

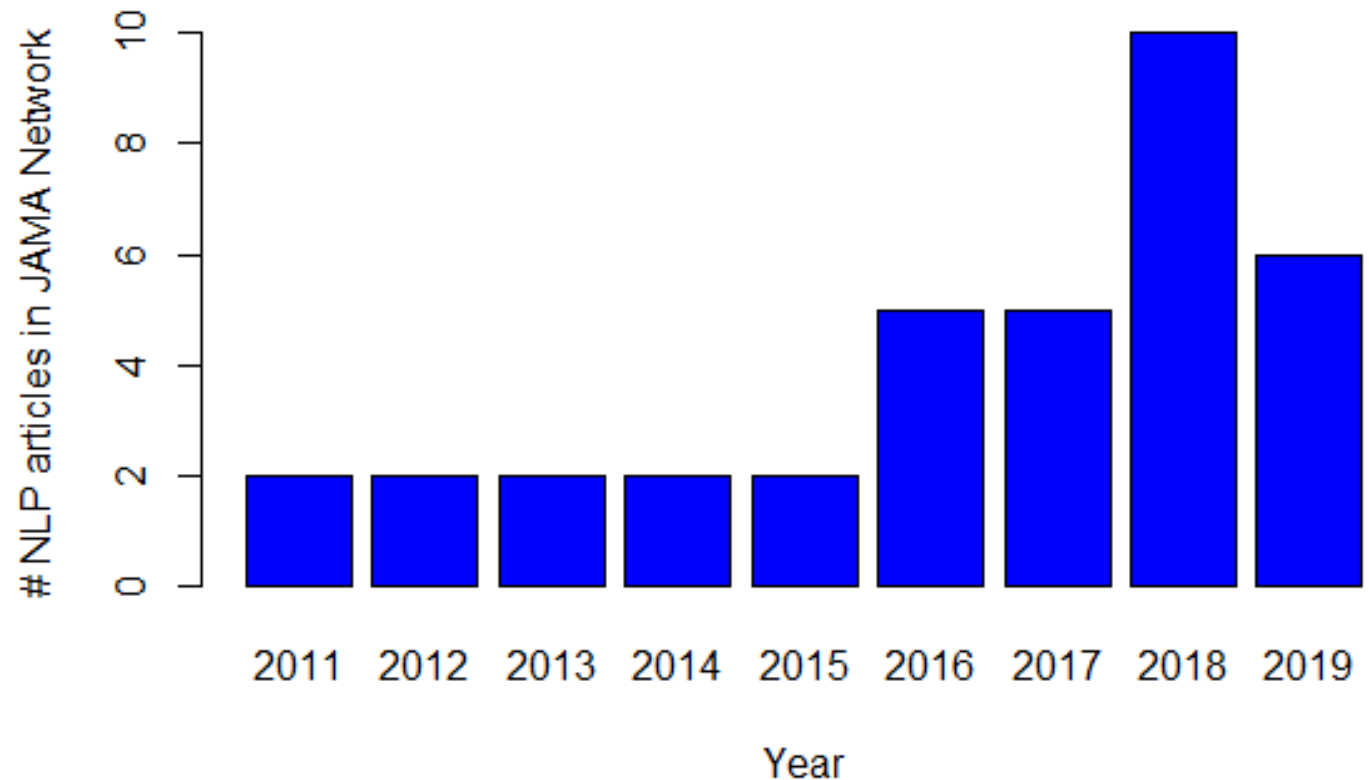
# Biomedical NLP in Practice

Matthew Engelhard

# In 2019, I did a brief survey of NLP in JAMA...

Of 28 articles:

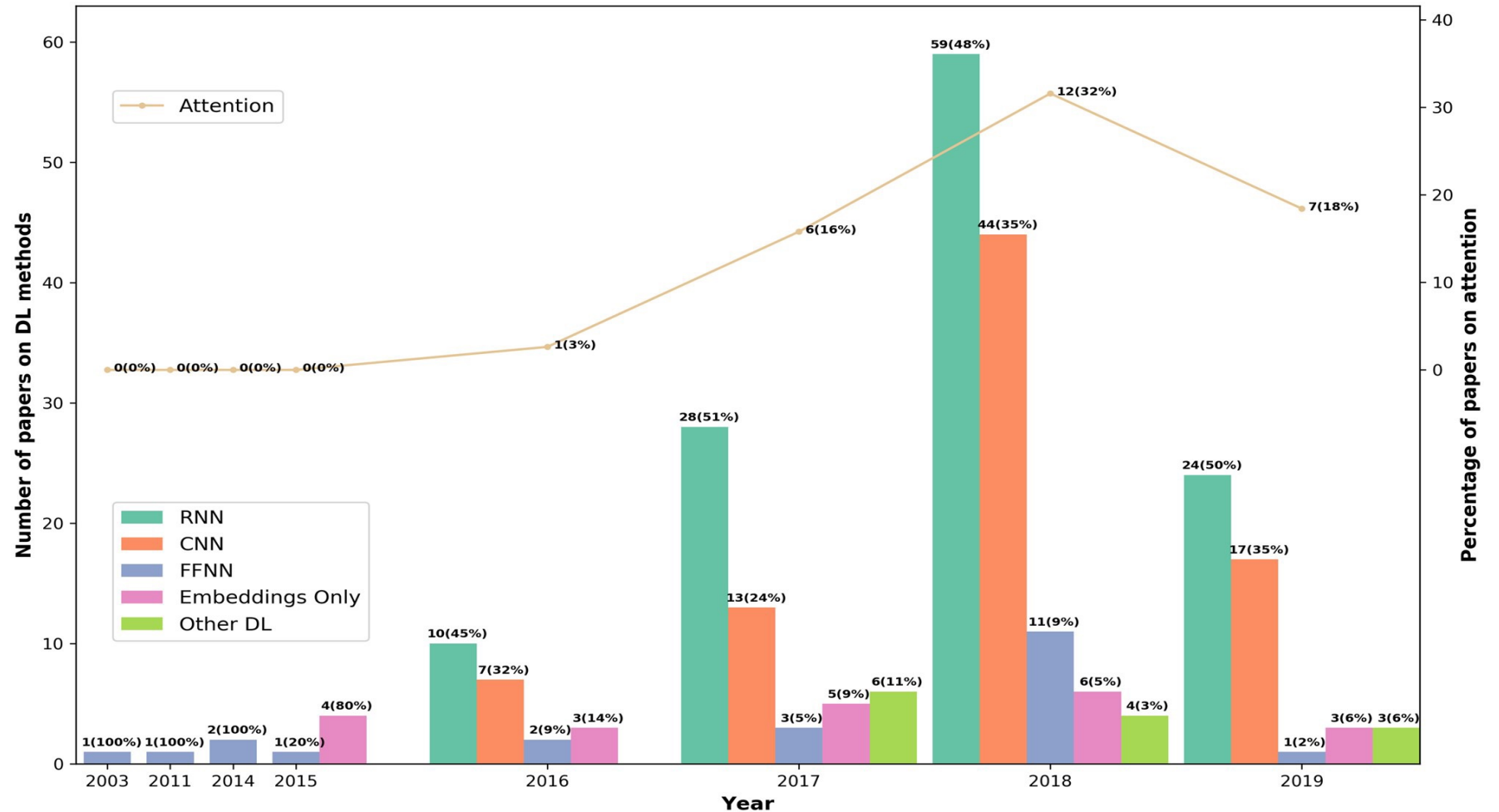
- 27 were based entirely on word counts
- Most focused on identifying specific diagnoses or events within notes
- Even today, these simple approaches often work best in practice



