NLP with Neural Networks: Introduction

THE UNIVERSITY OF UTAH

Words are a very fantastical banquet, just so many strange dishes

And yet, we seem to do fine

We can understand and generate language effortlessly.

Almost

Wouldn’t it be great if computers could understand language?

Wanted

Programs that can learn to understand and reason about the world via language

Processing Natural Language

Or:
An attempt to replicate (in computers) a phenomenon that is exhibited *only* by humans.

https://flic.kr/p/6fnqdv

Today: Why study neural networks for NLP

What makes language different from other applications?

Why neural networks?

Language is fun!

Language is *ambiguous*

Tuna Recall Blamed on Seal

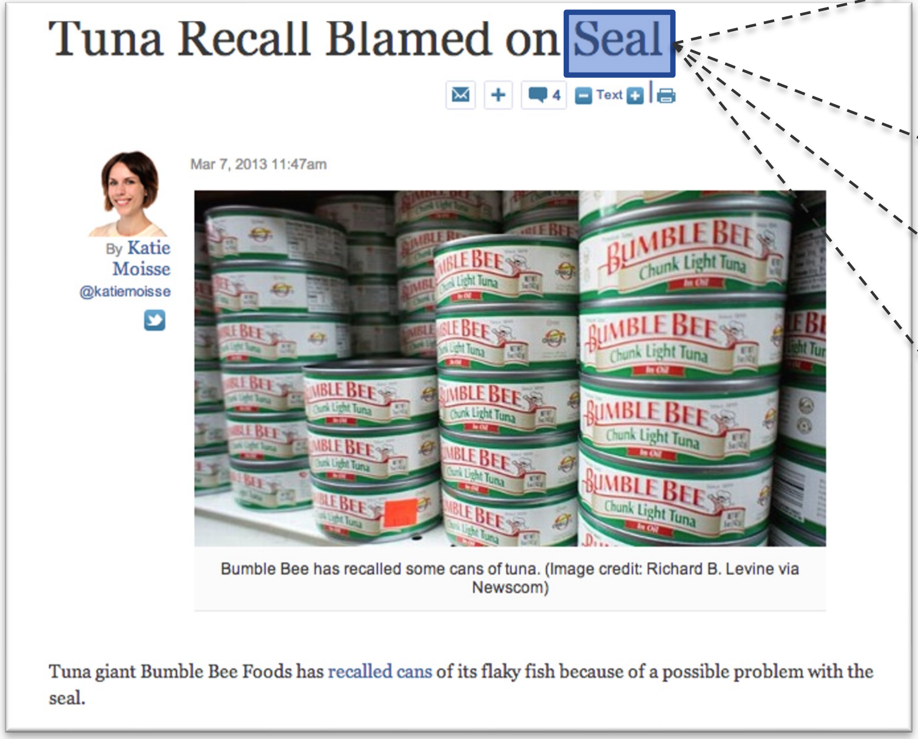
By Katie Moisse
@katiemoisse

Mar 7, 2013 11:47am

Bumble Bee has recalled some cans of tuna. (Image credit: Richard B. Levine via Newscom)

Tuna giant Bumble Bee Foods has recalled cans of its flaky fish because of a possible problem with the seal.

Language is *ambiguous*



Tuna Recall Blamed on Seal

Text

Mar 7, 2013 11:47am

By Katie Moisse
@katiemoisse

Bumble Bee has recalled some cans of tuna. (Image credit: Richard B. Levine via Newscom)

Tuna giant Bumble Bee Foods has recalled cans of its flaky fish because of a possible problem with the seal.

Language is *ambiguous*

I ate sushi with tuna.

I ate sushi with chopsticks.

I ate sushi with a friend.

I saw a man with a telescope.

Stolen painting found by tree.

Ambiguity can take many forms: Lexical, syntactic, semantic

Language has complex structure

Mary saw a ring through the window and asked John for ***it***. Why on earth did Mary ask for a window?

“My parents are stuck at Waterloo Station. There’s been a bomb scare.”
“Are **they** safe?”
“No, bombs are really dangerous.

Anaphora resolution: Which entity/entities do pronouns refer to?

Language has complex structure

Jan saw the children swim

Parsing: Identifying the syntactic structure of sentences

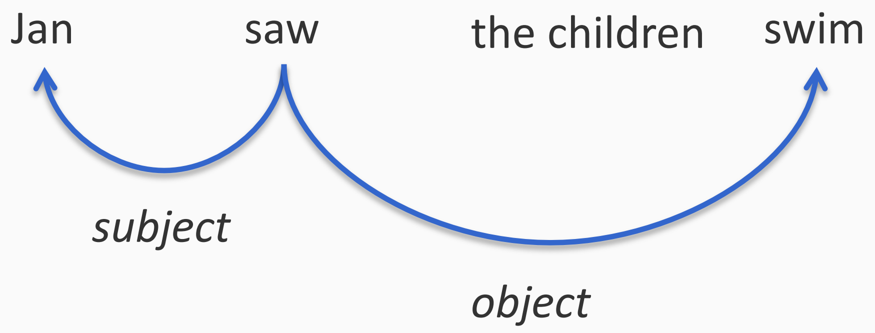
Language has complex structure

Jan saw the children swim

subject

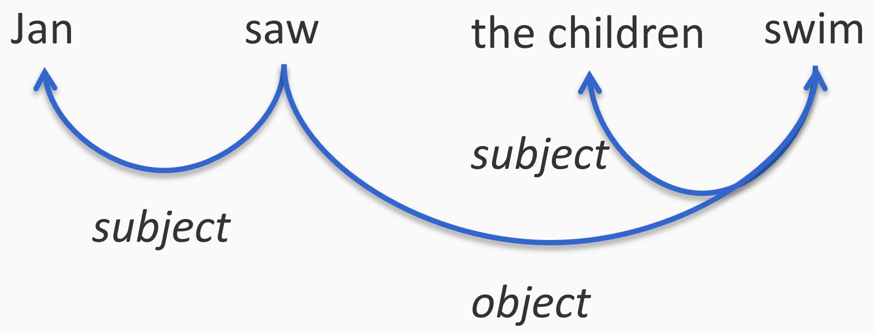
Parsing: Identifying the syntactic structure of sentences

