



AI & Smart Tech: Words, words, words



Thanks for having me!

Tom Metcalfe

*Machine learning engineer at [Rasa](#)
t.metcalfe@rasa.com*

[@MetcalfeTom](#) on Github



Rasa is the leading open source machine learning toolkit for developers to extend bots beyond answering simple questions

100k+

Downloads

1,000+

Community Support

100+

Contributors

ERGO

RAIFFEISEN

UBS



helvetia

YellowPages

friendsurance



Allianz

Structure of this talk

- I. Why words?
- II. Teaching words
- III. Teaching meaning
- IV. Research from Rasa
- V. Where will words go?

Why words?

How and why do we
use words as data?

Why words?

Machine learning is all about information

We have a **wealth of information** to gather from words and literature

Whether text or speech, people use words to communicate, so good word representation **facilitates AI's integration**

Talk  Books

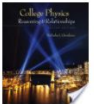
why is the sky blue?

Go!

Passages from books

Insight 25.3 **WHY IS THE SKY BLUE AND NOT VIOLET?** The Rayleigh scattering of light by molecules in the atmosphere gets stronger as the wavelength decreases. [\(view in book\)](#)

from *College Physics: Reasoning and Relationships*
by Nicholas Giordano



That is, the longer the wavelength, the less the scattering. **It is this scattering that gives rise to the blue color of the sky.** (See the Highlight: Blue Skies and Red Sunsets.) [\(view in book\)](#)

from *An Introduction to Physical Science*
by James Shipman, Jerry D. Wilson, Charles A. Higgins



...most strongly this time out of the beam reaching the eyes; therefore the loss of blue results in a yellow appearance. **It is because of this that the sky appears blue (Rayleigh scattering).** For particles of dimensions similar to the wavelength range of visible light (500nm), scattering achieves maximum intensity... [\(view in book\)](#)

from *Bioengineering of the Skin: Skin Imaging & Analysis*
by Klaus-Peter Wilhelm, Peter Elsner, et. al.



Google's Talk to Books

This is not an easy task

Why words?

Words are nonsense

“*Uksrukuaktak*”

{Iñupiat Eskimo dictionary}



Why words?

Vocabularies are vectors

Banana : [1 0 0 ... 0 0 0]

Bubblegum: [0 1 0 ... 0 0 0]

Caramel : [0 0 1 ... 0 0 0]

⋮

⋮

Vanilla : [0 0 0 ... 0 0 1]

Each word is a feature

Why words?

Why words?

- How do words even work anyway?

They are the smallest element of language that can be uttered in isolation with objective or practical meaning.

- Why use them?

They are convenient representations which we use to externalise our thoughts. There is a wealth of knowledge to be acquired from literature

- How do we use words as data?

Separate at whitespace and assign each word a linearly-independent vector

Teaching words

Why word vectors are
bad

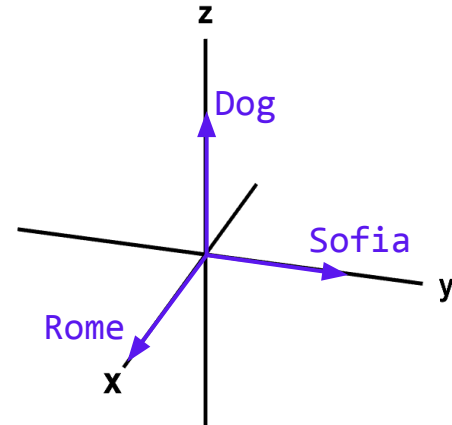
One-hot vectors are bad

Complete the sentence: “A holiday in ____”

Rome : [1 0 0]

Sofia : [0 1 0]

Dog : [0 0 1]



One-hot vectors are bad

Rome : [1 0 0 0]

Sofia : [0 1 0 0]

London : [0 0 0 1]

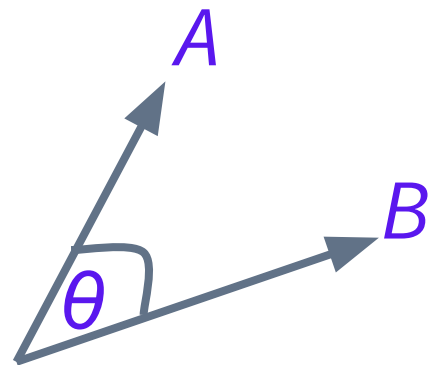
Dog : [0 0 1 0]