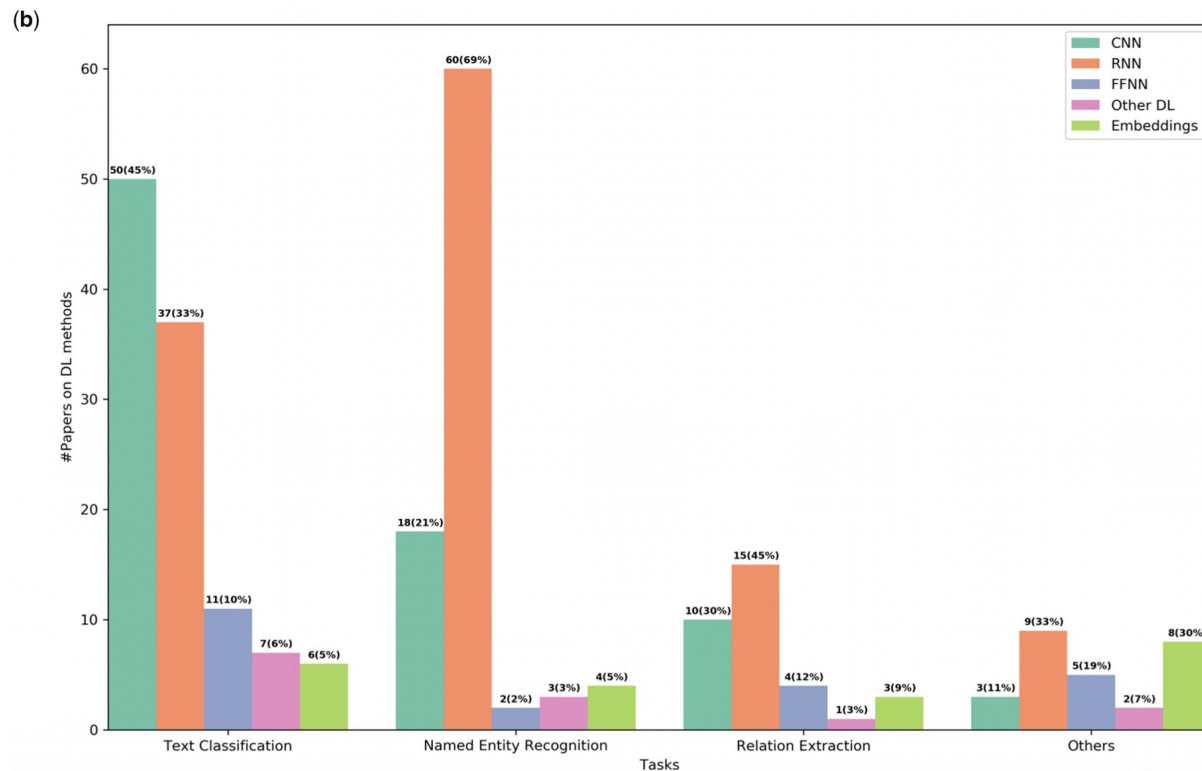
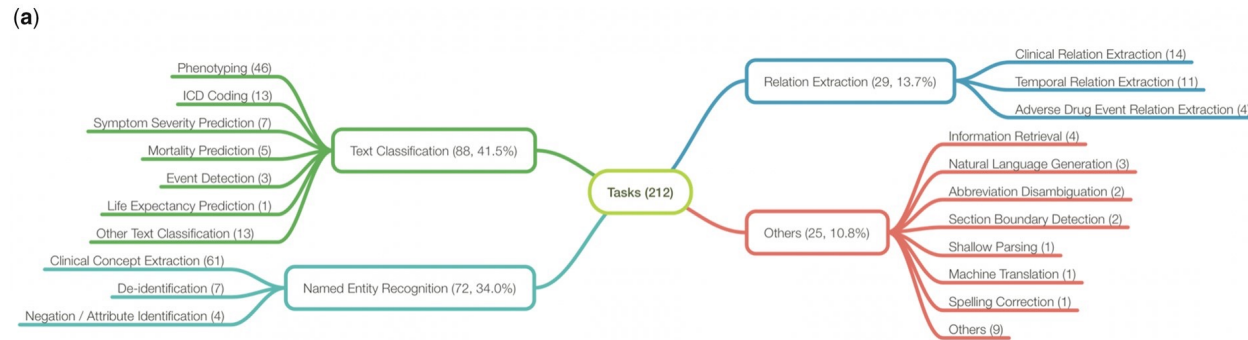
# In 2022, deep learning is becoming dominant

Why?

