UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Faculty of Engineering,
Built Environment and
Information Technology

Fakulteit Ingenieurswese, Bou-omgewing en
Inligtingtegnologie / Lefapha la Boetšenere,
Tikologo ya Kago le Theknolotši ya Tshedimošo

Special Topic: Word Embeddings + Language Models

Dr. Vukosi Marivate

Inputs:
- A Modupe [PhD Candidate]
- A Moodley [MIT Big Data Science Student]

Make today matter

Data Science for Social Impact

Word Embeddings + Language Models

Word Embeddings

“You shall know a word by the company it keeps”

“Tell me who your friends are, and I will tell you who you are.”

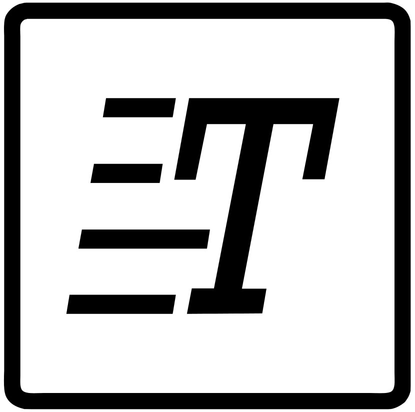
The Distributional Hypothesis is that words that occur in the same contexts tend to have similar meanings [2]

Word Embeddings

“You shall know a word by the company it keeps”

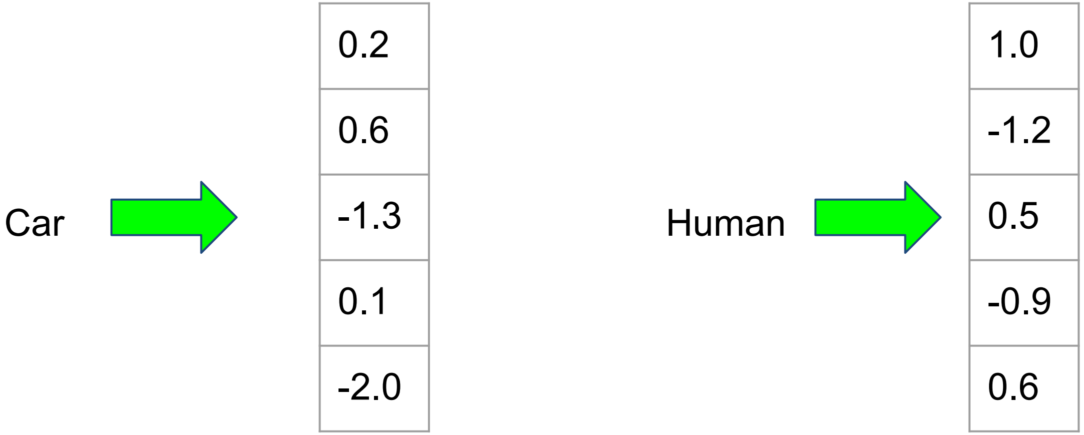
“Tell me who your friends are, and I will tell you who you are.”

The Distributional Hypothesis is that words that occur in the same contexts tend to have similar meanings [2]



Word Vectors

- Mapping from tokens to a continuous vector space
- Trained using a shallow neural network (not deep)



Car
0.2
0.6
-1.3
0.1
-2.0

Human
1.0
-1.2
0.5
-0.9
0.6

https://www.youtube.com/watch?v=UqRCEmrv1gQ

Word Vectors - The Idea!