Language	Words in the Dictionary
Korean	1,100,373
Japanese	500,000
Italian	260,000
English	171,476

Teaching words

Cosine similarity

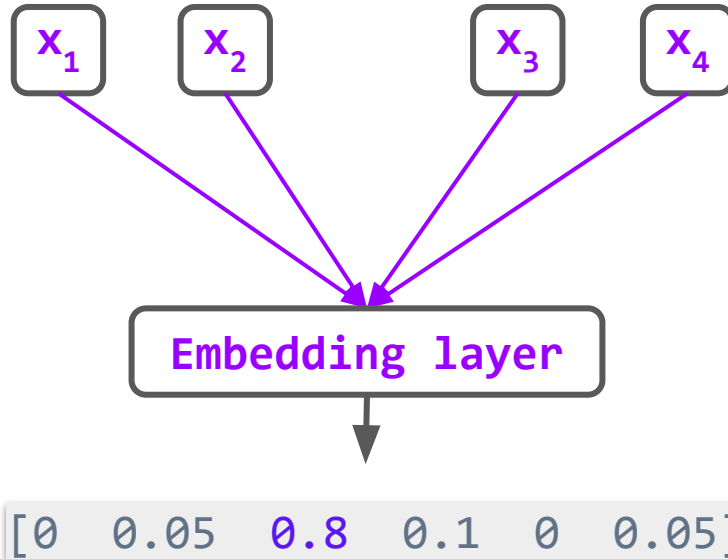
$$\text{sim}(A, B) = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|}$$



$$\text{sim}(\text{Rome}, \text{Sofia}) = 0$$

Learning word embeddings

“Bacon and ___ with toast”



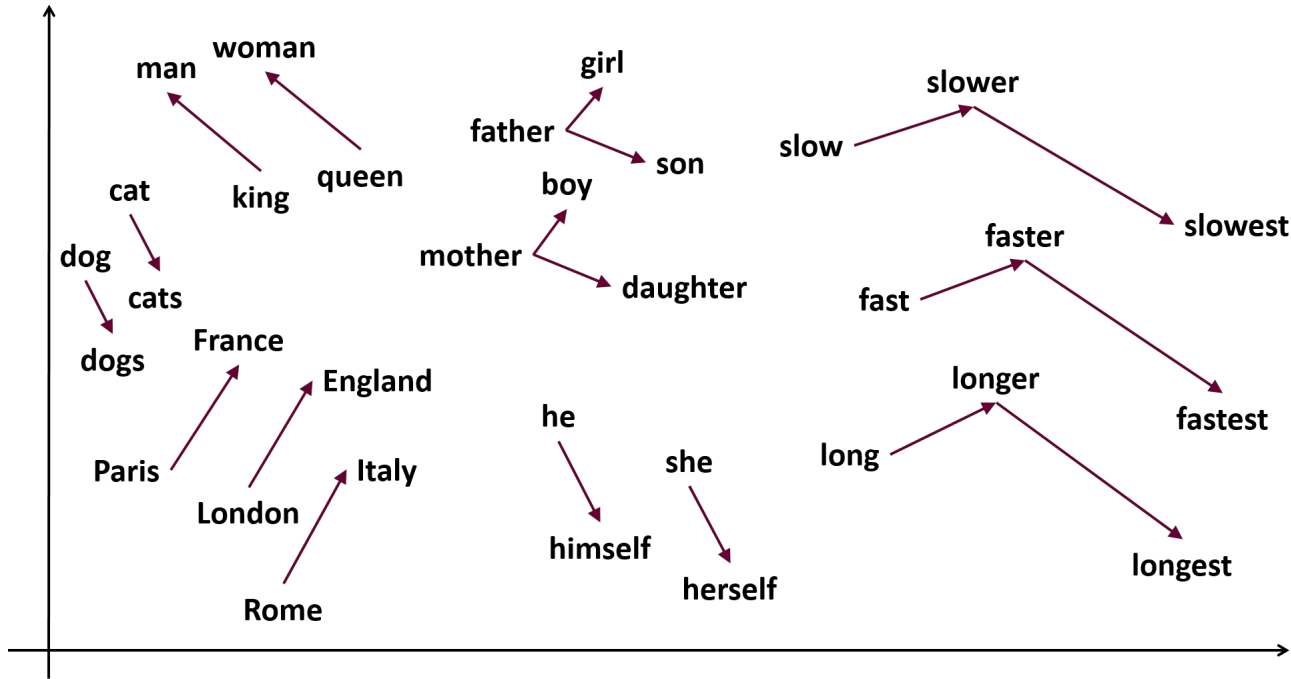
egg?