- Rarely worse than BoW and often better
- Can be very easy to use pre-trained models



# In 20223, deep learning is ~~becoming~~ dominant



## Common tasks:

- Text classification  
(classify a note)
- Named entity recognition  
(identify clinical concepts within a note)
- Relation Extraction  
(identify relationships between pairs of concepts within notes)
- Others...

# How does “deep” NLP work?

Answer, part 1: word vectors

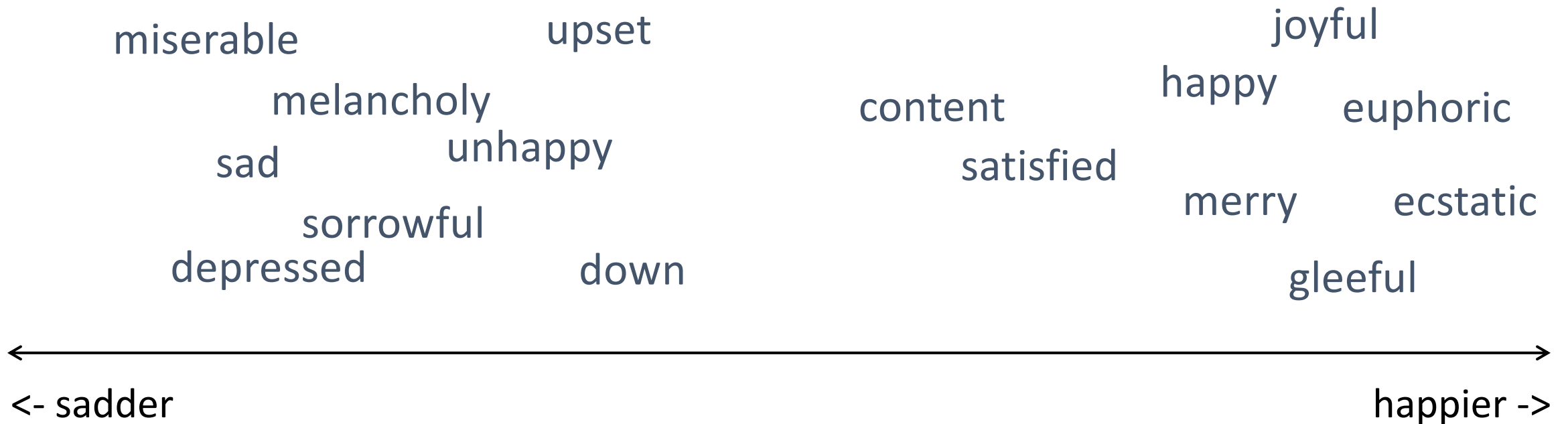
# ~~Old~~ Tried and true approach: word counts

I passed out and Mom said I was shaking

$x_3$

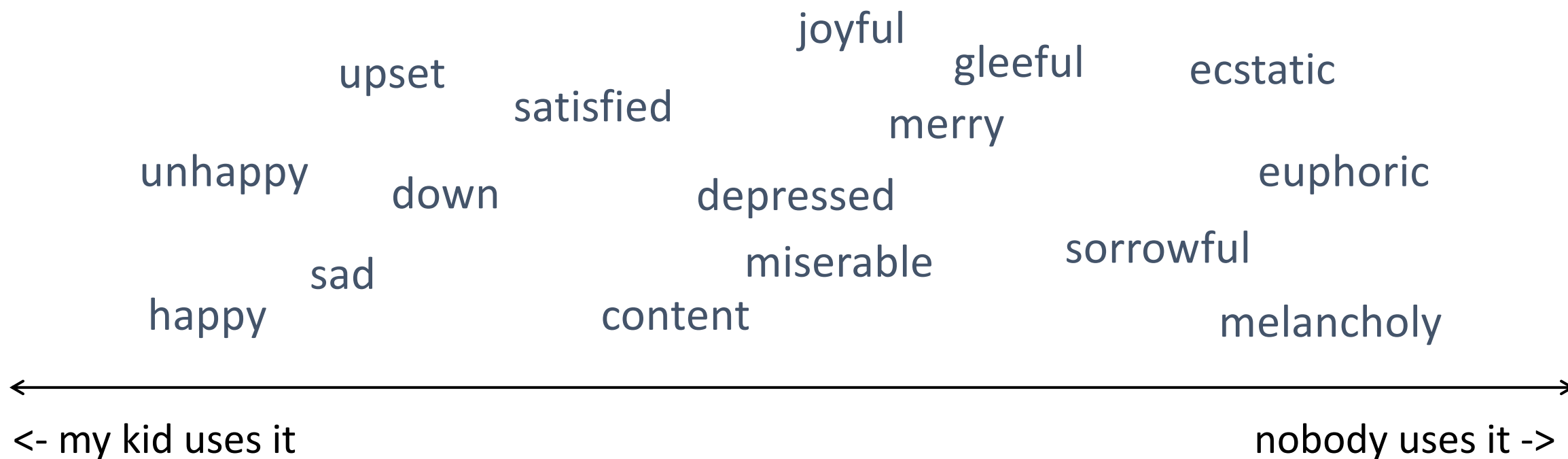
1	0	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	2	0	0	1	0	1	0
shaking	what	clinic	how	helps	was	nearest	many	with	said	months	the	morning	mom	should	sickness	and	I	is	how	out	breastfeed	passed	where

New approach: we'd like to *encode the meaning* of each word by assigning numeric attributes



Attribute 1: how happy or sad is the word?

New approach: we'd like to *encode the meaning* of each word by assigning numeric attributes



Attribute 2: how highfalutin is the word?



# We're putting all our words on a map...



- We'll use a few hundred attributes, not just two.
- The closer together two words are on this map, the more similar their meaning.