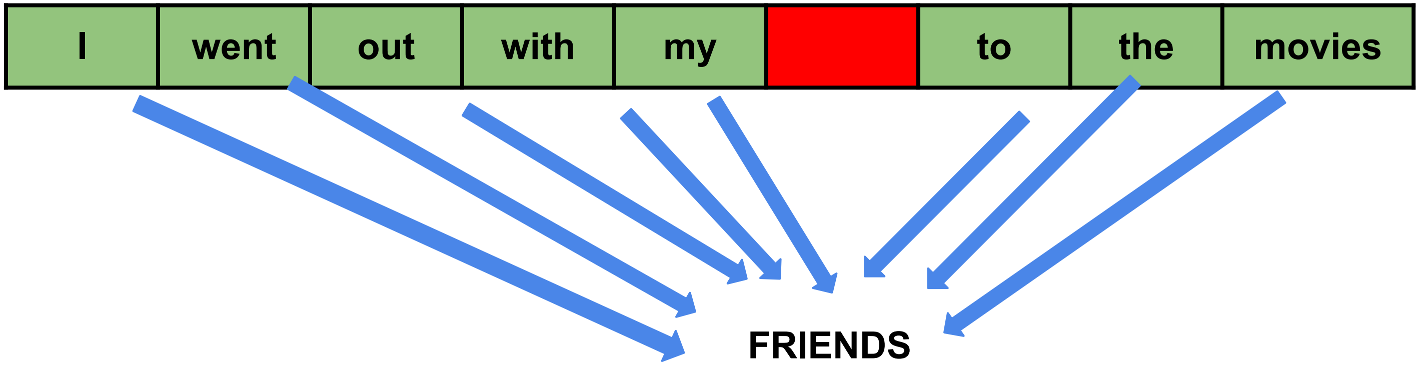
- I went out with my **BOYFRIEND** to the movies.
- I went out with my **GIRLFRIEND** to the movies.
- I went out with my **BAE** to the movies.
- I went out with my **FRIENDS** to the movies.

Words used in a similar fashion in the same context!!!!

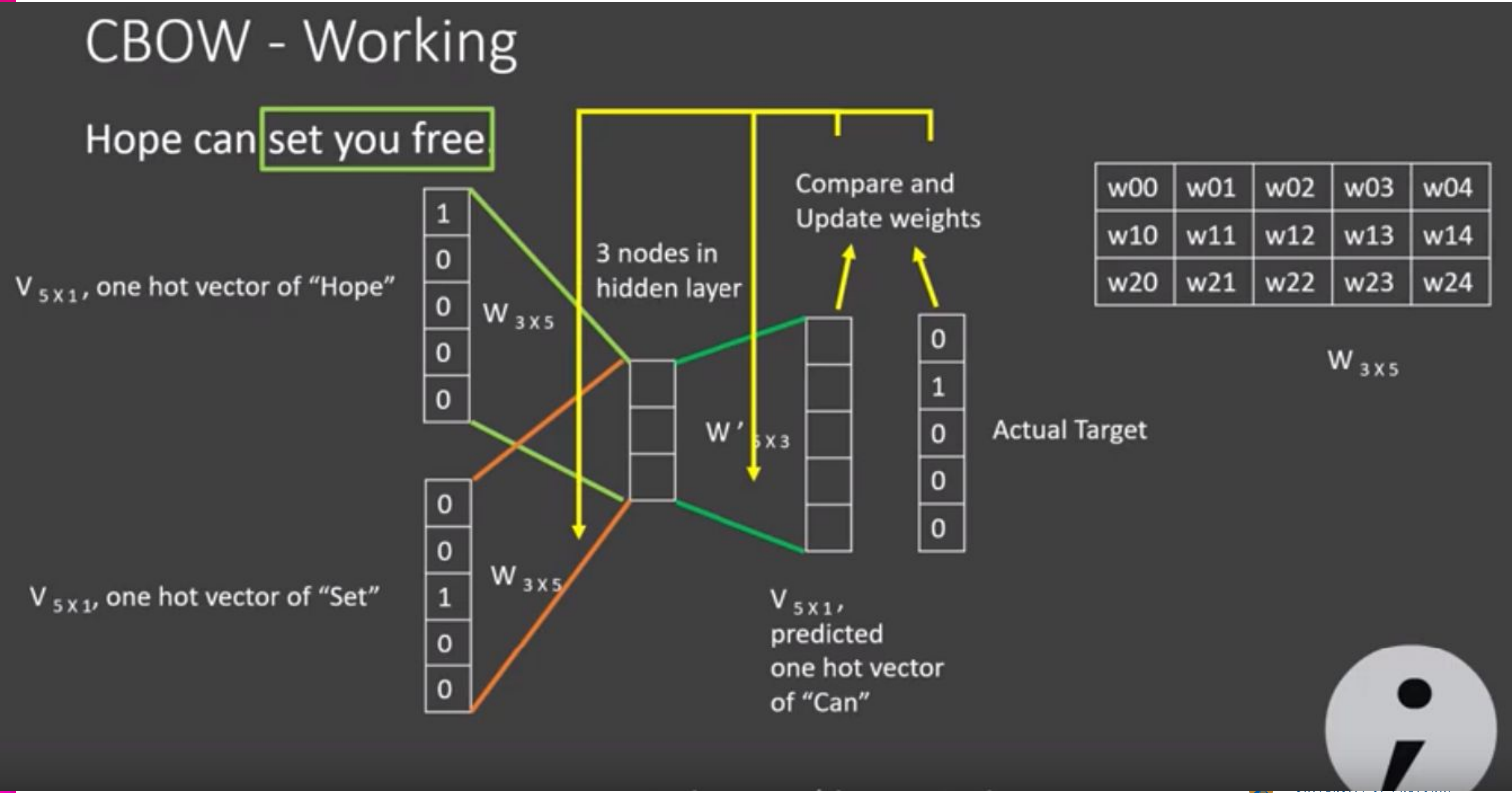
https://www.youtube.com/watch?v=UqRCEmrv1gQ