Using learned embeddings



$$\text{sim}(E(\text{Man}), E(\text{King})) = 0.7$$

Properties of word embeddings



Embedding space captures relations

OK, but does this make
things better?

Intent classification with Rasa NLU

Goal: classify user messages to a list of “intents”

```
greet
goodbye
thank_you
restaurant_search
book_table
confirm
```

Example intent list

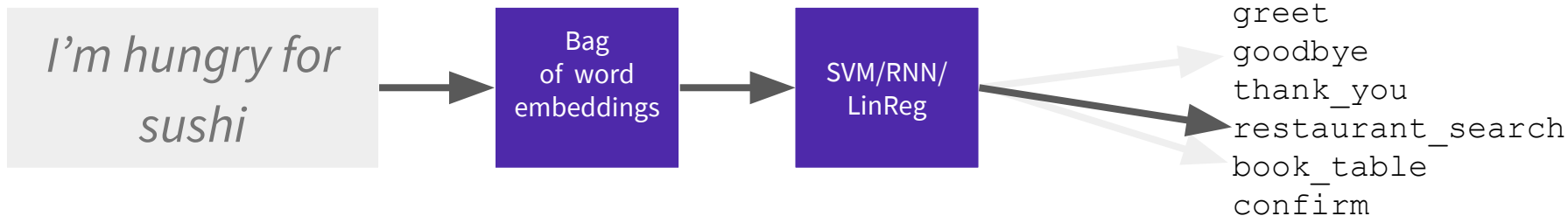


I'm hungry for sushi

Example user message

Intent Classification with word vectors

Embedded words sentence representation: $\{v_1, \dots, v_s\} \rightarrow \frac{1}{s} \sum_i v_i$



*This works **embarrassingly well**, and generalises from little data*

Teaching words

- Why one-hot vectors are bad

Being sparse causes slow learning, related words share no association

- What are word embeddings?

Word embeddings are a mapping from one-hots to a multi-dimensional space where synonymous words occupy similar regions

- What do word embeddings do?

Capture analogous information, increase task accuracy with low amounts of data

Teaching Meaning

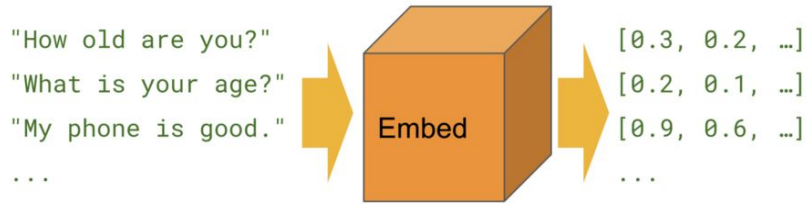
Why word embeddings
are bad

Word embeddings are bad

“No I don’t want sushi”

“I’d go hungry before eating sushi”