Language has complex structure



Parsing: Identifying the syntactic structure of sentences

Language has complex structure



Parsing: Identifying the syntactic structure of sentences

Language has complex structure

Jan	de kinderen	zag	zwemmen
Jan	*the children*	*saw*	*swim*

Language has complex structure

Jan	de kinderen	zag	zwemmen
Jan	*the children*	*saw*	*swim*

Jan	saw	the children	swim

Language has complex structure

Jan	de kinderen	zag	zwemmen
Jan	*the children*	*saw*	*swim*

Jan saw the children swim

Language has complex structure

Jan	de kinderen	zag	zwemmen
Jan	*the children*	*saw*	*swim*

Jan saw the children swim

Language has complex structure

Jan	de kinderen	zag	zwemmen
Jan	*the children*	*saw*	*swim*

Jan saw the children swim

Language has complex structure

Jan	Piet	de kinderen	zag	helpen	zwemmen
Jan	*Piet*	*the children*	*saw*	*help*	*swim*

Language has complex structure

Jan	Piet	de kinderen	zag	helpen	zwemmen
Jan	*Piet*	*the children*	*saw*	*help*	*swim*
Jan	saw	Piet	help	the children	swim

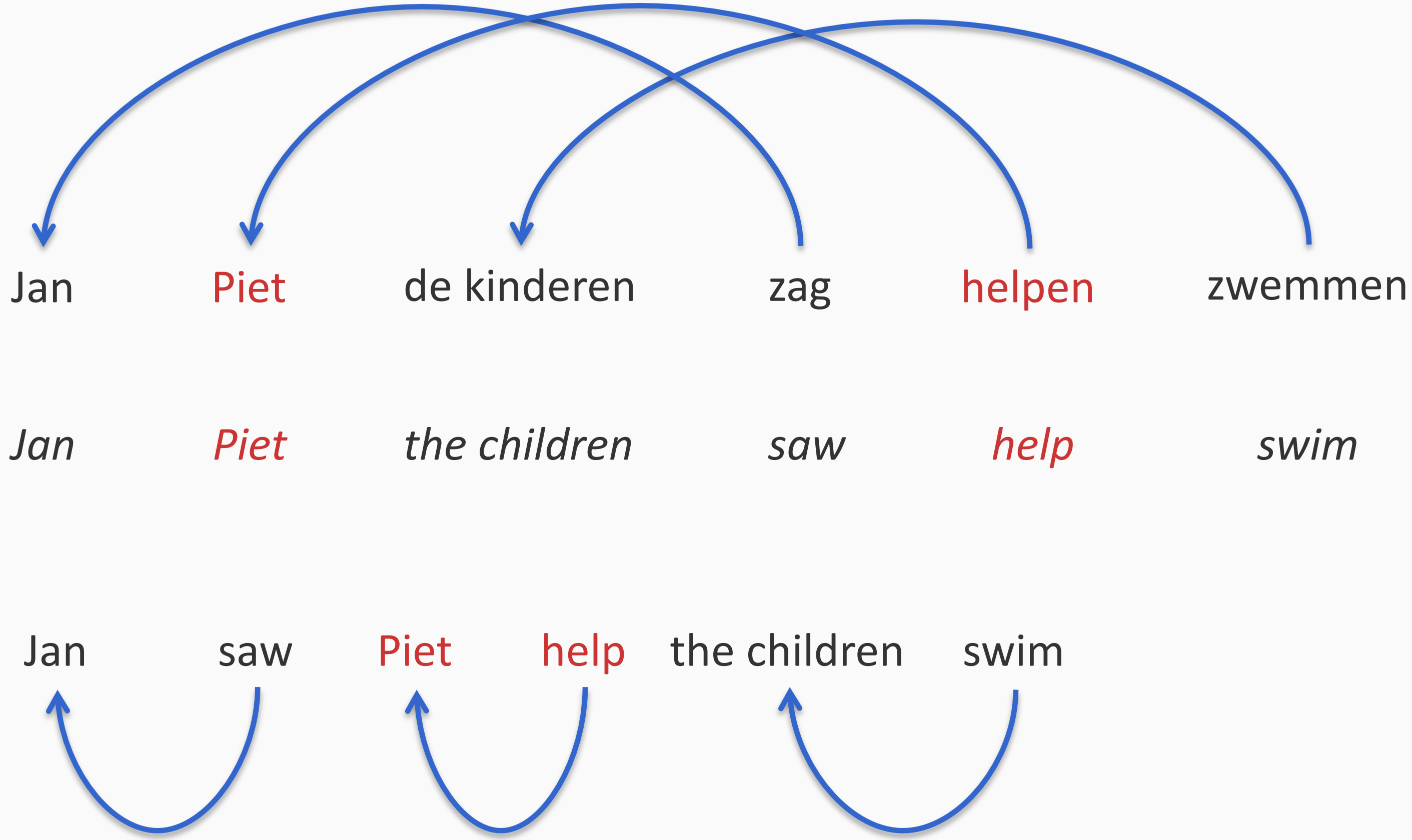
Language has complex structure

Jan Piet de kinderen zag helpen zwemmen

Jan Piet the children saw help swim

Jan saw Piet help the children swim

Language has complex structure



Language has complex structure

Jan	Piet	Marie	de kinderen	zag	helpen	leren	zwemmen
Jan	*Piet*	*Marie*	*the children*	*saw*	*help*	*teach*	*swim*

