CMPT 413/713: Natural Language Processing

Introduction 2

Zhiyun (Richard) Peng

2026-09-09

Adapted from slides from Angel Chang, Anoop Sarkar, Danqi Chen and Karthik Narasimhan

Last lecture

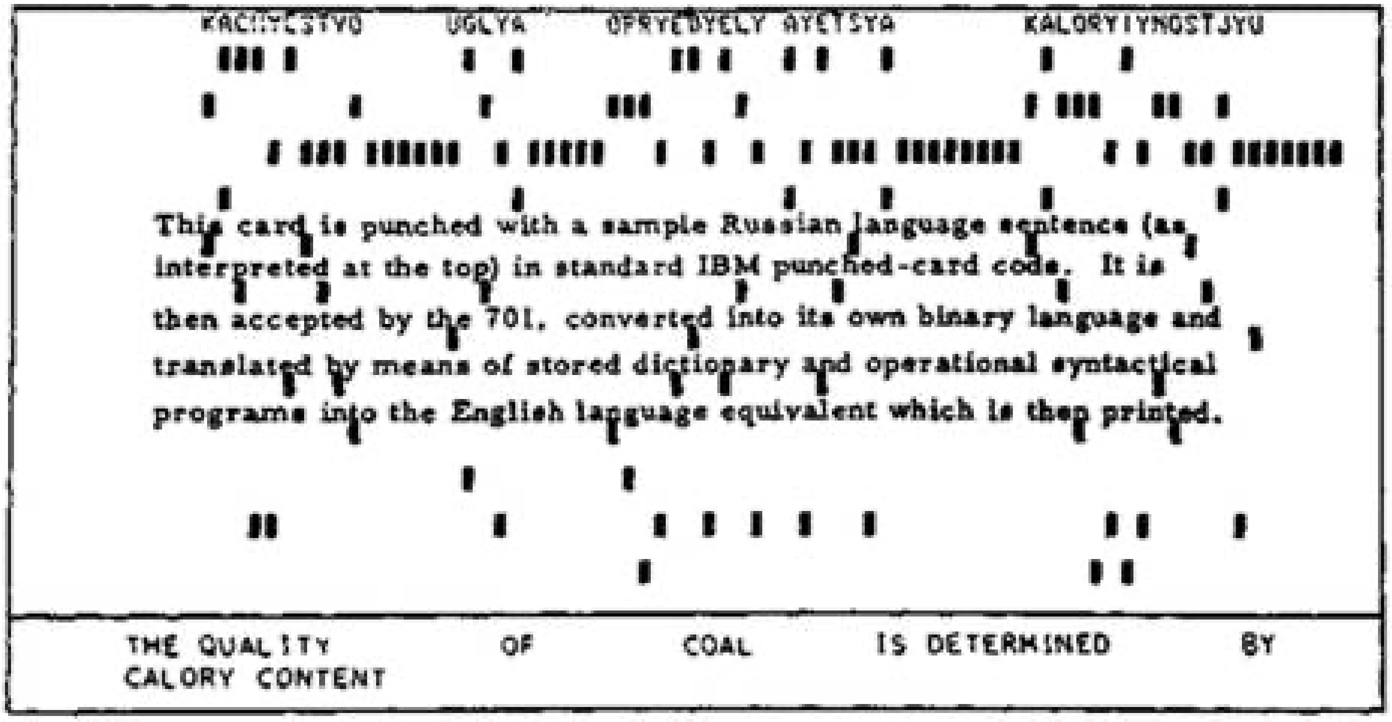
- Intro to NLP
  - What is language? NLP vs computational linguistics.
- Course Logistics
- Quiz 30 min.

Today

- Intro to NLP
  - What is language? NLP vs computational linguistics.
- Course Logistics
- Quiz 30 min.

Brief history of NLP

Beginnings



Specimen punched card and below a strip with translation, printed within a few seconds