Sentence Embeddings



*Universal Sentence Encoder
{Google Research, 2018}*

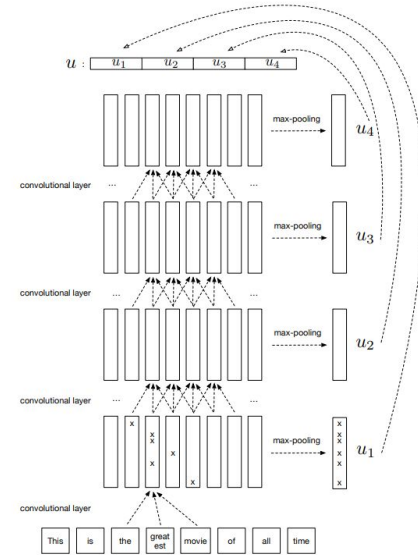
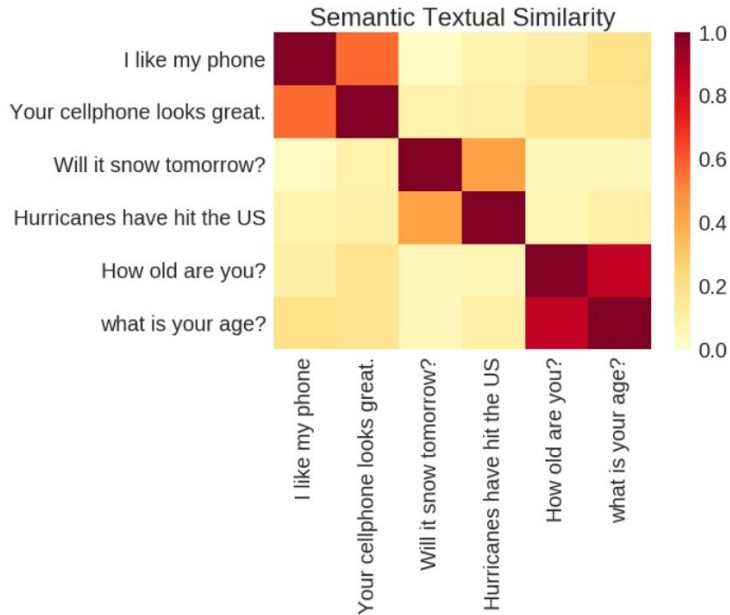


Figure 4: **Hierarchical ConvNet architecture.**

*InferSent encoder architecture
{Facebook Research, 2017}*

How good are sentence embeddings?



*They don't fall into the **saliency trap***

They capture meaning rather than word similarity

*Universal Sentence Encoder
{Google Research, 2018}*

How good are sentence embeddings?

*They have interesting **task specialisation***

They can capture information on a larger scope than word embeddings alone

premise	hypothesis	label
"Two women are embracing while holding to go packages."	"Two woman are holding packages."	entailment
A man is typing on a machine used for stenography	The man isn't operating a stenograph	contradiction
"A man is singing a song and playing the guitar"	"A man is opening a package that contains headphones"	1.6
"Liquid ammonia leak kills 15 in Shanghai"	"Liquid ammonia leak kills at least 15 in Shanghai"	4.6

*Inferent embedding tasks
{Facebook Research, 2017}*

Teaching meaning

- Why word embeddings are bad

Don't capture information from sentence structure like negation. Spelling errors and out-of-vocabulary words still cause problems

- How do we embed whole sentences?

Google use a deep averaging network, Facebook AI use multi-level representations, Skipthought takes the hidden representation of a whole sentence

- What do sentence embeddings do?

Allow for sentiments to be captured rather than synonymous relevance

Time for a break 🤖

Research from Rasa

Embedding sentences
for FAQ detection

FAQ detection: hypothesis

Can we use sentence embeddings to classify duplicate questions?

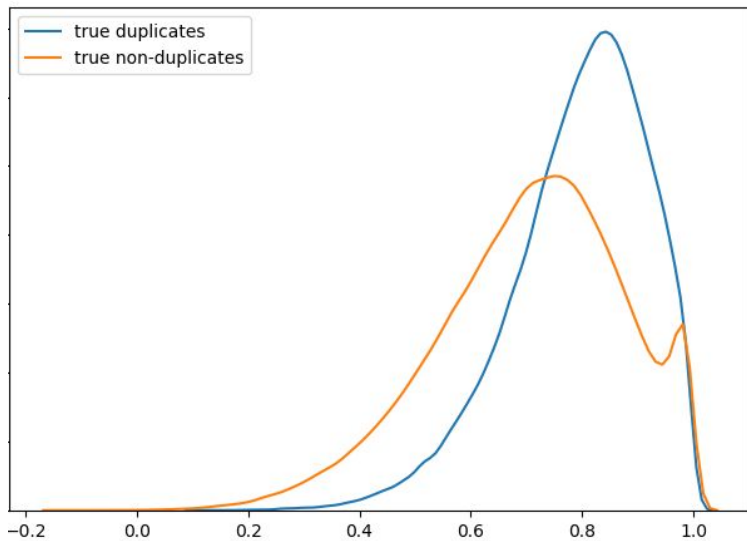
Question 1	Question 2	is_duplicate
How has your life changed since you became HIV positive?	How did your life change after you were diagnosed with HIV positive?	1
What is the defence mechanism in plants?	Is there any immune system in plants?	1
How long does WhatsApp keep the delivered data like images or etc. in the server?	How do I share GIF images in WhatsApp without converting the GIF into a video?	0
Within the context of dispute resolution, what is natural justice?	You earn a living to live well not to kill yourself with stress. What thoughts went through your mind after reading this sentence?	0

