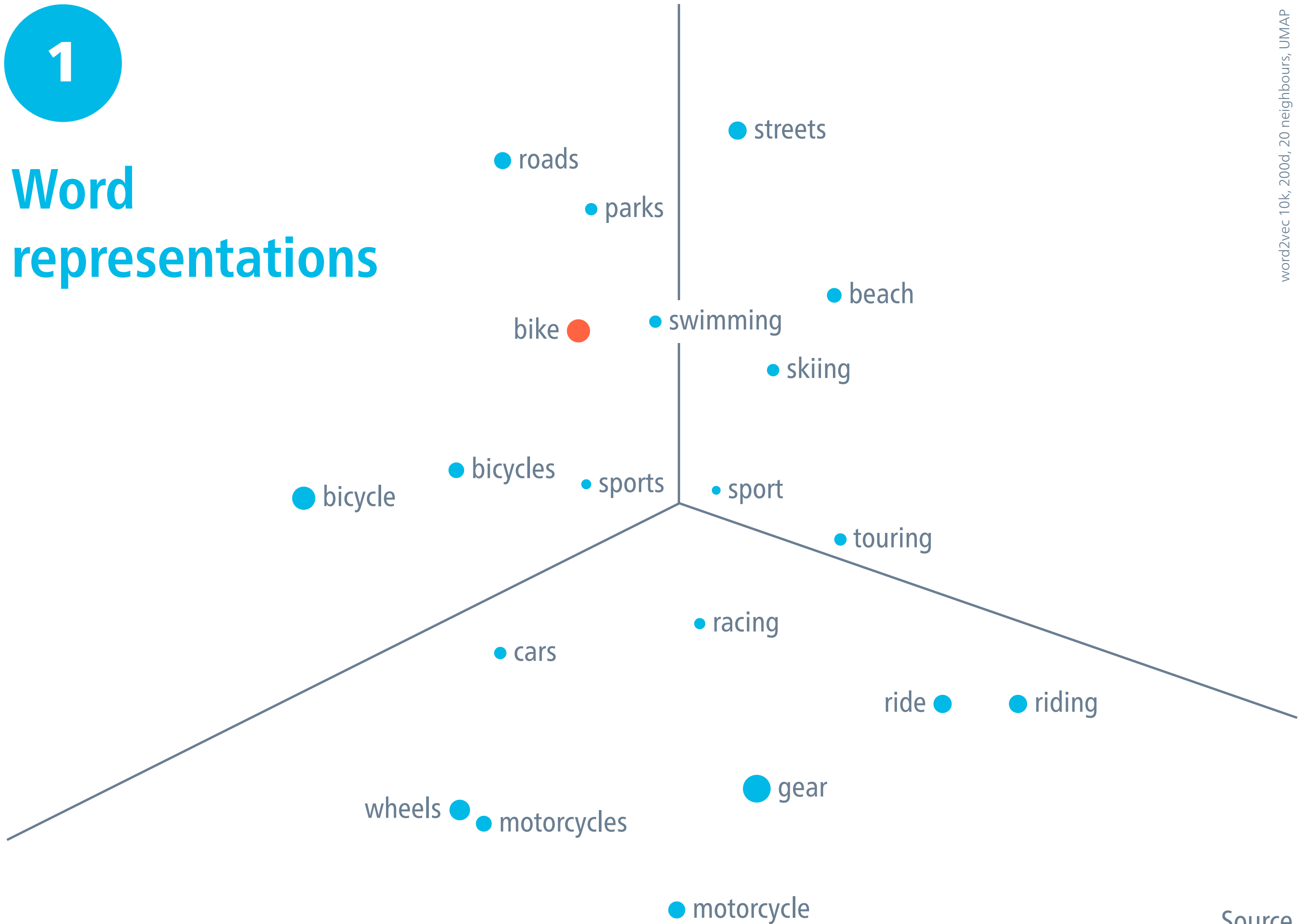
Course overview

Marco Kuhlmann

Department of Computer and Information Science

1

Word representations

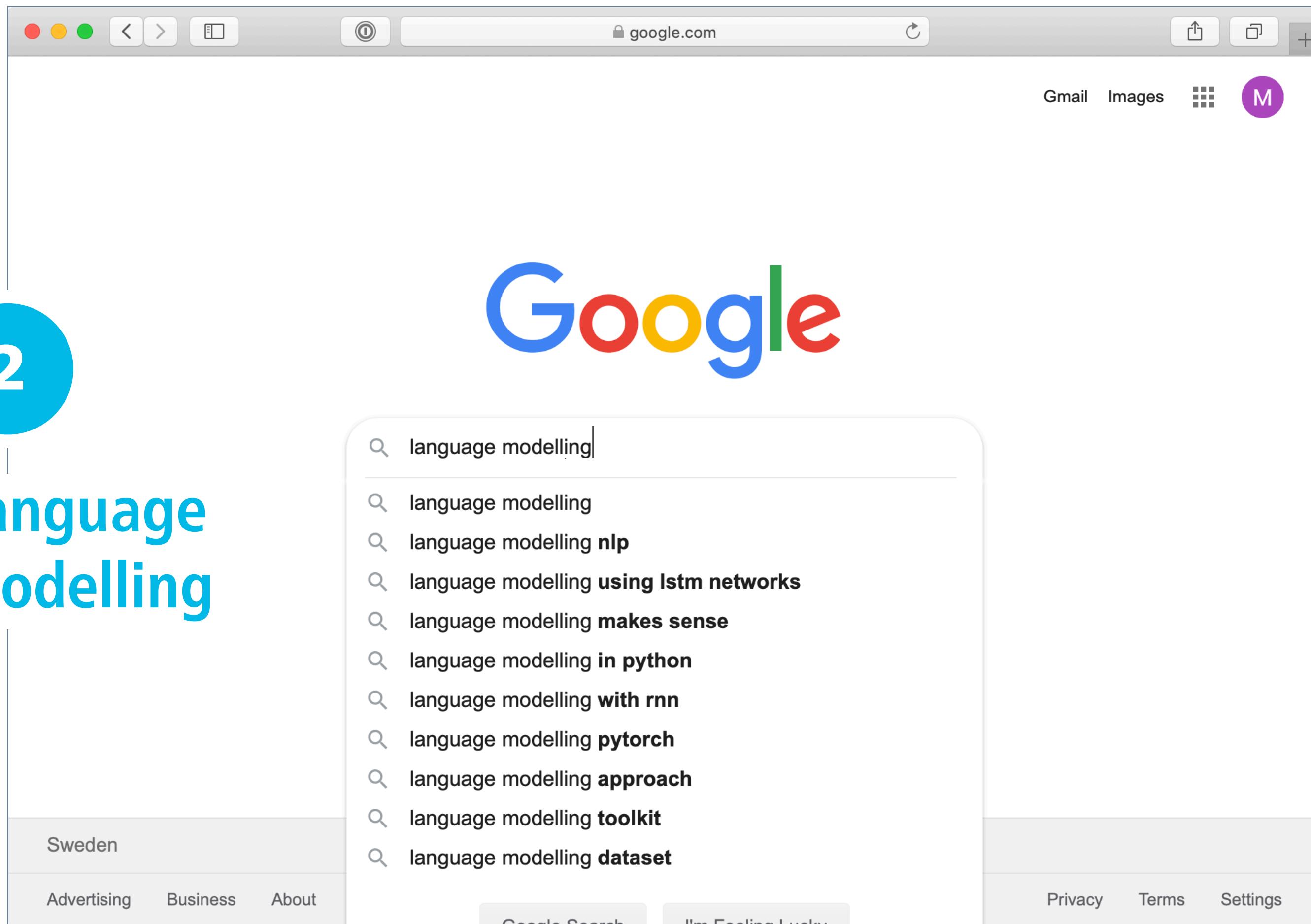