Language has complex structure

Jan Piet Marie de kinderen zag helpen leren zwemmen

Jan Piet Marie the children saw help teach swim

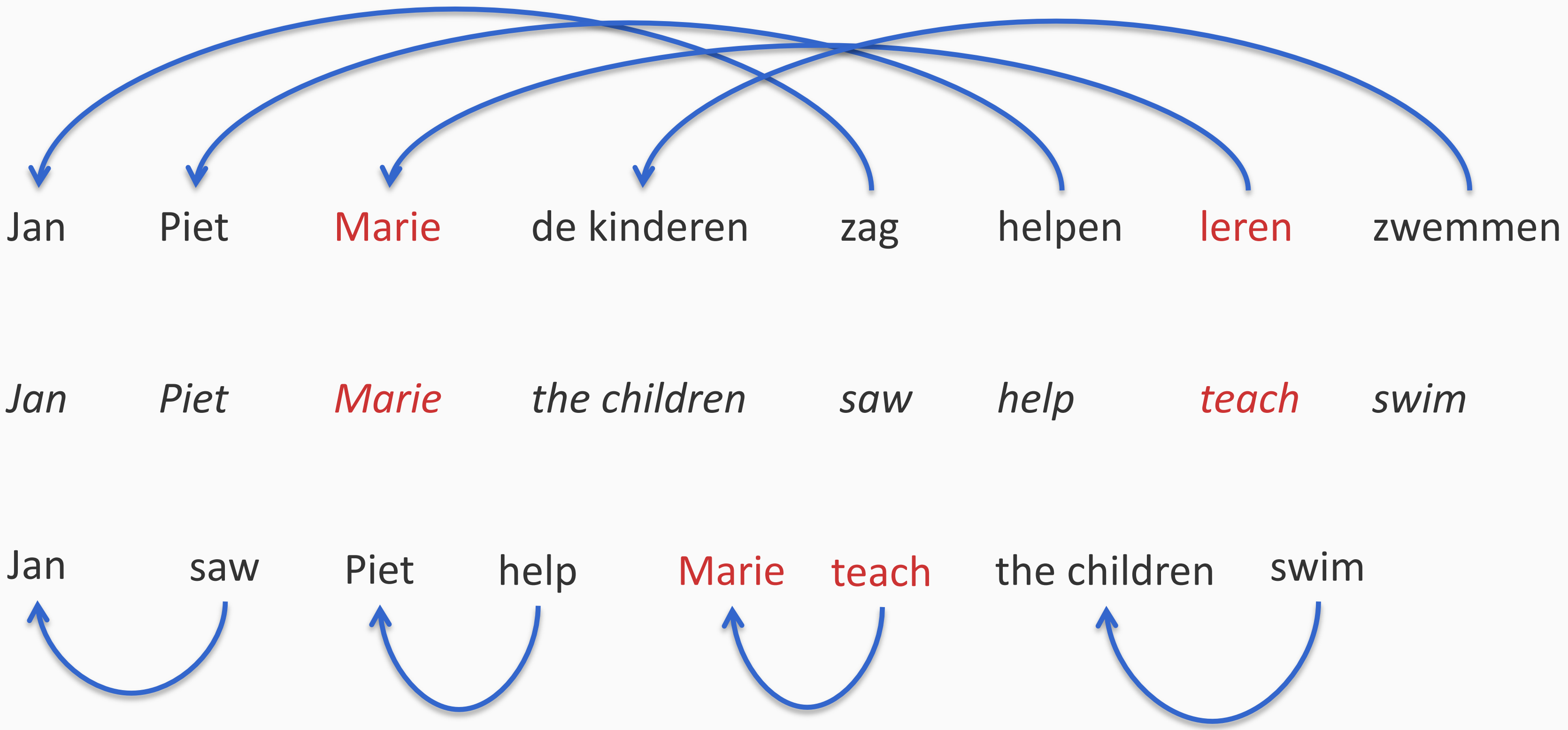
Jan saw Piet help Marie teach the children swim

Language has complex structure

Jan	Piet	Marie	de kinderen	zag	helpen	leren	zwemmen
Jan	*Piet*	*Marie*	*the children*	*saw*	*help*	*teach*	*swim*

Jan saw Piet help Marie teach the children swim

Language has complex structure



Language has complex structure

Natural language is not a context free language!