Georgetown-IBM experiment, 1954

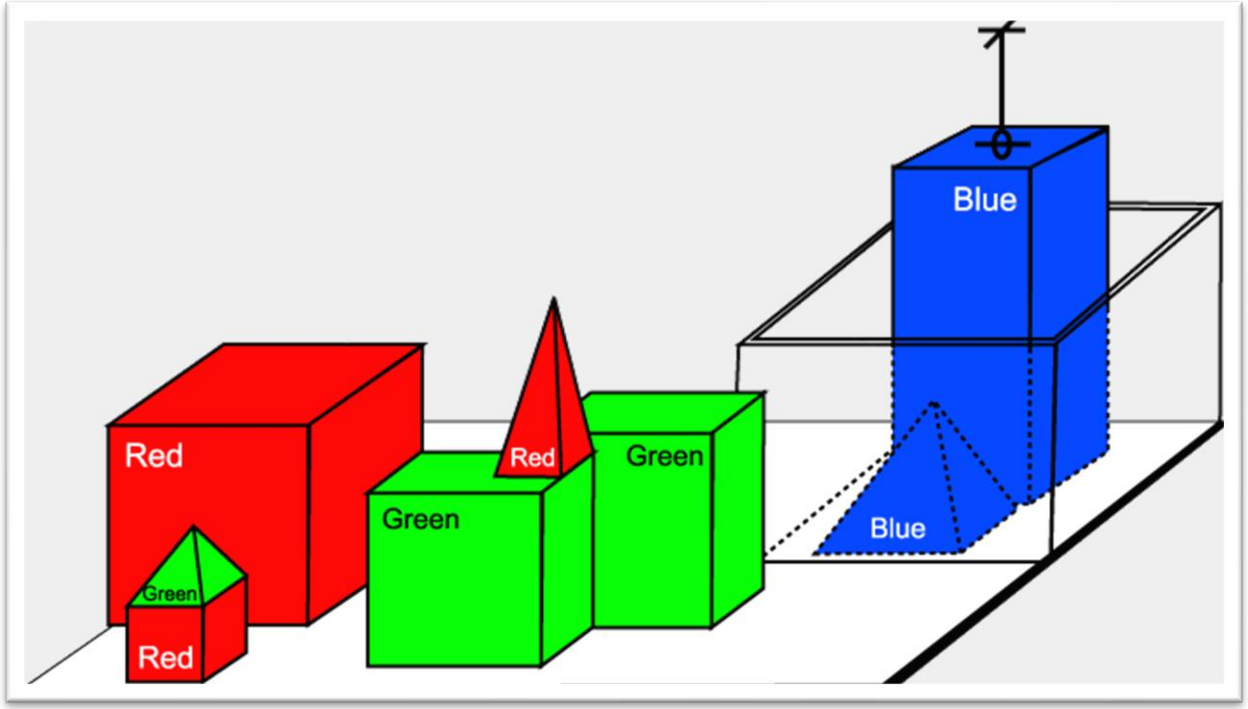
“Within three or five years, machine translation will be a solved problem”

SHRDLU (Winograd, 1968)

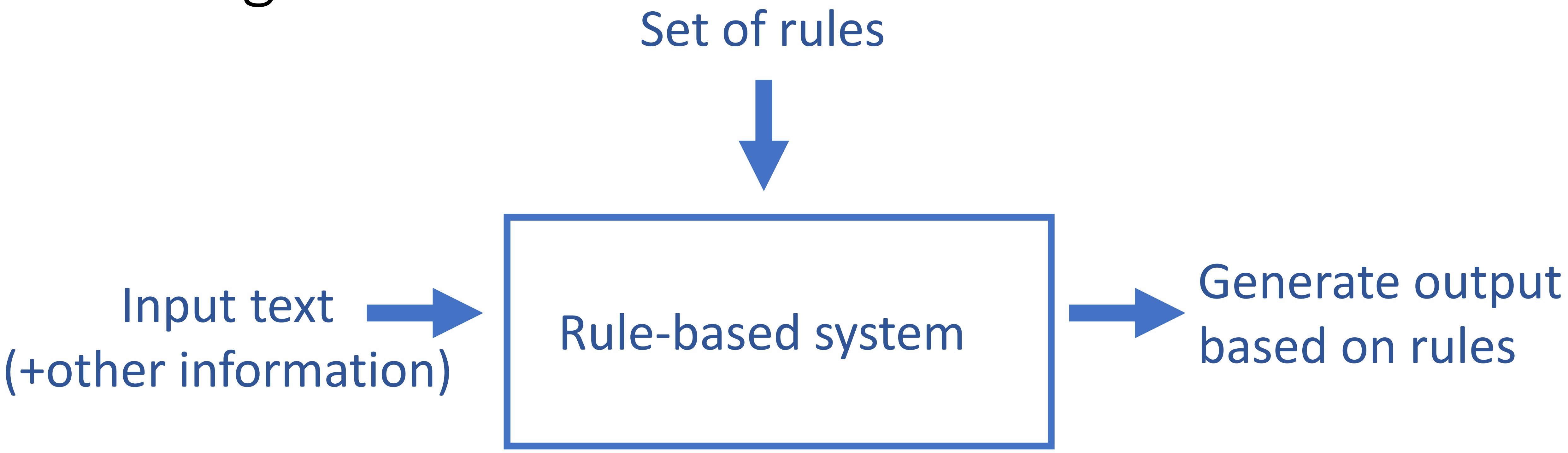
Video of actual system: https://www.youtube.com/watch?v=bo4RvYJYOzI

Person: Pick up a big red block.
Computer: OK.
Person: Grasp the pyramid.
Computer: I don't understand which pyramid you mean.
…

Lots of rules!



Using rules for NLP



Match a pattern

- Regular expressions

    .*ed => past tense verb

Contain keyword: delicious

Why is natural language processing hard?