QUORA duplicate questions dataset

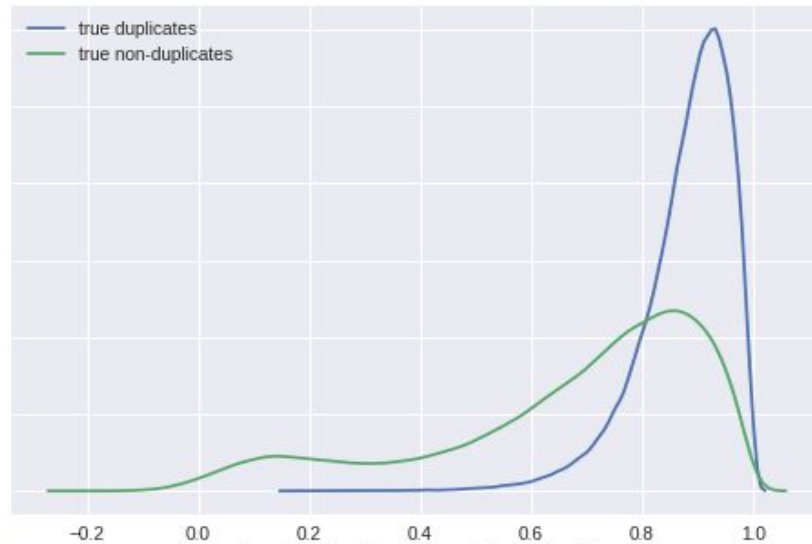
FAQ detection: method

1. *Embed the questions using Google's pre-trained **Universal Sentence Encoder***
2. *Measure **cosine similarity** between duplicate and non-duplicate questions*
3. *Determine if classification is possible*