Jan Piet Marie de kinderen zag helpen leren zwemmen

Jan Piet Marie the children saw help teach swim

Jan saw Piet help Marie teach the children swim

Many, many linguistic phenomena

Metaphor

- *makes my blood boil, apple of my eye, etc.*

Metonymy

- *The White House said today that …*

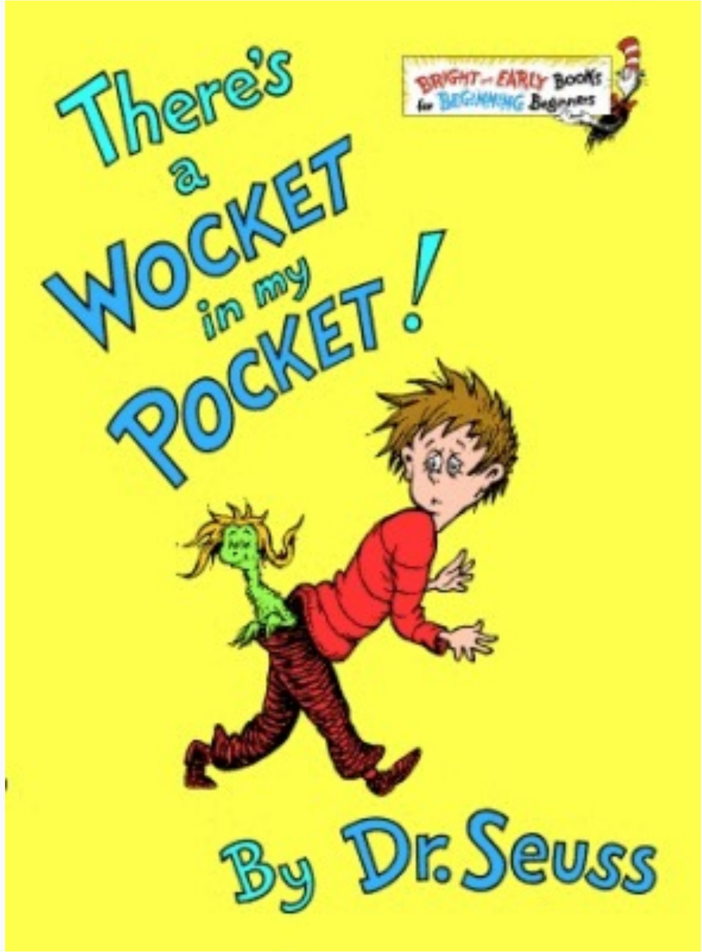
A very long list…

And, we make up words all the time

"If not actually disgruntled, he was far from being gruntled."

"The colors … only seem really real when you viddy them on the screen."

"Twas brillig, and the slithy toves
Did gyre and gimble in the wabe:
All mimsy were the borogoves,
And the mome raths outgrabe."



Language can be problematic to understand

Ambiguity and variability

Language is ambiguous and can have variable meaning

But machine learning methods can excel in these situations

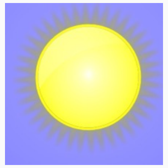
There are other issues that present difficulties:

1. Inputs are discrete, but numerous (words)
2. Both inputs and outputs are compositional

1. Inputs are discrete

What do words mean?
How do we represent meaning in a computationally convenient way?

bunny and *sunny* are only one letter apart but very far in meaning



bunny and *rabbit* are very close in meaning, but look very different

And can we *learn* their meaning from data?

2. Compositionality: Piecing meaning together from parts

Inputs to programs are compositional

– characters form words, which form phrases, clauses, sentences, and entire documents

Outputs are also compositional

– Several NLP tasks produce structures
    - Outputs are trees or graphs (e.g., parse trees)

– Or they produce language
    - E.g., translation, generation, summarization

Discrete + compositional = sparse

Compositionality allows infinite productivity from finite means

- Think of linguistic creativity
- How many words, phrases, sentences have you encountered that you have never seen before? How did you know what they meant?

No dataset has all inputs/outputs possible

NLP has to generalize to novel inputs and also generate novel outputs

Machine learning to the rescue

Modeling language: Power to the data

Understanding and generating language are challenging computational problems

Supervised machine learning offers perhaps the best known methods

- Essentially teases apart patterns from labeled data

Example: The company words keep

I would like to eat a ________ of cake

peace or *piece*?

An idea

- Train a *binary classifier* to make this decision
- Use indicators for neighboring words as features

Works surprisingly well!

Data + features + learning algorithm = Profit!

Example: The company words keep

I would like to eat a ________ of cake

peace or *piece*?

An idea

- Train a *binary classifier* to make this decision
- Use indicators for neighboring words as features

Works surprisingly well!

Data + $\boxed{\text{features}}$ + learning algorithm = Profit!

What features?

The problem of representations

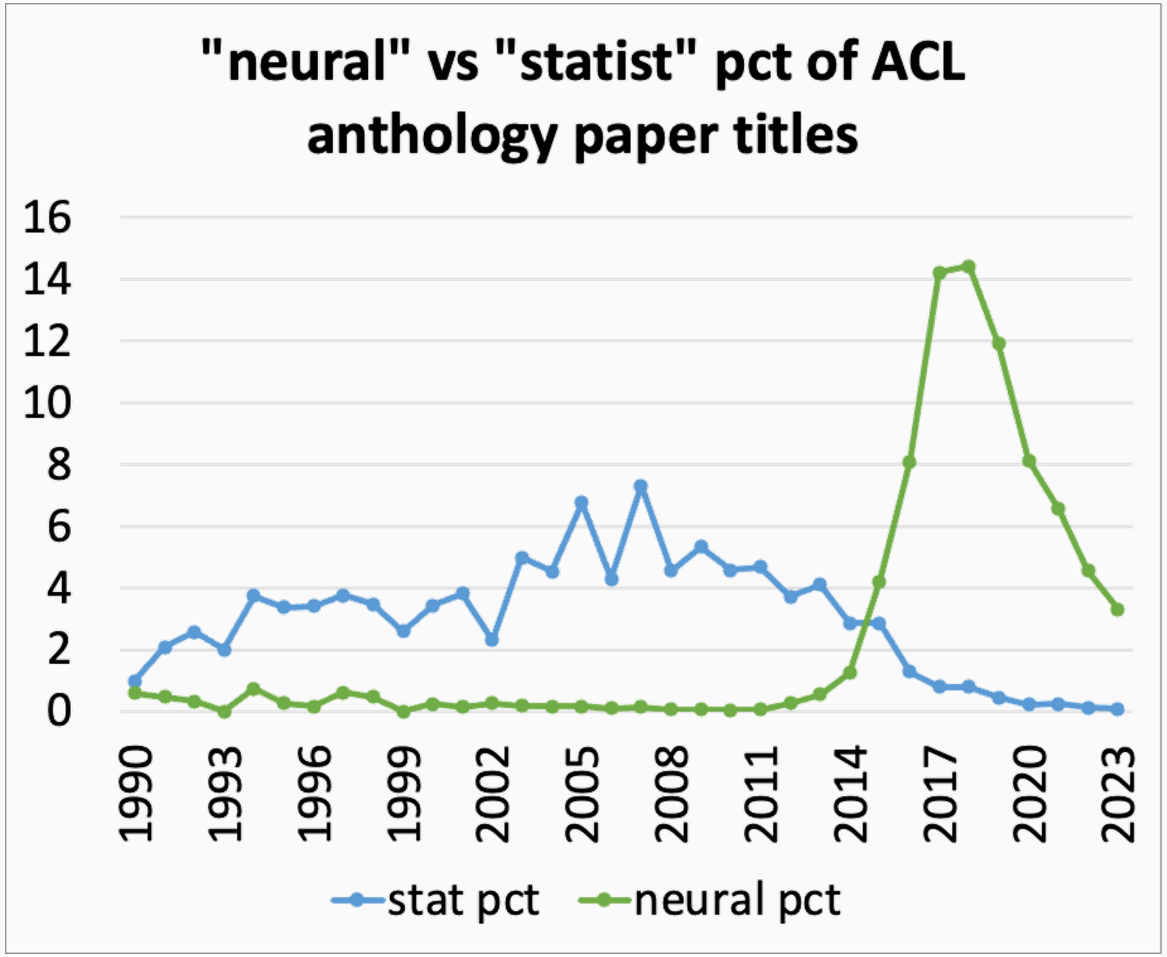
"Traditional NLP"