Programming Languages
C, C++, Java, Python, …

- Unambiguous
- Fixed
- Designed
- Learnable?
- Known simple semantics

Natural Languages
French, English, Korean, Chinese, Tagalog, …

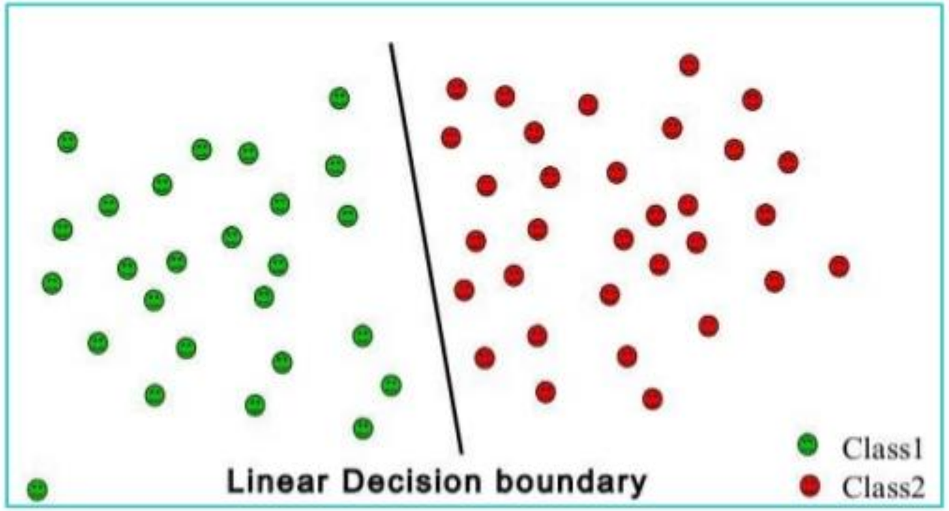
- Ambiguous
- Evolving
- Transmitted
- Learnable
- Complex semantics

Coming up rules is hard!

Let’s learn from data!

Rise of statistical learning

(late) 1980s to 2000s: Statistical MT, 2000s-2014: Statistical Phrase-Based MT



- Use of machine learning techniques in NLP
- Increase in computational capabilities
- Availability of electronic corpora
- Data-driven evaluation of models

Promise of deep learning

- Most NLP works in the past focused on human-designed representations and input features

Var	Definition	Value in Fig. 5.2
x_1	count(positive lexicon) $\in$ doc)	3
x_2	count(negative lexicon) $\in$ doc)	2
x_3	$\begin{cases} 1 & \text{if ``no''} \in \text{doc} \\ 0 & \text{otherwise} \end{cases}$	1
x_4	count(1st and 2nd pronouns $\in$ doc)	3
x_5	$\begin{cases} 1 & \text{if ``!''} \in \text{doc} \\ 0 & \text{otherwise} \end{cases}$	0
x_6	log(word count of doc)	$\ln(64) = 4.15$

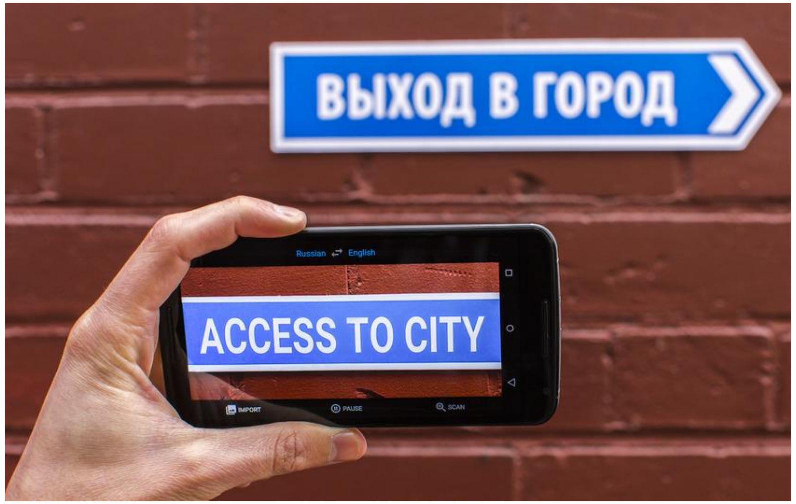
- **Representation learning attempts to automatically learn good features and representations**
- **Deep learning attempts to learn multiple levels of representation on increasing complexity/abstraction**

Credits: Greg Durrett

Deep learning era

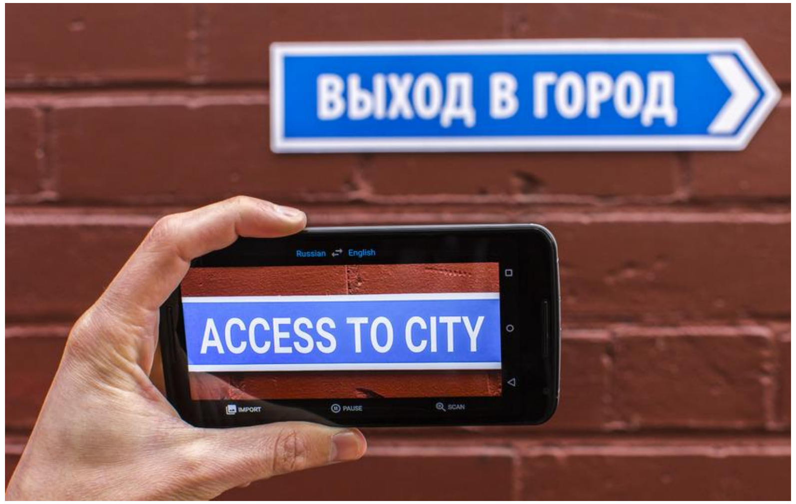
2014-Present: Neural Machine Translation

- Led to significant advances in core NLP technologies



Deep learning era

- Significant advances in core NLP technologies
- Essential ingredient: large-scale supervision, lots of compute
- Reduced manual effort - less/zero feature engineering



36M sentence pairs

Russian: Машинный перевод - это круто!

English: Machine translation is cool!

RESEARCH › PUBLICATIONS ›

Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation

https://ai.googleblog.com/2016/09/a-neural-network-for-machine.html

Side-by-side:
"Double-blinded",
Rate 0 to 6.