# To see why this might be a good idea, let's think about training our new robot to buy groceries



Example from Anand Chowdhury, MMCI 2019

## Grocery List

- ☐ granulated sugar
- ☐ vanilla extract
- ☐ dark brown sugar
- ☐ carrots
- ☐ table salt
- ☐ eggs

Instead of identifying items individually, we can now identify items by their attributes (including previously unseen items)

Dimension	1	10
State	Liquid	Solid
Sweetness	Bland	Sweet
Color	Light	Dark
Size	Small	Large
Carrotiness	Not really	Platonic essence of carrot



# Why does this help us?

- The model can make sense of words it hasn't seen before (weren't used in training)
- Similar words (e.g. synonyms) will have similar attributes, and therefore will have similar effect on model predictions
- (more complicated) Now we can convert text to a sequence of vectors; and we were already very good at making predictions from sequences of vectors

# How do we learn these attributes?

-> In brief, for now, but there's an additional, optional lecture on this

**KEY IDEA:** words are *defined* by the context in which they appear

A **man** strolls down the street

A **woman** strolls down the street

A **child** strolls down the street

A **crocodile** strolls down the street

A **banana** strolls down the street

A **concept** strolls down the street

# How do we learn these attributes?

KEY IDEA: words are *defined* by the context in which they appear

-> if words are always exchangeable, they must have very similar meaning



learn word meaning like an adult:  
explicit definitions

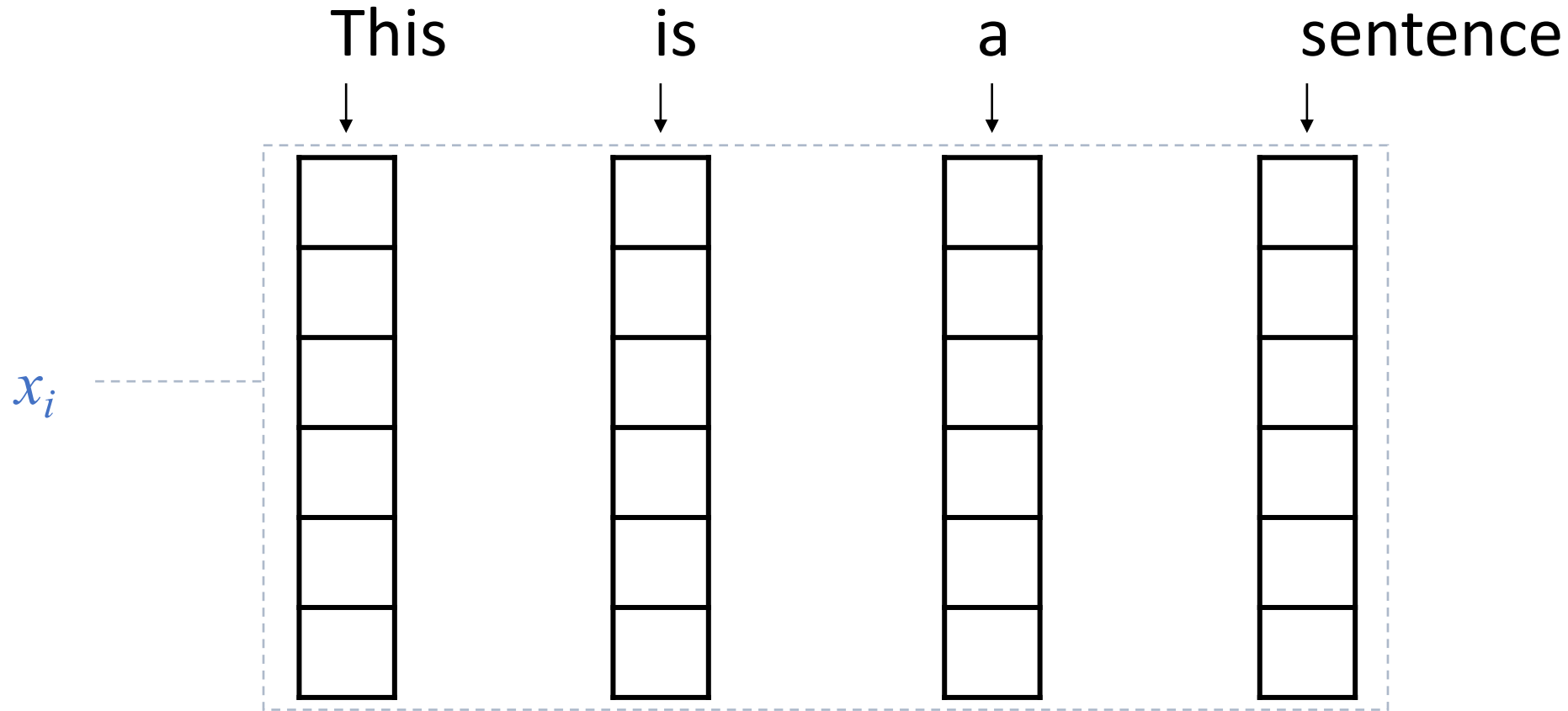
<https://www.parenting.com/activities/baby/teach-baby-to-talk/>