- Hand designed features: words, parts-of-speech, etc
- Linear models

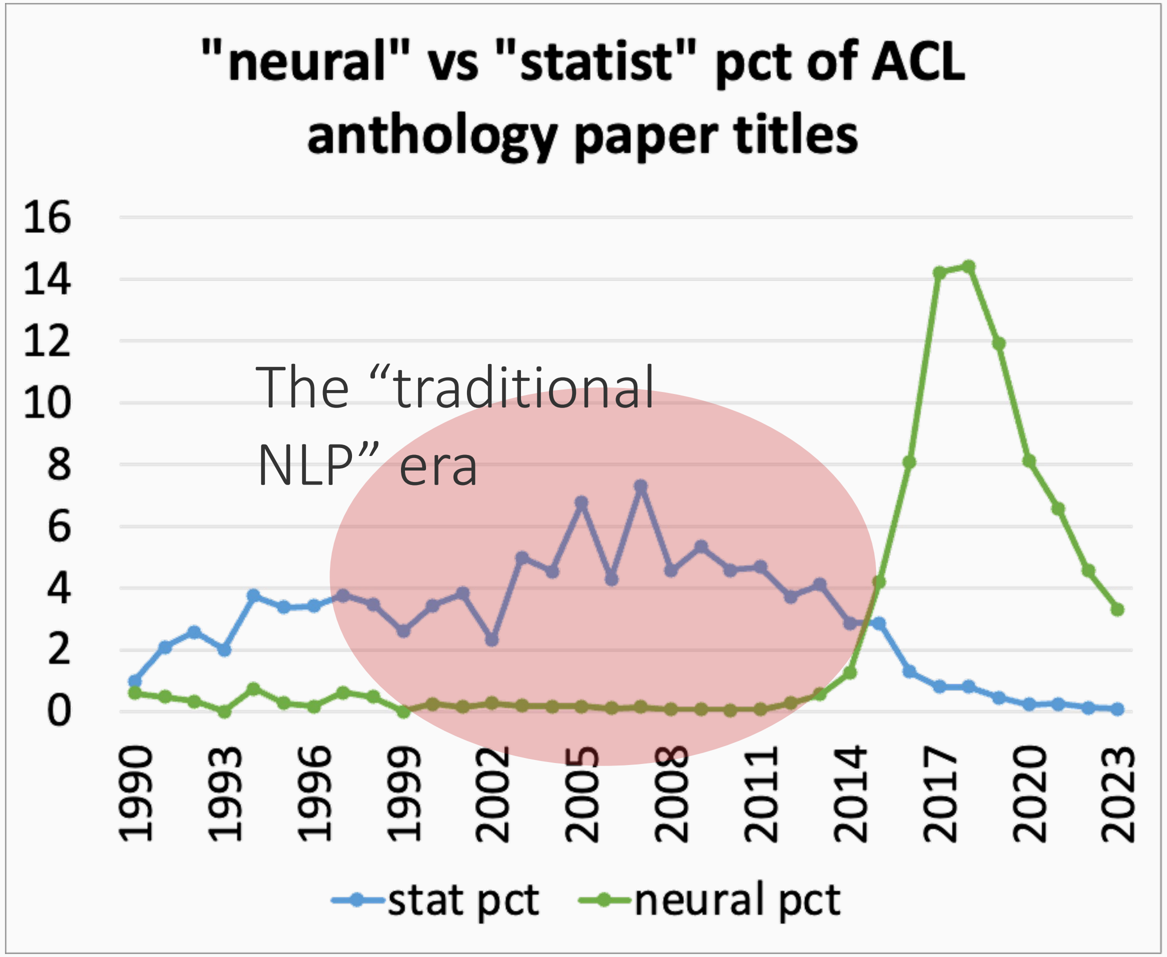
Manually designed features could be incomplete or overcomplete