Table 10: Mean of side-by-side scores on production data

	PBMT	GNMT	Human	Relative Improvement
English → Spanish	4.885	5.428	5.504	87%
English → French	4.932	5.295	5.496	64%
English → Chinese	4.035	4.594	4.987	58%
Spanish → English	4.872	5.187	5.372	63%
French → English	5.046	5.343	5.404	83%
Chinese → English	3.694	4.263	4.636	60%

(Wu et al., 2016)

Power of Data

Meena (Google, 2020)

Conversation D

Human:	*I like how cute cows are*
Meena:	*Ah, interesting. You like how they are cute, right?*
Human:	*yeah, and they're realy smart*
Meena:	*I heard they go to college*
Human:	*cows go to college?*
Meena:	*I heard that a cow went to Harvard.*
Human:	*what did the cow study?*
Meena:	*Bovine sciences.*
Human:	*do horses go to Harvard?*
Meena:	*Horses go to Hayvard.*
Human:	*that's a pretty good joke, I feel like you led me into it*
Meena:	*You were trying to steer it elsewhere, I can see it.*

How it works:

- Corpus of conversational turns (over 40B words)
- Train huge neural network (2.6 billion parameters) for 30 days on 2048 TPUs cores
- Predict response given a sentence

https://ai.googleblog.com/2020/01/towards-conversational-agent-that-can.html

Power of Data

ChatGPT (OpenAI, 2022)

https://openai.com/blog/chatgpt/

write a sonnet about pokemon

A world of wonder, filled with creatures wild,
A realm of magic, where adventures start,
From Kanto to Kalos, far and wide,
The world of Pokémon, alive and heart.

With Pikachu as guide, we set out to roam,
To catch and train them all, to be the best,
Battling Gym Leaders, overcoming foes,
Our bond with creatures strong, a bond we'll never test.

But it's not just battles that define,
The world of Pokémon, so full of cheer,
It's friendships formed, and lessons learned,
A world that's full of love, and never fear.

So let's go on a journey, you and I,
Through the world of Pokémon, beneath the sky.

How it works:

- Trained on conversational data
- Uses reinforcement learning with human feedback

The Brilliance and Weirdness of ChatGPT

A new chatbot from OpenAI is inspiring awe, fear, stunts and attempts to circumvent its guardrails.

https://www.nytimes.com/2022/12/05/technology/chatgpt-ai-twitter.html

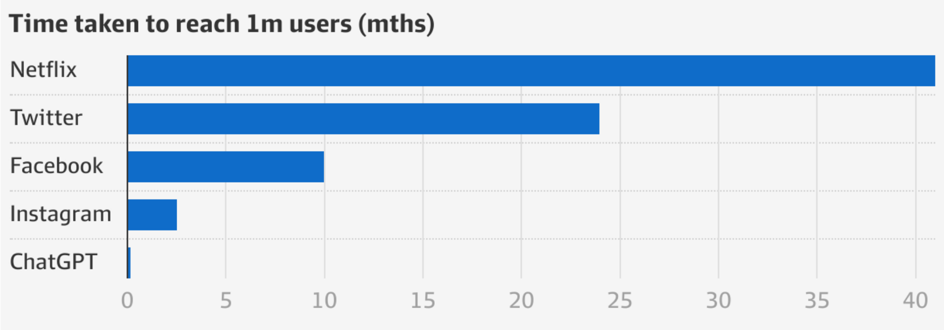
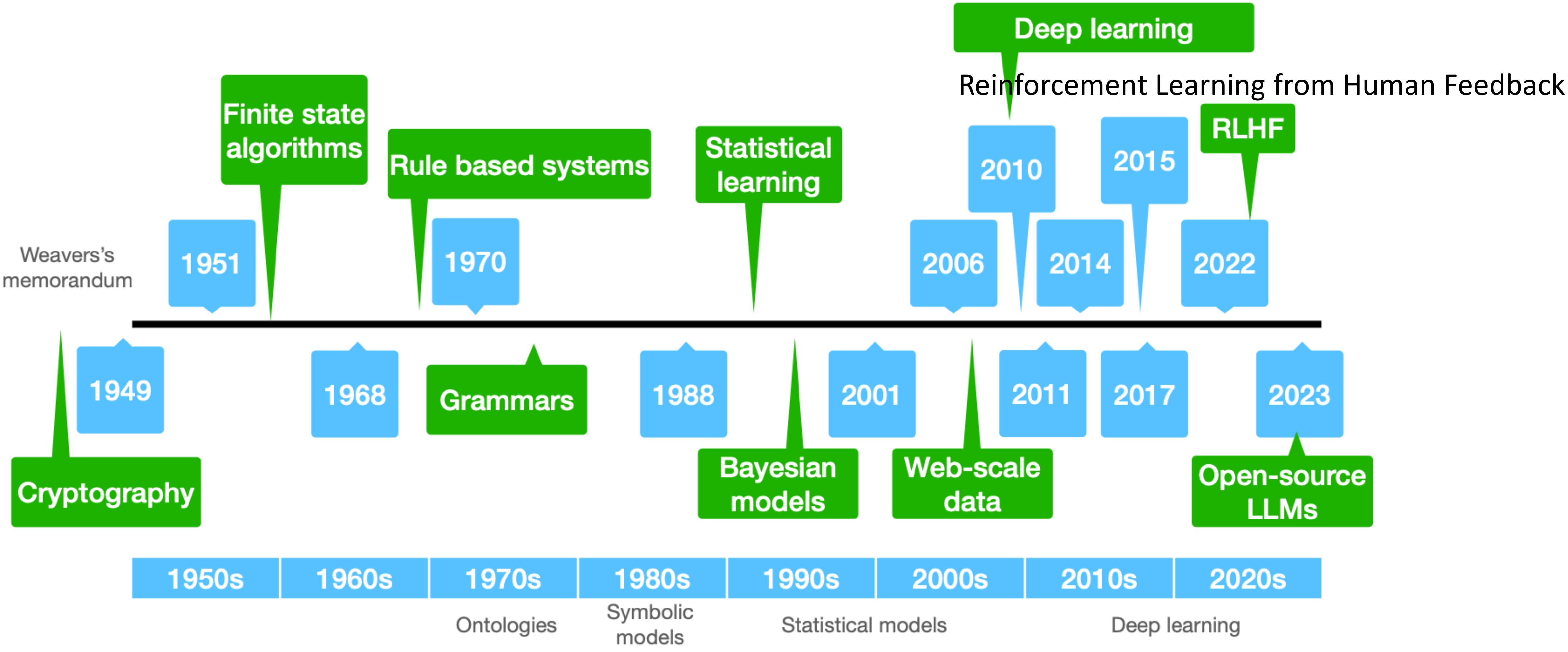