word2vec 10k, 200d, 20 neighbours, UMAP

Source

2

Language modelling



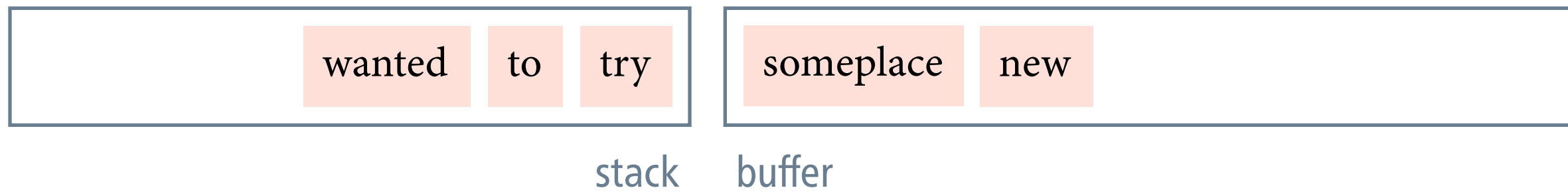
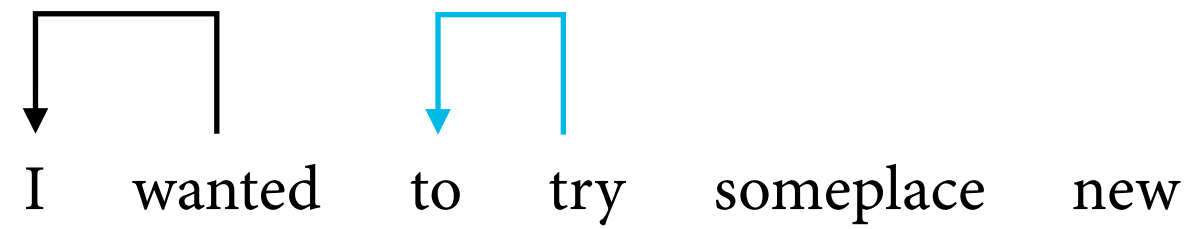