Deep learning to learn representations

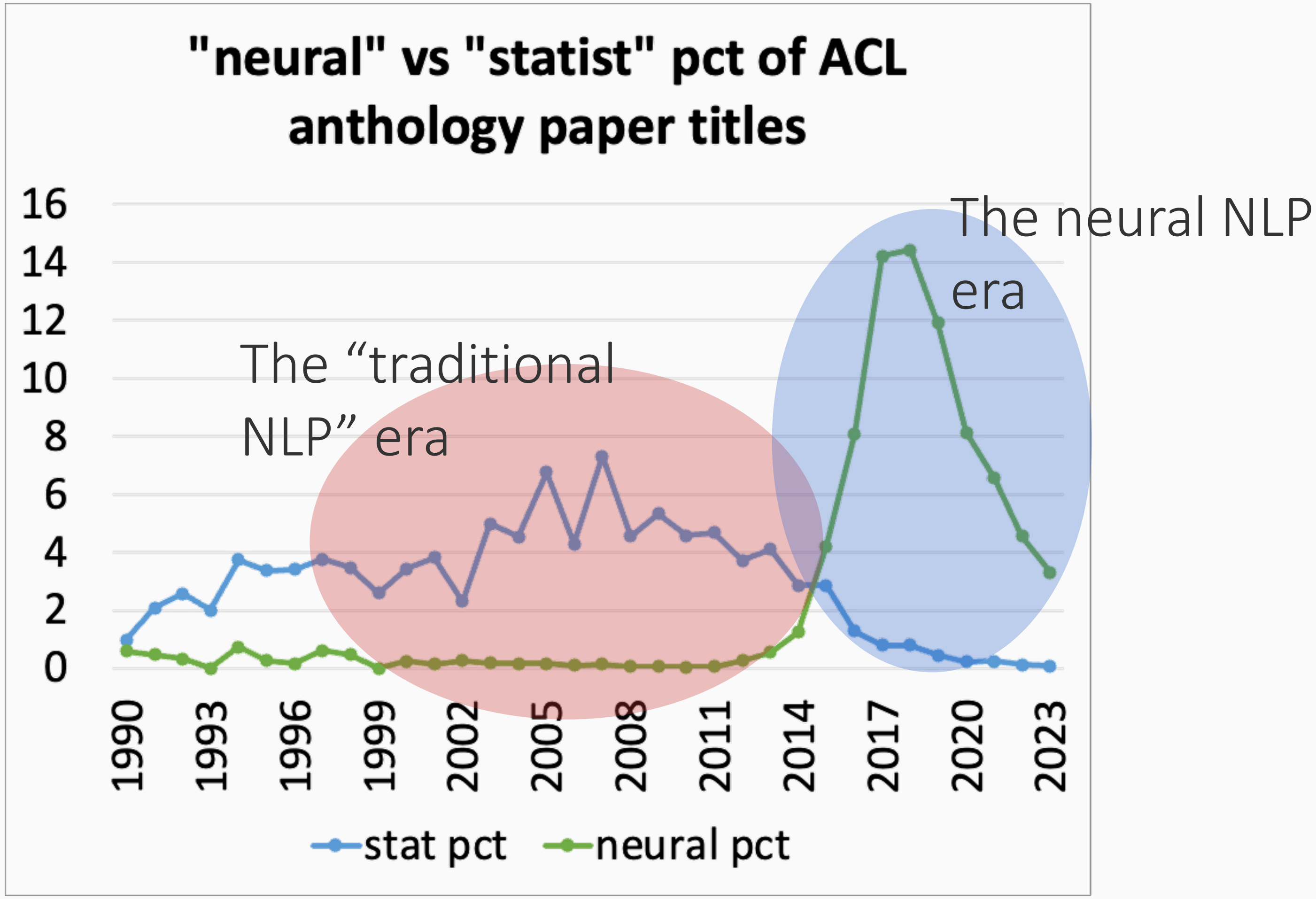
- Promises the ability to learn good representations (i.e. features)
- Typically vectors, also called distributed representations