Chart: Financial Review • Source: Genevieve Roch-Decter, CFA

Key developments in NLP

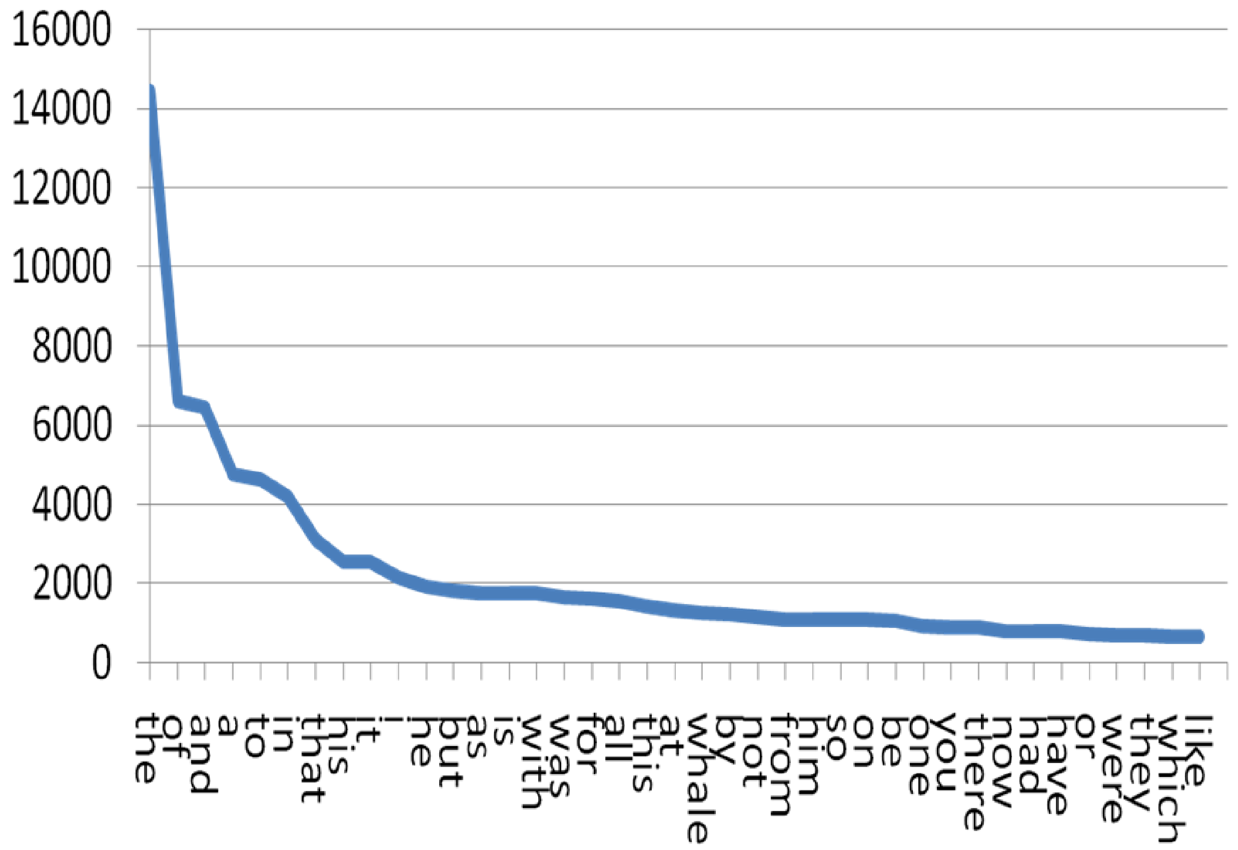


https://anoopsarkar.github.io/nlp-class/assets/slides/intro-history-of-nlp.pdf

Remaining challenges in NLP?

Challenges in modern NLP

- *Scale*: Large number of phenomena
- *Sparsity*: Text data is often heavy-tailed



- ***Solved by ingesting the Internet?***
- ***But what about low-resource languages? Specialized domains?***

What will this course cover?

What will this course cover?