3

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

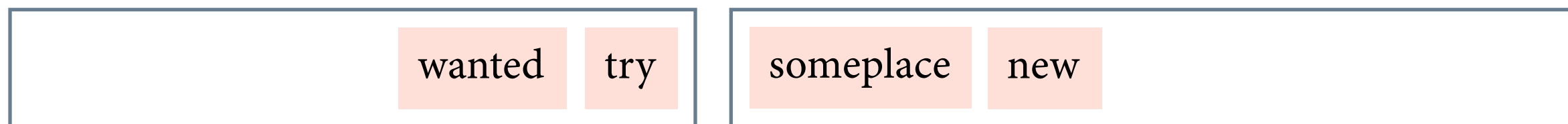
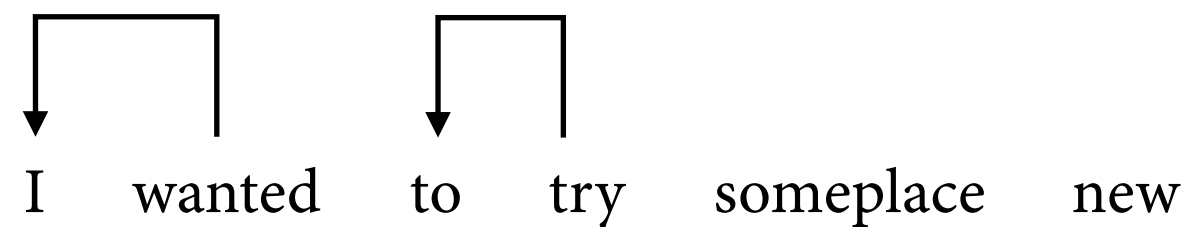
4 Dependency parsing