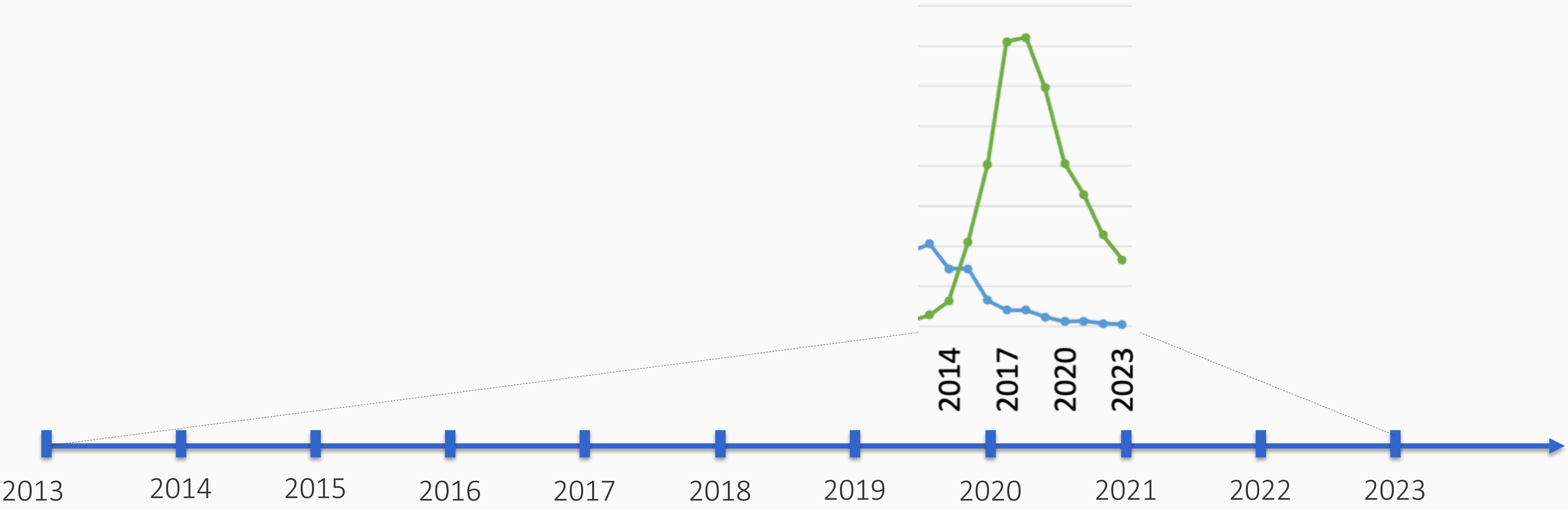
Plot courtesy Jon May



Plot courtesy Jon May



Plot courtesy Jon May


2014
2017
2020
2023
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023

Word2vec, Glove

Static word embeddings

2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023

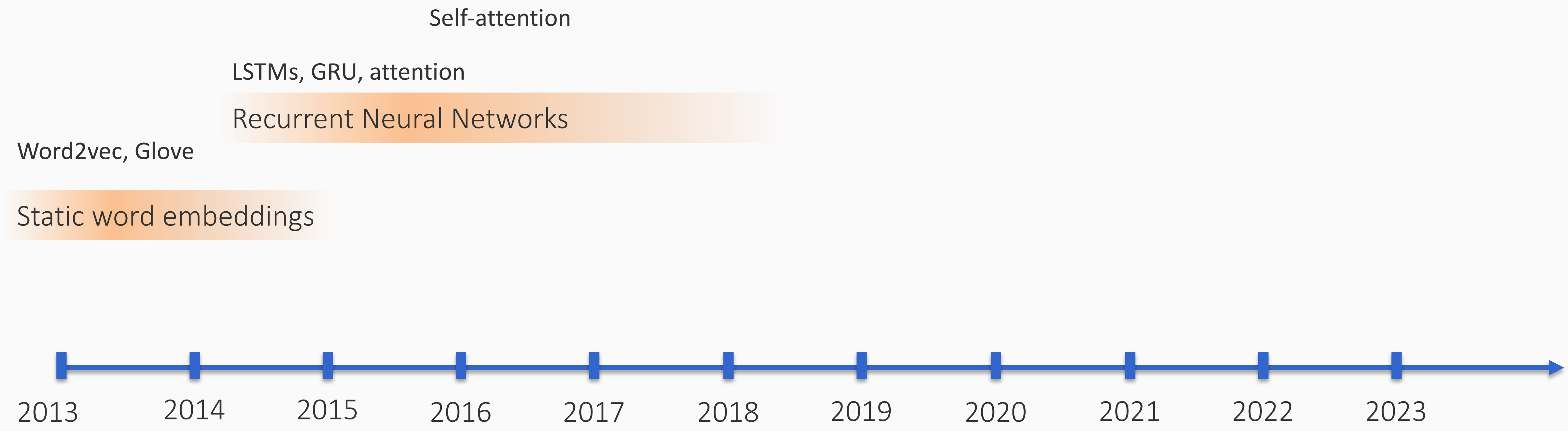
LSTMs, GRU, attention

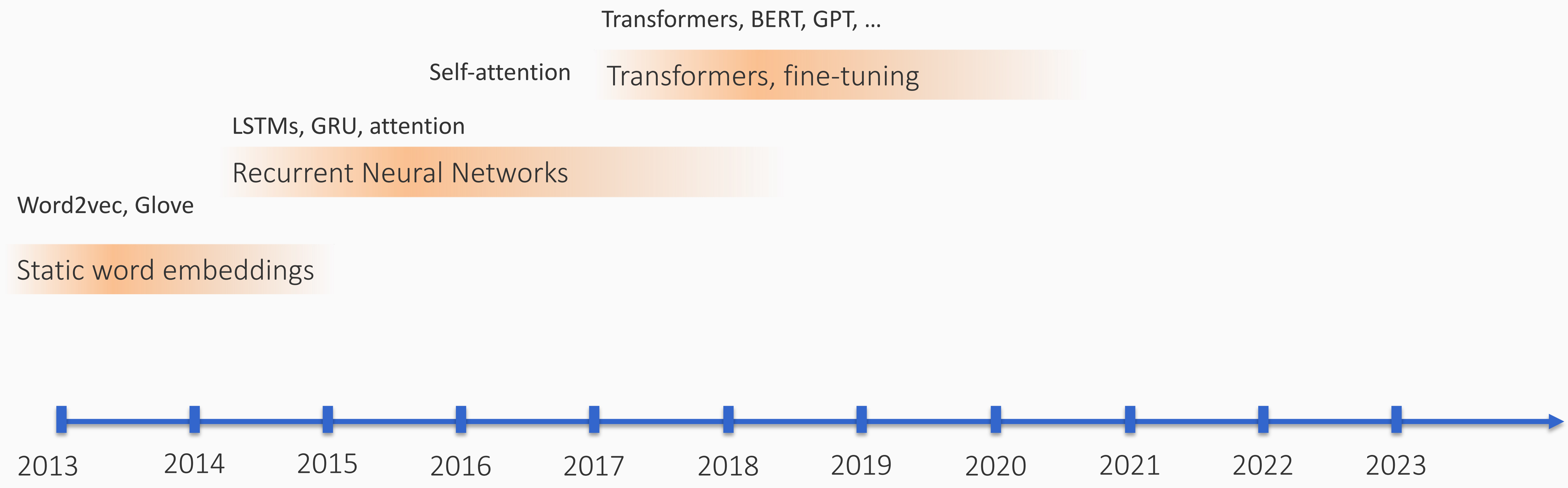
Recurrent Neural Networks

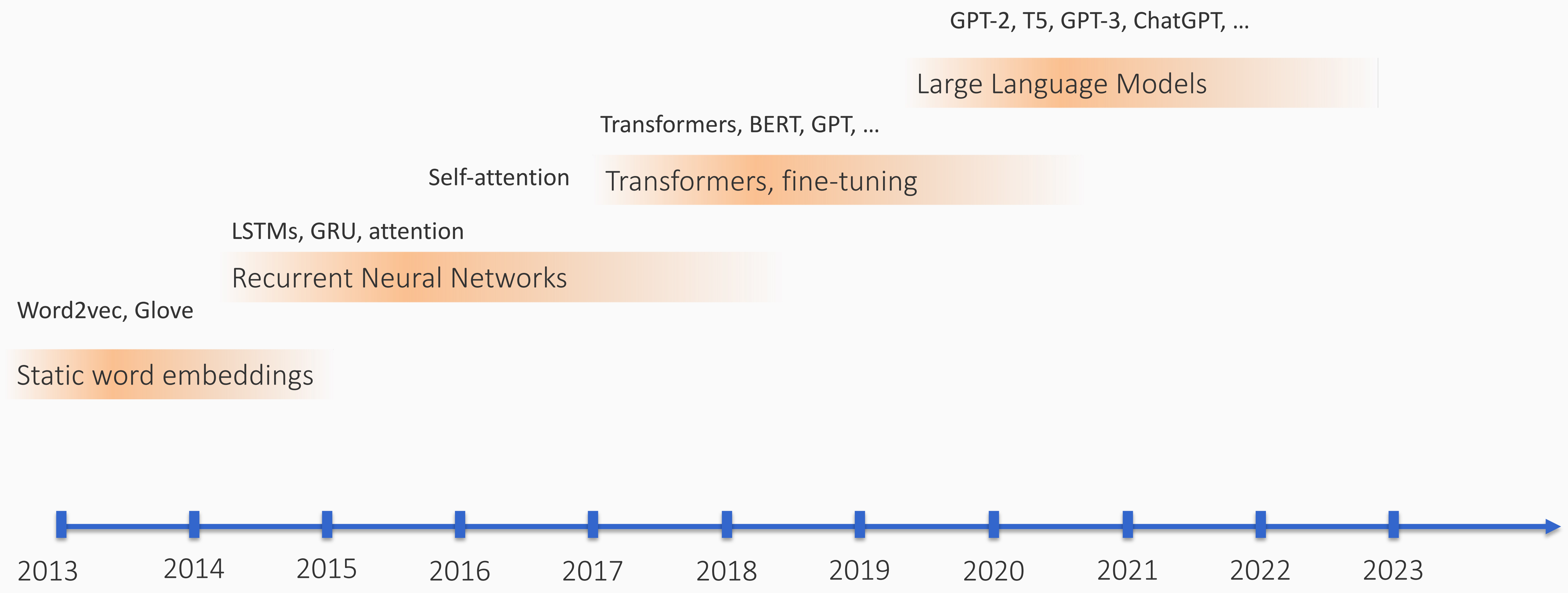
Word2vec, Glove

Static word embeddings

2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023


Self-attention
LSTMs, GRU, attention
Recurrent Neural Networks
Word2vec, Glove
Static word embeddings
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023


Transformers, BERT, GPT, …
Self-attention
Transformers, fine-tuning
LSTMs, GRU, attention
Recurrent Neural Networks
Word2vec, Glove
Static word embeddings
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023


GPT-2, T5, GPT-3, ChatGPT, …
Large Language Models
Transformers, BERT, GPT, …
Self-attention
Transformers, fine-tuning
LSTMs, GRU, attention
Recurrent Neural Networks
Word2vec, Glove
Static word embeddings
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023

Several successes of deep learning

- Word embeddings
  - A general purpose feature representation layer for words

- The de facto tool for building NLP tools today for any task
  - Syntax, semantics, coreference, …

- Language modeling
  - Starting with Bengio, 2003, several advances since then

More successes

- Machine translation
  - Neural machine translation is the norm
- Text understanding tasks
  - Natural language inference
  - Reading comprehension
- A new interface to compute and data that generates text, answers questions, writes poetry, even code
  - Even reasoning (almost)

Deep learning for NLP

Techniques that integrate

1. Neural networks for NLP, trained end-to-end