FAQ detection: results

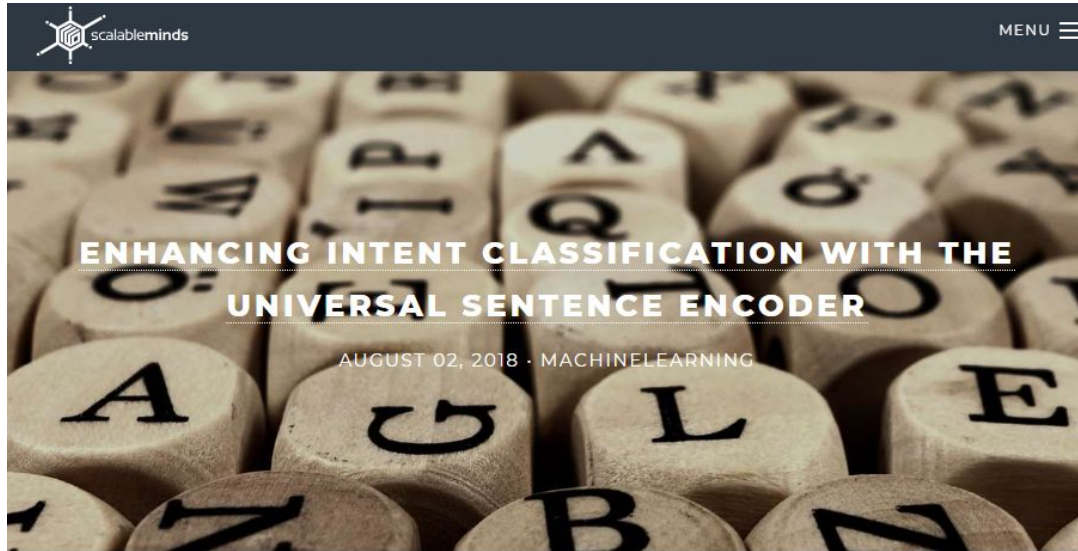


Similarity between word embeddings



Similarity between sentence embeddings

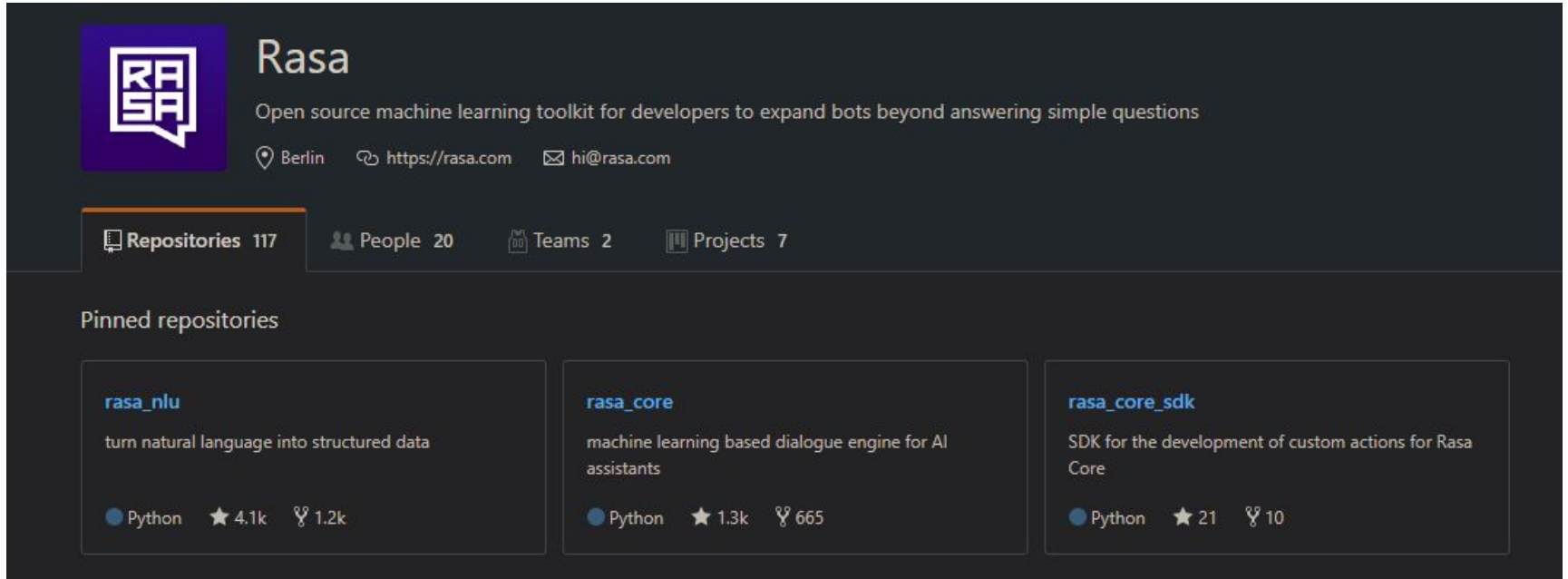
If you try 'em out, let us
know



TLDR: We built a custom [Rasa Featurizer](#) which appends a [Universal Sentence Encoding](#), improving intent classification performance by overcoming inherent limitations of bag-of-words models. See [this Gist](#) for the code.

When building chatbots, the first step is usually to build an *intent classifier* in order to put arbitrary user messages (“*Can you reserve me a restaurant table?*”) to a pre-defined *intent* (`request_restaurant_reservation`). These are usually bag-of-words (BoW) models because they require little training data and are a surprisingly effective baseline.

Rasa is open source



Rasa
Open source machine learning toolkit for developers to expand bots beyond answering simple questions

Berlin <https://rasa.com> hi@rasa.com

Repositories 117 People 20 Teams 2 Projects 7

Pinned repositories

- rasa_nlu**
turn natural language into structured data
Python ★ 4.1k 🍴 1.2k
- rasa_core**
machine learning based dialogue engine for AI assistants
Python ★ 1.3k 🍴 665
- rasa_core_sdk**
SDK for the development of custom actions for Rasa Core
Python ★ 21 🍴 10

Where will words go?