LA
classifier

4

Dependency parsing

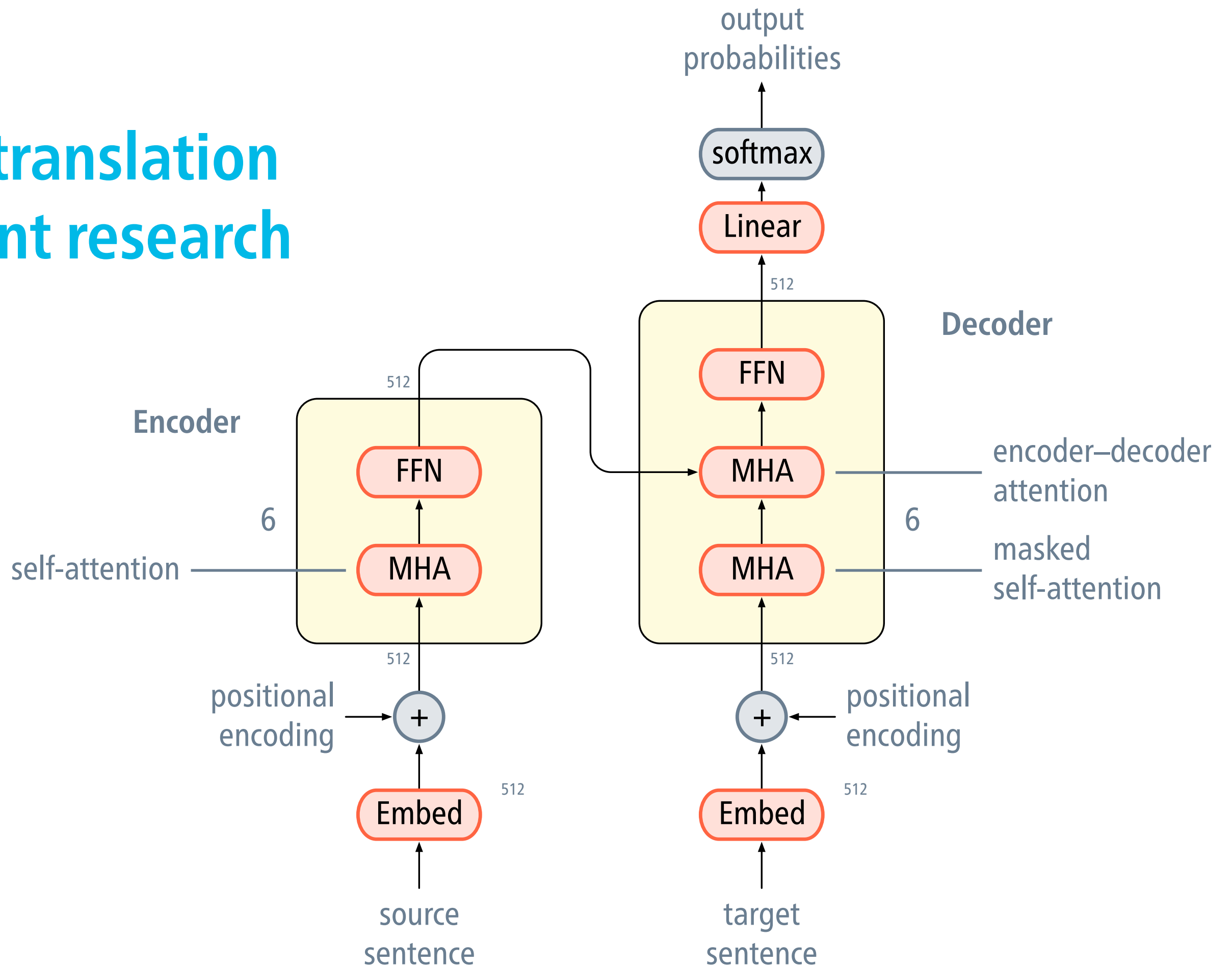


stack buffer

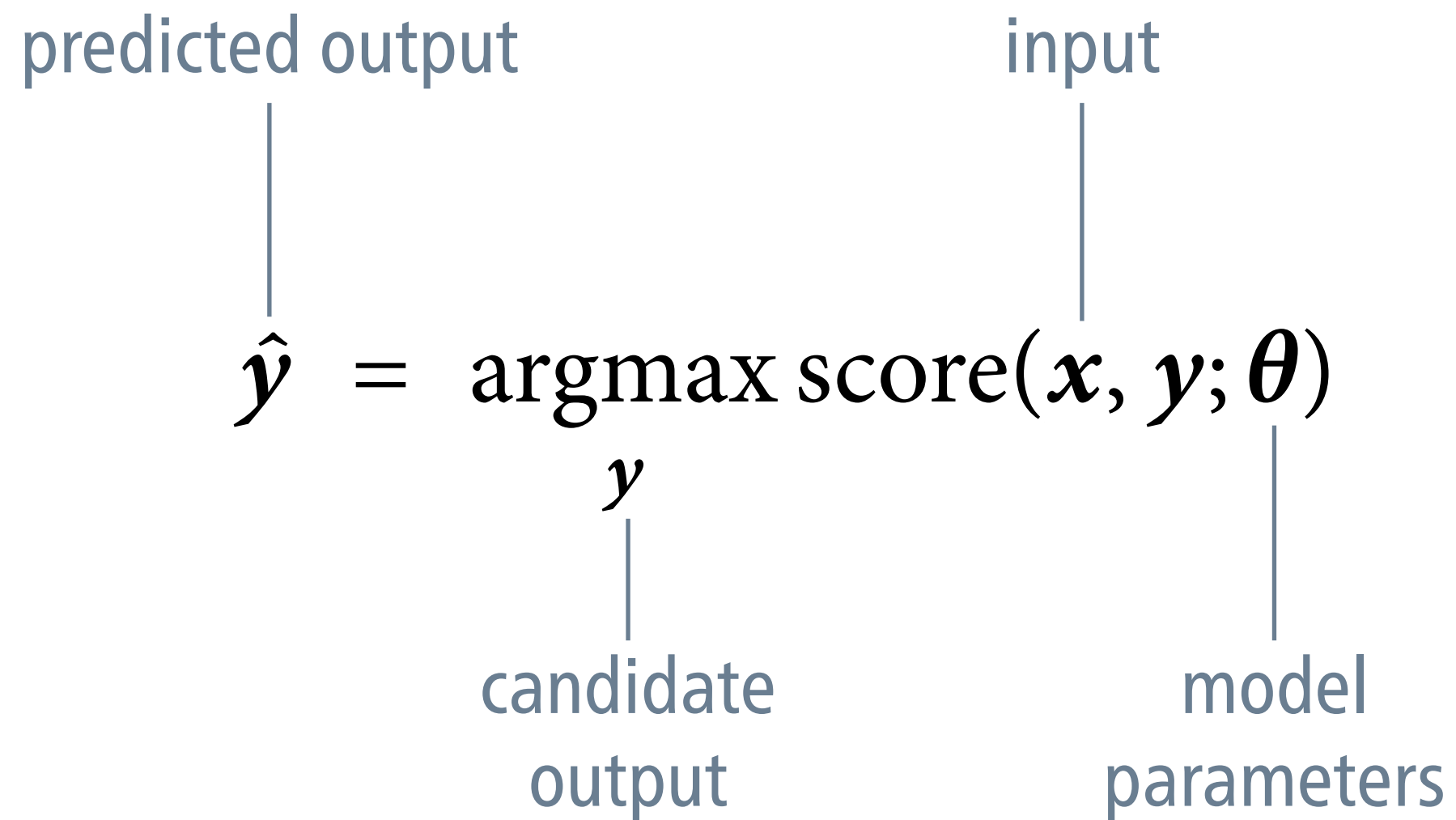
classifier

5

Machine translation and current research



Search and learning



Search and learning

Eisenstein (2019), § 1.2.2

- **Search module**