Representations

- Embeddings
- Compositional
- Structured

Methods

- Language Models
- Statistical Methods
- Neural Models
- Dynamic programming

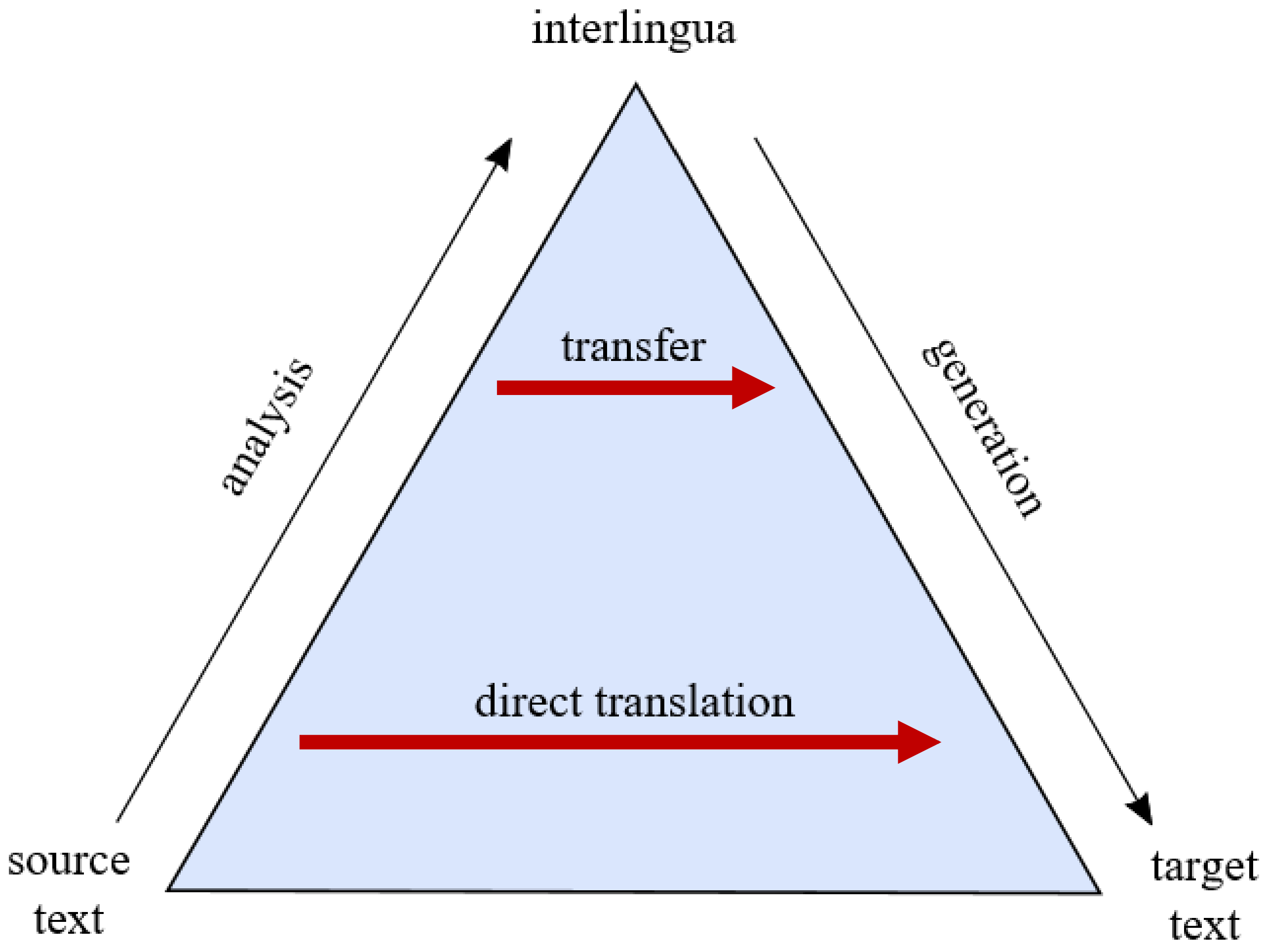
LLMs

- Pretrained models
- Post-training
- Prompting
- Efficient fine-tuning
- LLM agents

Applications and tasks

- Machine translation, text generation, question answering, grounding

Vauquois Triangle for translation



the cat sat on the table

η γάτα κάθισε στο τραπέζι

Key Questions

- How to represent language?
- How to compose meaning?
- How do we translate from the ``surface form'' of language to a representation that a computer can process, and back?
  - How do we encode / parse the meaning of a sentence?
  - How do we decode / generate text from a meaning representation?
- What are the key methods / algorithms for learning these representations?