...are words bad?

Where will words go?

Uksrukuaktak



Where will words go?

The Eskimo library



Where will words go?

The Eskimo library



Emily M. Bender (@emilymbender)

I am not interesting LM for LM sake --- I'm saying that if the "embeddings" that are sold as a representation of sentence (meaning) are learned in the context of an LM task they can't possibly represent meaning. →

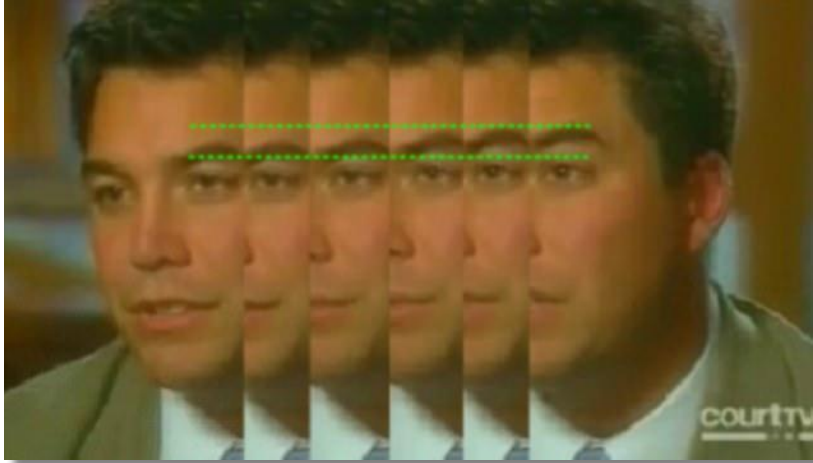
Where will words go?

Grounding words in reality

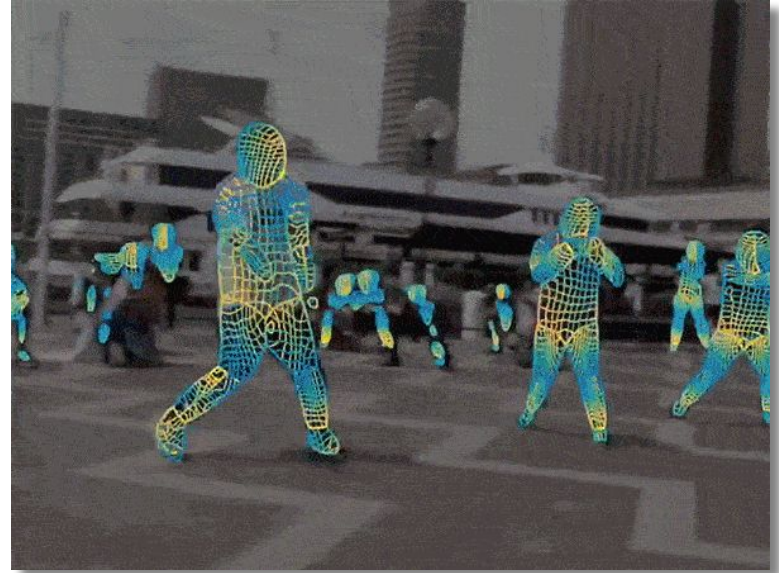
“Go grab my lunch”

Where will words go?