learn word meaning like an child:  
implicit definitions from context

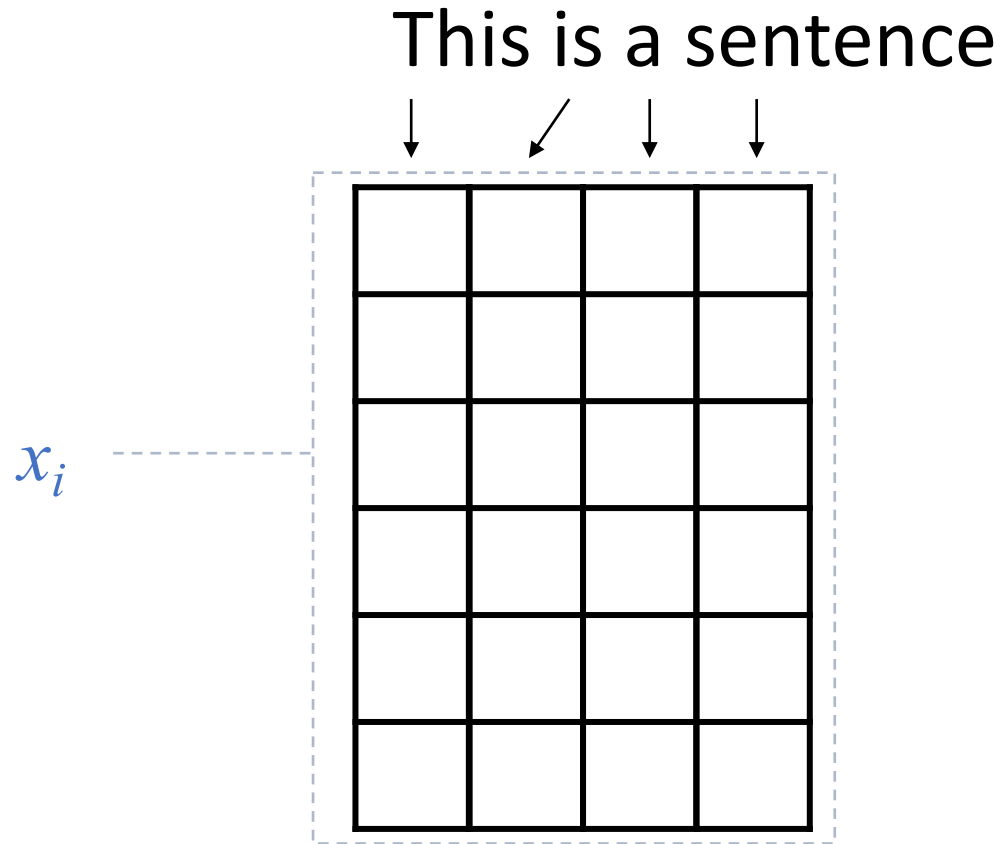
# Now that we have word vectors, how do we use them?

- Look up words individually to obtain their vectors
- Construct a sequence of vectors



# Now that we have word vectors, how do we use them?

- Look up words individually to obtain their vectors
- Construct a sequence of vectors



Now we have a grid of numbers  
Similar in many ways to an image



# How does “deep” NLP work?

Answer, part 2: hierarchical feature extraction (like in image processing)

Now we can use deep learning to build our hierarchy of features.