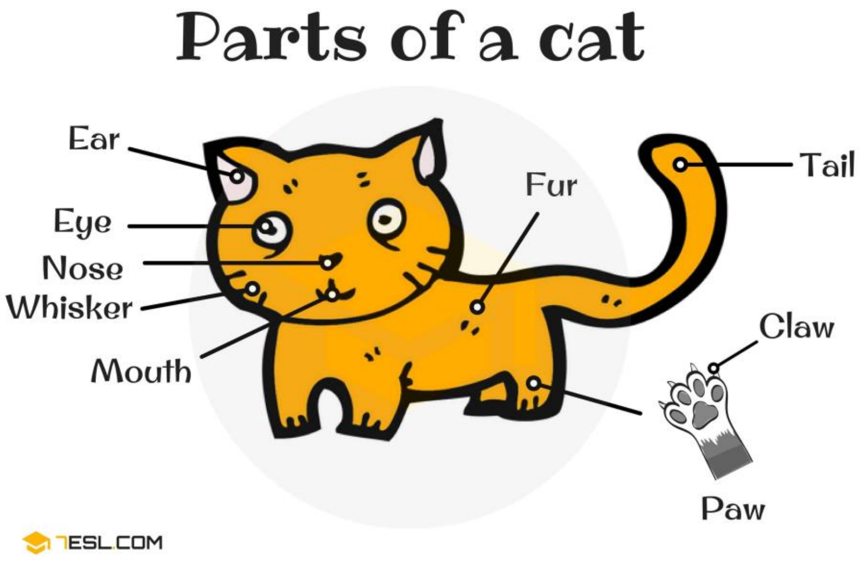
Representations

How to represent the meaning of something?

"cat"

cat: a small domesticated carnivore, *Felis domestica* or *F. catus,* bred in a number of varieties.

```
cat → {
  isMammal: true
  hasFur: true
  hasLegs: true
  meows: true
  barks: false
  height: 9.1 – 9.8 in
  weight: 7.9 – 9.9 lbs
  …
}
```



Attributed representation

Representations

“cat”

“dog”

Representing meaning as vectors

- common representation space
- enables information sharing
- can be learned from data

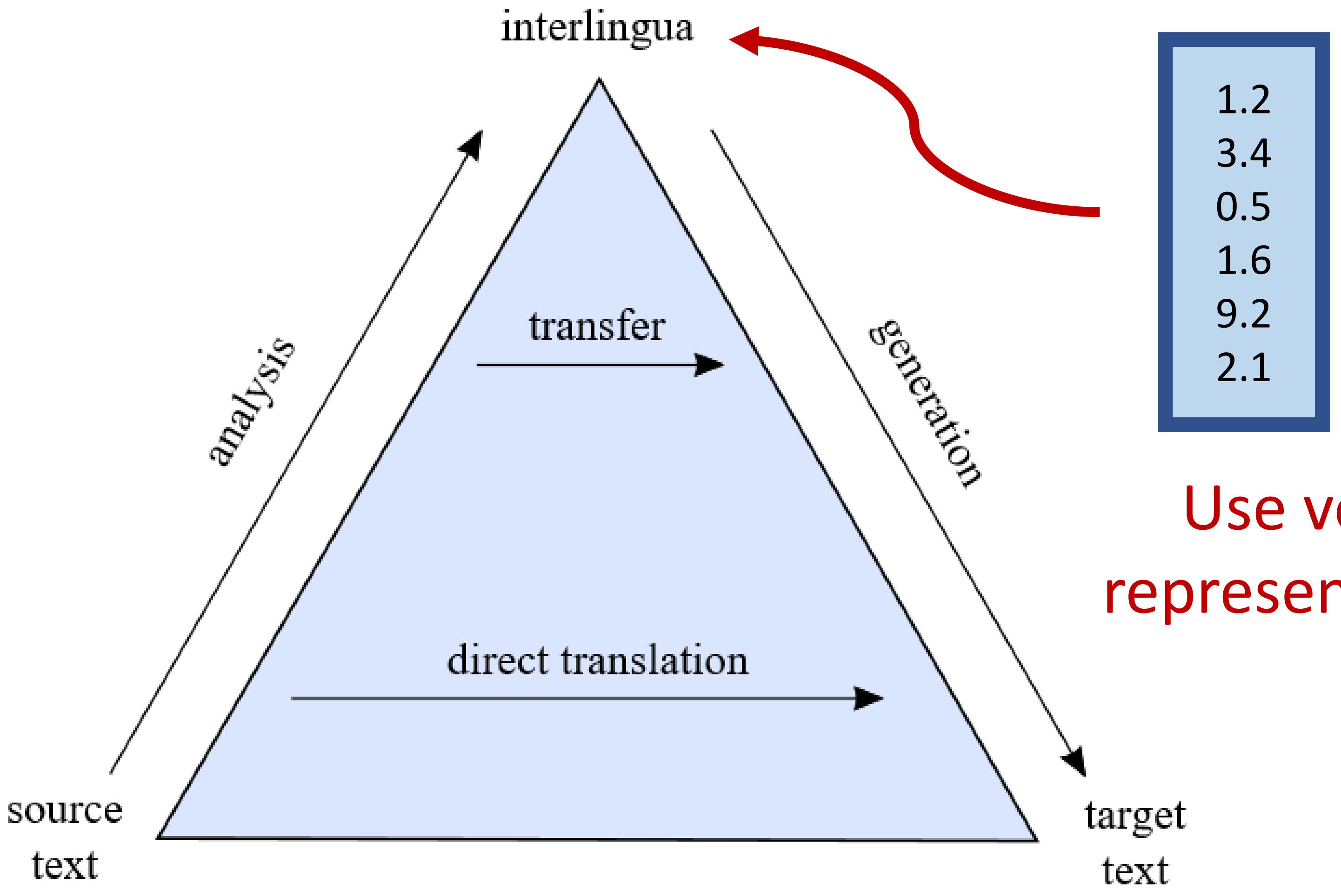
- One-hot

  cat = [0 0 0 1 0 0 0]
  dog = [0 0 0 0 0 1 0]