The search module is responsible for finding the candidate output y with the highest score relative to the input x .

requires efficient algorithms

- **Learning module**

The learning module is responsible for finding the model parameters θ that maximize the predictive performance.

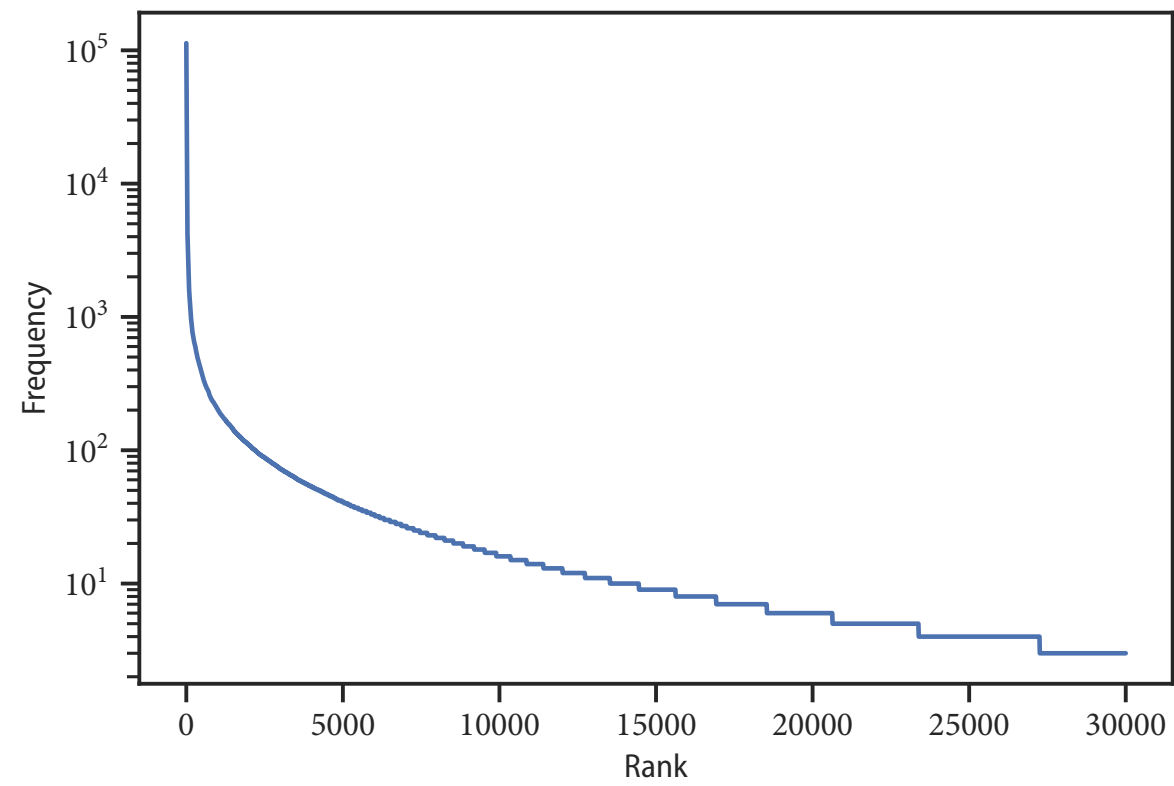
for example, using supervised machine learning