pixels

low-level  
motifs

- edges
- shapes
- textures

high-level  
motifs

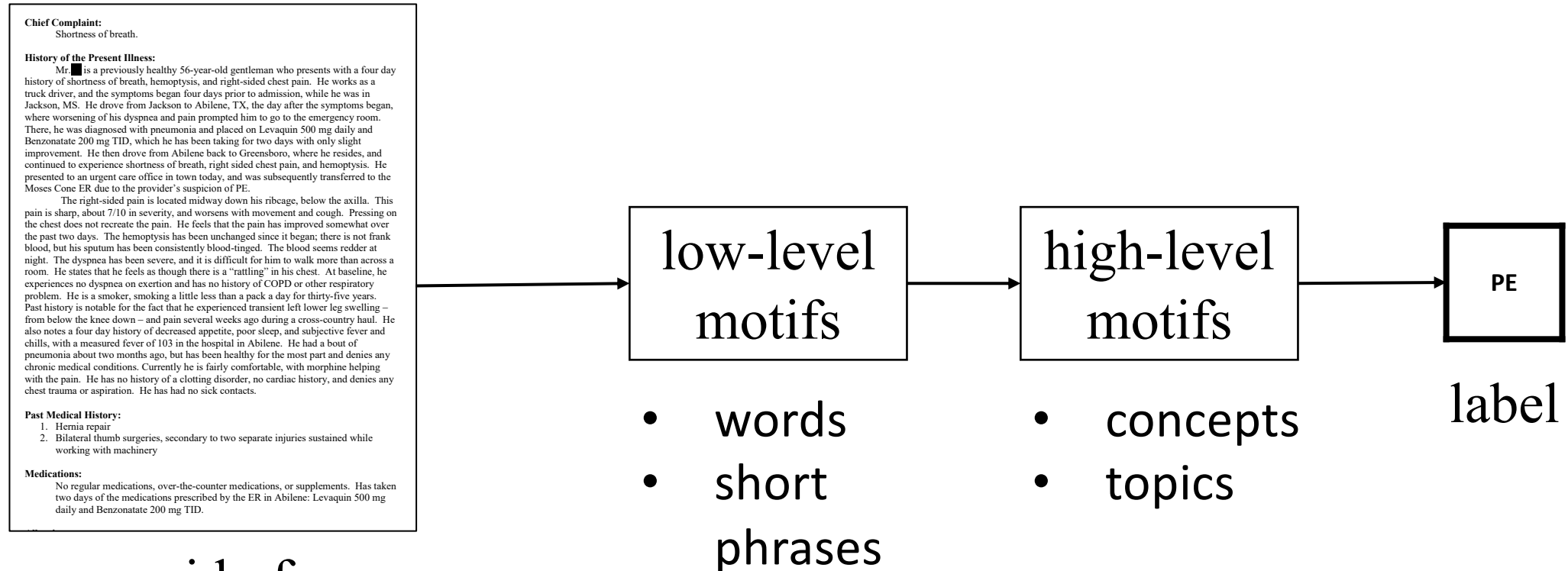
- eyes
- ears
- paws

dog

label

End goal: predict *dog* from *pixels*

# Now we can use deep learning to build our hierarchy of *semantic* features.

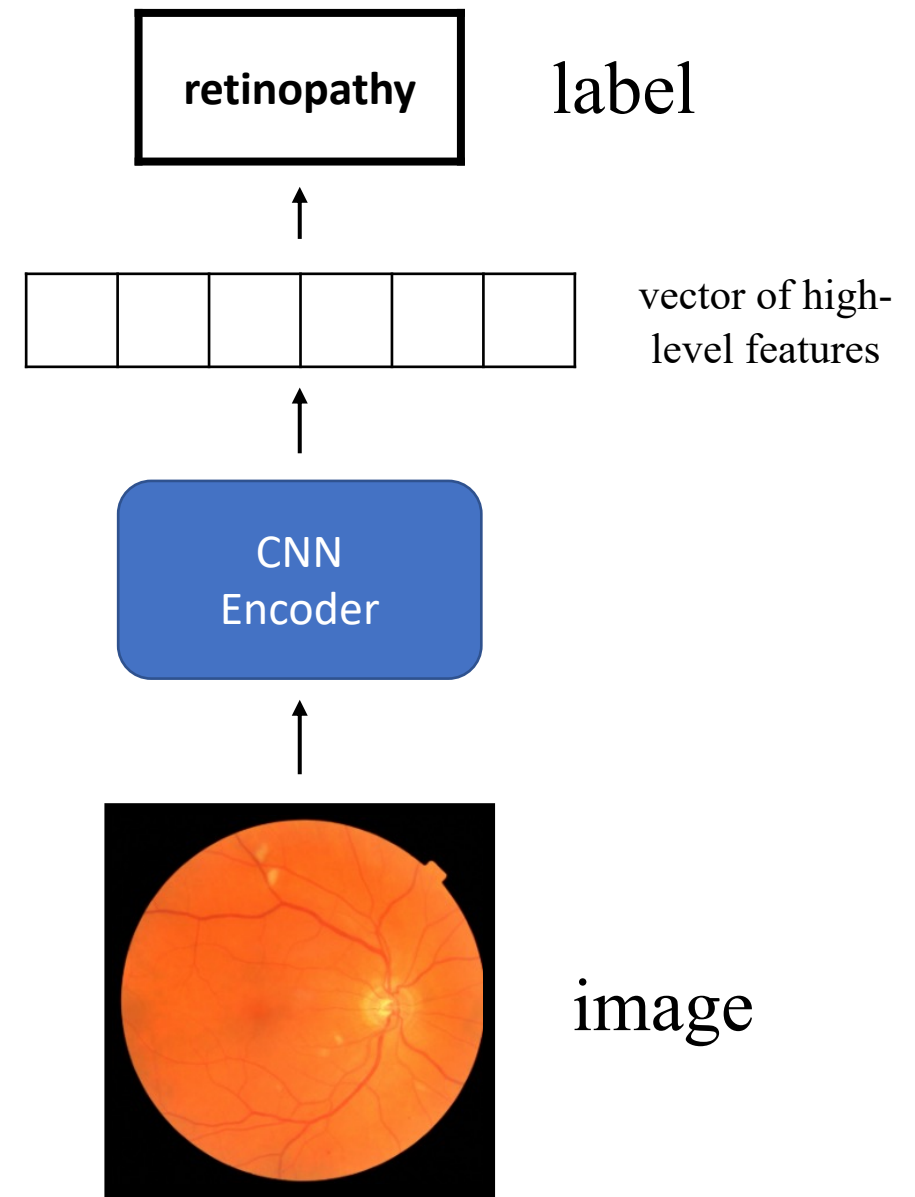


grid of  
semantic  
attributes

End goal: predict *pulmonary embolism* from *text*

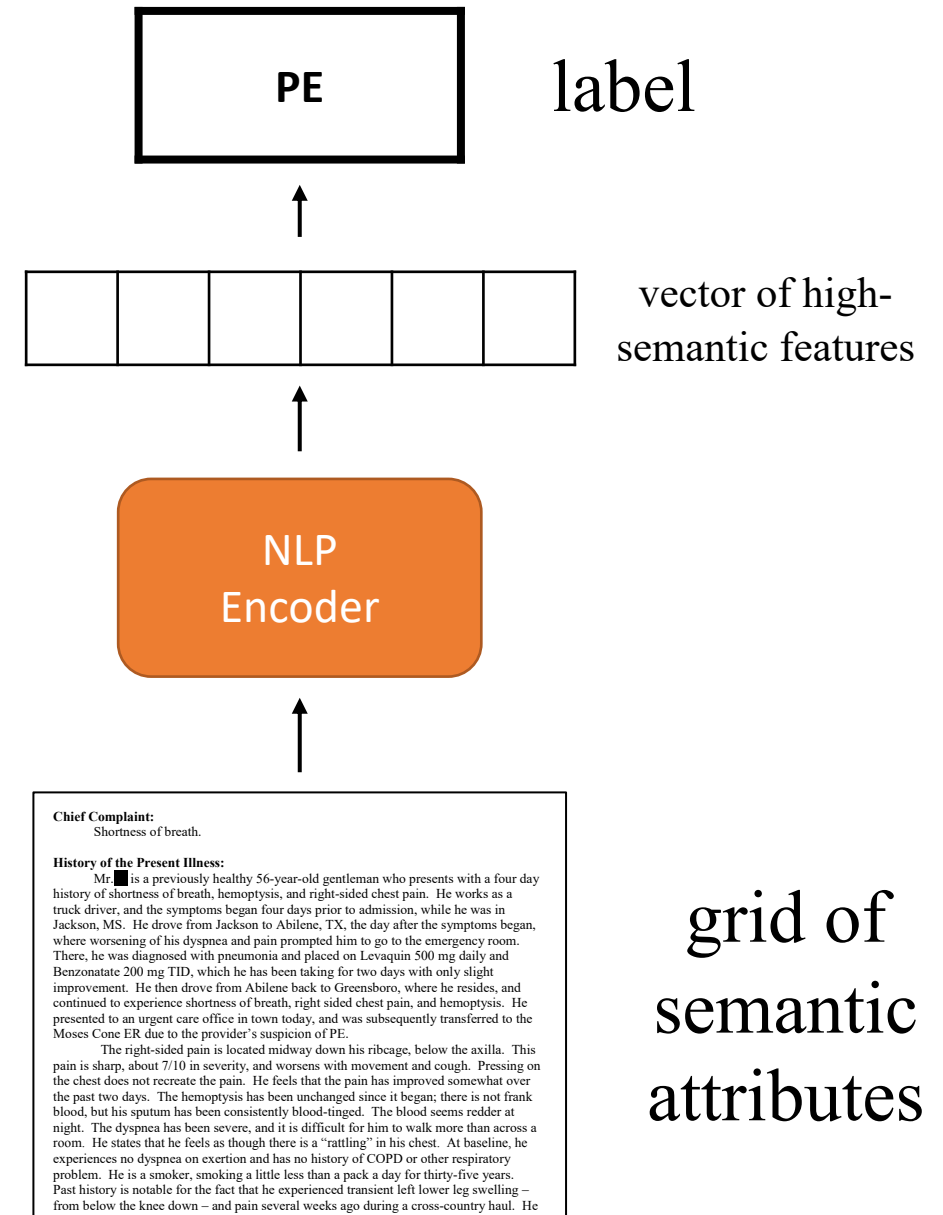
Recall: in image processing, we start with a pre-trained *encoder*

1. A CNN *image encoder* that converts the raw image to a vector of high-level motifs / features.
  2. A *final layer*, or *prediction head* – this is a logistic regression model – that makes predictions about the label from these high-level features.
- We will reuse the encoder but replace the prediction head, since it is specific to the previous (non-medical) task.