- Embeddings

  cat = [0.04 1.79 -1.79 1.07 0.48]
  dog = [0.61 1.84 -1.12 0.52 0.53]

Vauquois Triangle for translation

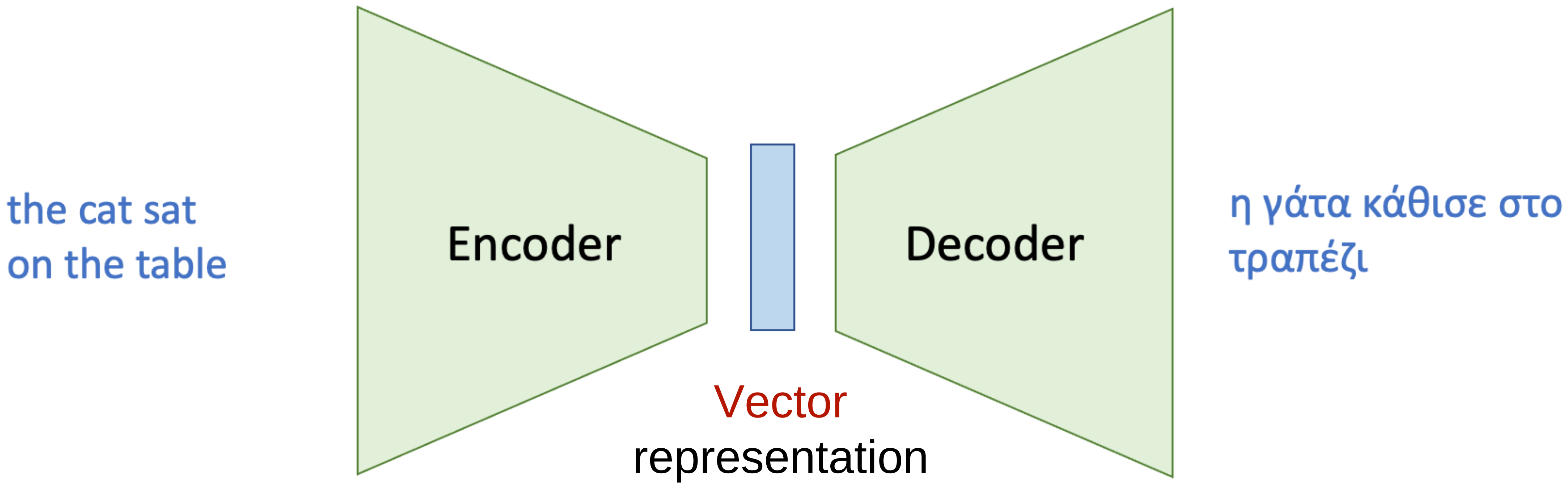


Use vector to represent meaning

the cat sat on the table

η γάτα κάθισε στο τραπέζι

Vector based representation



Common choice: flat unstructured vector of real numbers
- easy to work with
- well defined mathematical operations
- can be used to measure if two things are similar or not

How to learn representations for words?

“cat” “dog”

start with representing words as discrete tokens

- **One-hot**
  - cat = [0 0 0 1 0 0 0]
  - dog = [0 0 0 0 0 1 0]
- **Embeddings**
  - cat = [0.04 1.79 -1.79 1.07 0.48]
  - dog = [0.61 1.84 -1.12 0.52 0.53]

learn about neural models that use vectors

use context to represent words

learn a dense vector representation that can be used to predict words from their context

- Predict what goes next

The cat sat on the _____

- Fill in the blank

A bottle of ______ is on the table.

How to compose language?

Sequence of words

(or characters)

Composition

function

1.2
3.4
0.5
1.6
9.2
2.1

Bag of Words (BOW)

Window BoW (bag of ngrams)

RNN

Transformers

CNN

GraphNN

Neural networks!

How to compose (aggregate) language?

Sequence of words

(or characters)

Composition
(aggregation)
function

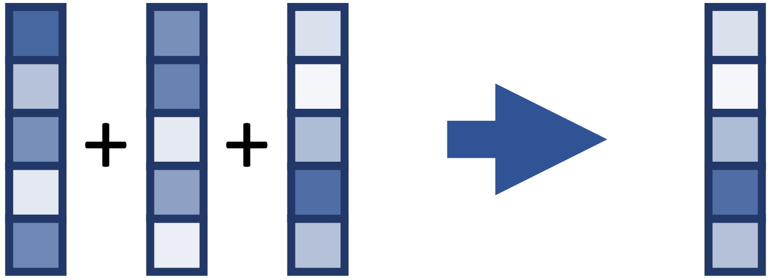
1.2
3.4
0.5
1.6
9.2
2.1

Simplest: take weighted average

Example:

Movie was terrible

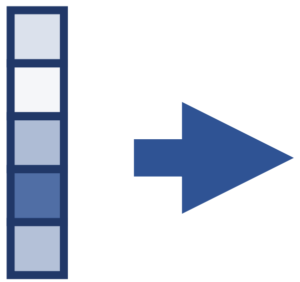
Sentence
representation



Use in downstream task

How to use the learned representation?

What do we do with this vector representation after we get it?



Use in downstream task

Vector representation

Do prediction: classify as spam/not spam

Use to generate text (also known as **decoding**)

Can also be used to generate other stuff (like images!)