2. Learned features providing distributed representations

3. Ability to handle varying input/output sizes

What we will cover this semester

A mix of NLP and deep learning ideas

Deep learning for language

Static and contextualized embeddings of language

Recurrent neural networks

Attention, self-attention

Transformer networks

Fine-tuning transformer language models

Prompting language models

Instruction tuning, RLHF

A mix of NLP and deep learning ideas

Deep learning for language

Static and contextualized embeddings of language

Recurrent neural networks

Attention, self-attention

Transformer networks

Fine-tuning transformer language models

Prompting language models

Instruction tuning, RLHF

NLP tasks

Language modeling

Dependency parsing

Translation

Summarization

Semantic role labeling

Reading comprehension

Natural language inference

Multilinguality

Semester overview

Part 0: Introduction

- Review of neural networks
- The computation graph abstraction and gradient-based learning

Semester overview

Part 1: Representing words

- Distributed representations of words, i.e. word embeddings
- Training word embeddings using the distributional hypothesis and feed-forward networks
- Evaluating word embeddings

Semester overview

Part 2: Recurrent neural networks

- Sequence prediction using neural networks
- LSTMs and their variants
- Applications
- Word embeddings revisited

Semester overview

Part 3: Transformer networks

- The basic architecture
- Transformer language models
- BERT and its siblings
- Fine tuning

Semester overview

Part 4: The LLM era

- The impact of scale
- Prompting instead of fine-tuning
- Instruction fine-tuning
- Reinforcement learning with human feedback

Semester overview

Part 5: Where we are today and what we may see next

- Computational efficiency of transformer networks
- Long documents
- Ethical concerns
- Your favorite topic?