Untapped languages



Deception Detection in Videos
{University of Maryland, 2018}

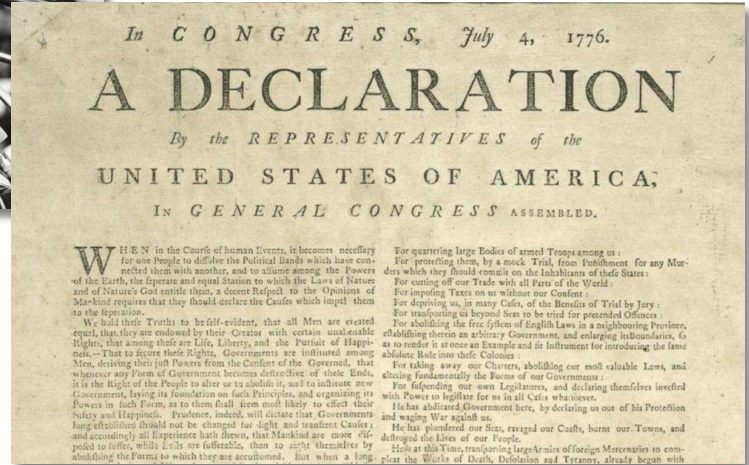
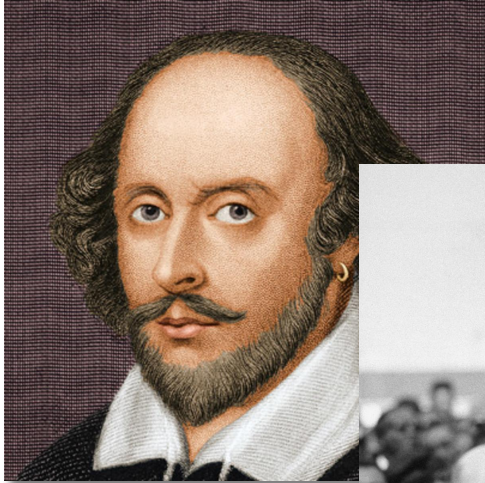


DensePose: Dense Human Pose Estimation in the Wild
{Facebook, 2018}

Words are beautiful

Where will words go?

Words are beautiful



Words are disastrous

Where will words go?

Words are disastrous

*Language
barriers*

Dyslexia

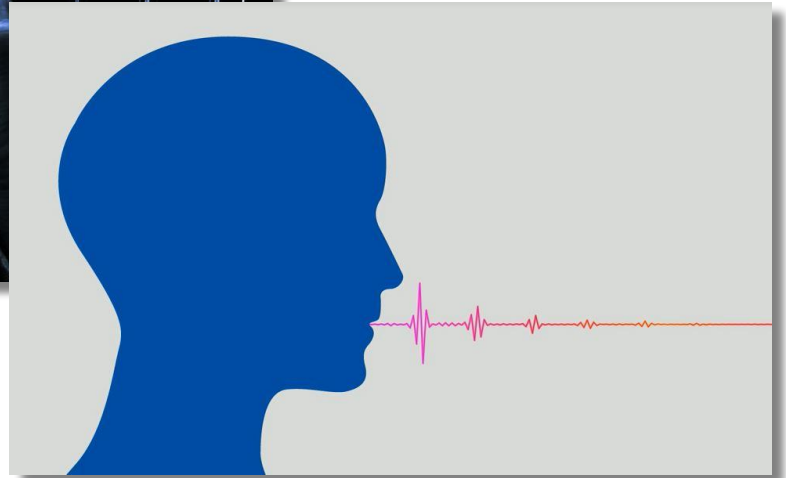
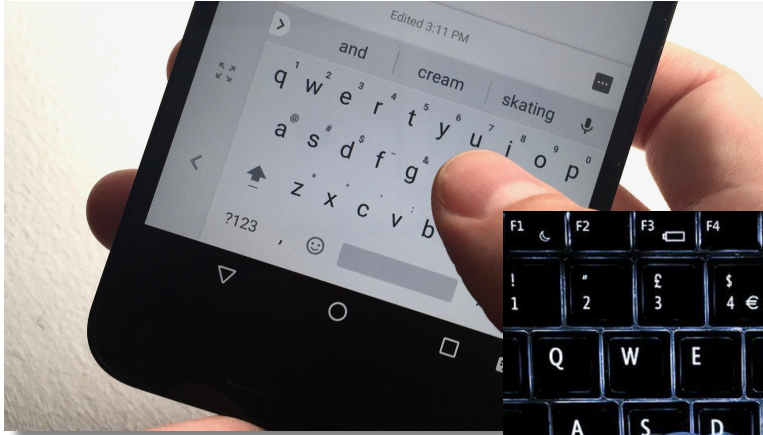
Misunderstandings

Aphasia

*Speech
impediments*

Where will words go?

Text, type, talk



Words are emissaries

**Thanks for
listening to my
words!**



Tom Metcalfe

Machine Learning Engineer

t.metcalfe@rasa.com