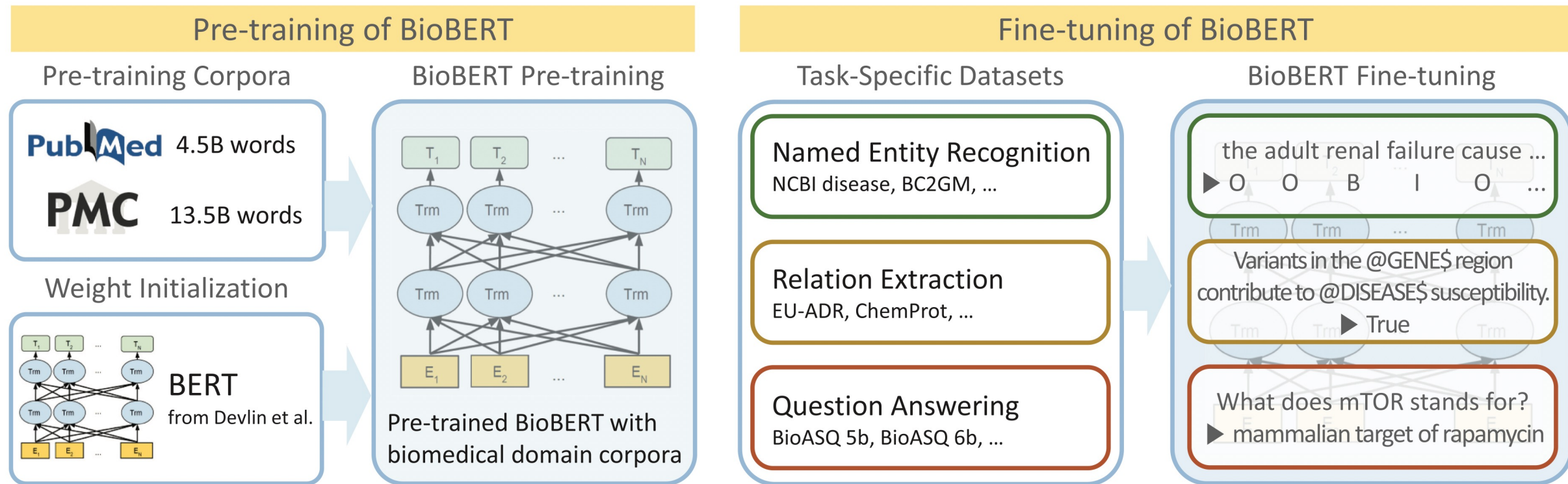


# In modern (deep) NLP, we also start with a pre-trained *encoder*

1. A transformer network *image encoder* that converts the raw semantic attributes to a vector of high-level motifs / features.
  2. A *final layer, or prediction head* – this is a logistic regression model – that makes predictions about the label from these high-level features.
- We will reuse the encoder but replace the prediction head, since it is specific to the previous task.



# Our encoder (& word vectors) is pre-trained on biomedical corpora.



**Fig. 1.** Overview of the pre-training and fine-tuning of BioBERT

Lee J, Yoon W, Kim S, Kim D, Kim S, So CH, Kang J. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. Bioinformatics. 2020 Feb 15;36(4):1234-40.

# Our encoder (& word vectors) is pre-trained on biomedical corpora.

- Common encoders (e.g. BERT, GPT3) have millions or billions of parameters (up to 1T)
- However, the principles remain the same: neural networks performing hierarchical feature extraction
- Different tasks require slightly different final modifications to the encoder
- Deep NLP is becoming more accessible (and common in the clinical literature) as tools to acquire and use these encoders continue to improve

# Named Entity Recognition

## The ~~old~~ tried and true way:

- Unified Medical Language System (UMLS)
- Apache cTAKES
- Rules-based systems to extract medical concepts from free text
- Can then build predictive models based on presence or absence of specific medical concepts

## Example: NILE

- Best approach in many cases
- Fast, easy to implement

### Narrative Information Linear Extraction (NILE)

#### Introduction

**NILE** is an efficient and effective software for natural language processing (NLP) of clinical narrative texts. It uses a prefix tree algorithm for named entity recognition, and finite-state machines for semantic analysis, both of which were inspired by the natural reading behavior of humans. The design aims to directly translate linguistic and clinical knowledge to code, allowing for the development of functions to parse complex language patterns.

The software was developed by Sheng Yu and Tianxi Cai at Harvard T.H. Chan School of Public Health and Tianrun Cai at The Brigham and Women's Hospital. It is distributed free of charge for academic and non-commercial research use by the President and Fellows of Harvard College.





# Named Entity Recognition

**The new way:** deep NLP encoder

Choose Sample Text

The patient is a 30-year-old female with a long history of insulin dependent diabetes, type 2; coron...