Language is special

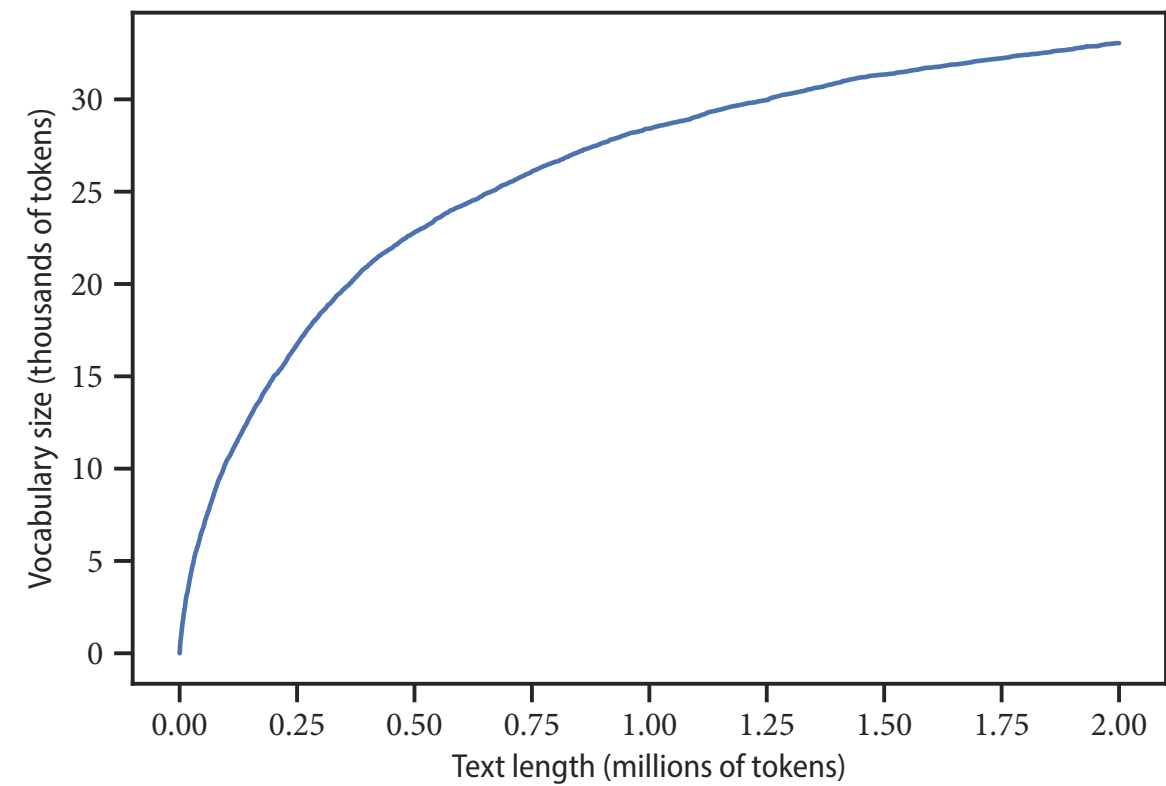
Eisenstein (2019), § 1.1

- 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