Text annotated with identified Named Entities

The patient is a 30-year-old female with a long history of **insulin dependent diabetes, type 2** ; **coronary artery disease** ; **chronic renal insufficiency** ; **peripheral vascular disease** , also secondary to **diabetes** ; who was originally admitted to an outside hospital for what appeared to be **acute paraplegia** , lower extremities. She did receive a course of **Bactrim** for 14 days for **UTI** . Evidently, at some point in time, the patient was noted to develop **a pressure-type wound** on the sole of her left foot and left great toe. She was also noted to have **a large sacral wound** ; this is in a similar location with **her previous laminectomy** , and this continues to receive daily care. The patient was transferred secondary to inability to participate in full physical and **occupational therapy** and continue **medical management** of **her diabetes** , the sacral decubitus, **left foot pressure wound** , and associated **complications of diabetes** . She is given **Fragmin** 5000 units subcutaneously daily, **Xenaderm** to **wounds** topically b.i.d., **Lantus** 40 units subcutaneously at bedtime, **OxyContin** 30 mg p.o. q.12 h., **folic acid** 1 mg daily, **levothyroxine** 0.1 mg p.o. daily, **Prevacid** 30 mg daily,

[https://demo.johnsnowlabs.com/healthcare/NER\\_CLINICAL/](https://demo.johnsnowlabs.com/healthcare/NER_CLINICAL/)

# Fine-tuning LLMs is *rapidly* becoming more accessible.

 Meta AI

Research

## Introducing LLaMA: A foundational, 65-billion-parameter large language model

February 24, 2023

### Join the Hugging Face community

and get access to the augmented documentation experience



Collaborate on models, datasets and Spaces



Faster examples with accelerated inference



Switch between documentation themes

[Sign Up](#)

to get started

### PEFT

😊 PEFT, or Parameter-Efficient Fine-Tuning (PEFT), is a library for efficiently adapting pre-trained language models (PLMs) to various downstream applications without fine-tuning all the model's parameters. PEFT methods only fine-tune a small number of (extra) model parameters, significantly decreasing computational and storage costs because fine-tuning large-scale PLMs is prohibitively costly. Recent state-of-the-art PEFT techniques achieve performance comparable to that of full fine-tuning.

PEFT is seamlessly integrated with 😊 Accelerate for large-scale models leveraging DeepSpeed and [Big Model Inference](#).

#### Get started

Start here if you're new to 😊 PEFT to get an overview of the library's main features, and how to train a model with a PEFT method.

#### How-to guides

Practical guides demonstrating how to apply various PEFT methods across different types of tasks like image classification, causal language modeling, automatic speech recognition, and more. Learn how to use 😊 PEFT with the DeepSpeed and Fully Sharded Data Parallel scripts.

# Conclusions

- Text data are central to clinical medicine, so the potential for NLP impact is high (but *not yet realized*)
- Simple, count-based NLP models are surprisingly effective in most clinical applications.
- Complex, deep learning NLP models have exceeded human performance. In these models, words are converted to vectors of semantic attributes, and increasingly complex, hierarchical semantic features are then extracted.
- Similar to image processing, we can take advantage of complex NLP models by repurposing them for a specific clinical task via fine-tuning of parameters.

# A brief note on interpretability...

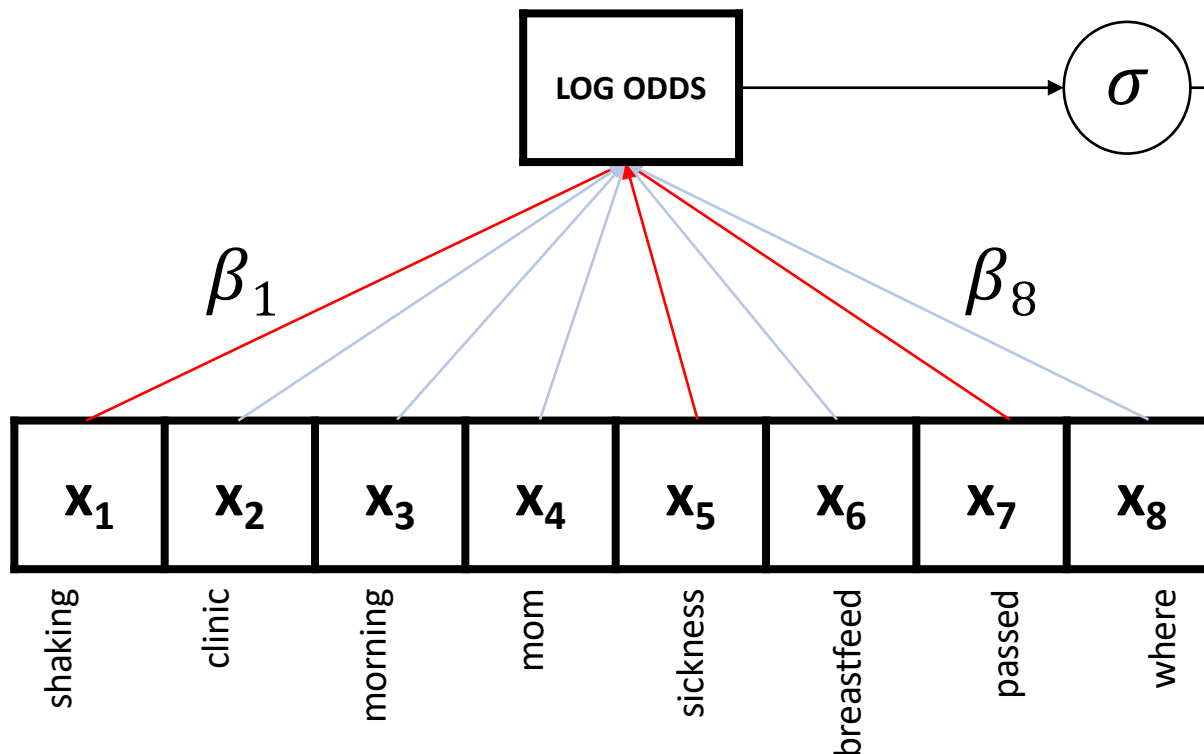
More on this next time.

# We can *interpret* a count-based NLP model

- Suppose you use logistic regression with count-based features, and your model predicts that that an SMS you receive is urgent.
- **Q:** Is it hard to figure out why it made that prediction?

# We can *interpret* a count-based NLP model

- Suppose you use logistic regression with count-based features, and your model predicts that that an SMS you receive is urgent.
- **Q:** Is it hard to figure out why it made that prediction?
- **A:** No. You can look at the coefficients to see which words increased and decreased the predicted probability.



$y$ , associated label:  
(0 = not urgent, 1 = urgent)

# Can we interpret a deep learning NLP model?

- Suppose you apply a deep neural network to a sequence of word vectors, and your model predicts that that the SMS you receive is urgent.
- **Q:** Is it hard to figure out why it made that prediction?