CBOW - Continuous Bag of Words

Predict word from context



https://www.youtube.com/watch?v=UqRCEmrv1gQ



CBOW - Working

Hope can set you free

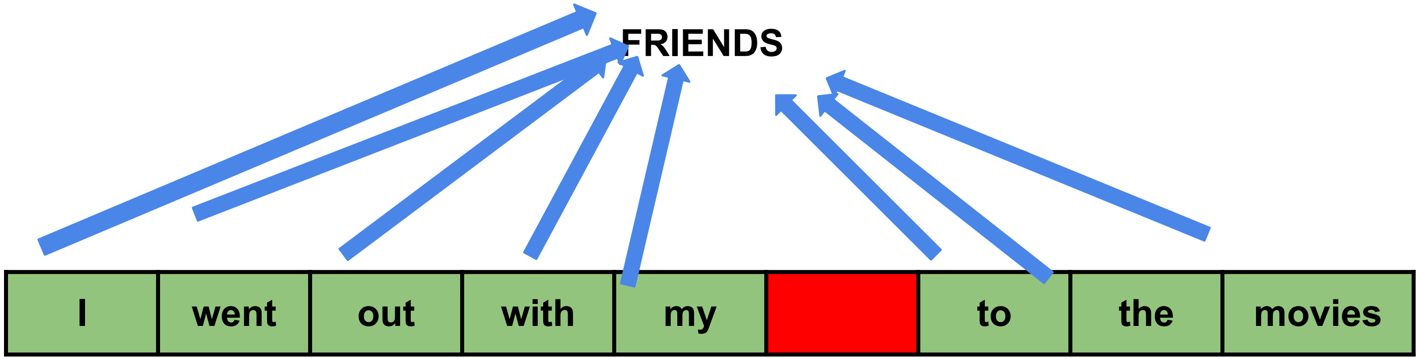
w00	w01	w02	w03	w04
w10	w11	w12	w13	w14
w20	w21	w22	w23	w24

$V_{5 \times 1}$, one hot vector of "Hope"
1
0
0
0
0
$W_{3 \times 5}$
$V_{5 \times 1}$, one hot vector of "Set"
0
0
1
0
0
$W_{3 \times 5}$
3 nodes in hidden layer
$W'_{5 \times 3}$
Compare and Update weights
0
1
0
0
0
Actual Target
$V_{5 \times 1}$, predicted one hot vector of "Can"
$W_{3 \times 5}$

https://www.youtube.com/watch?v=UqRCEmrv1gQ

Skip-GRAM: Crazy idea, but works!!

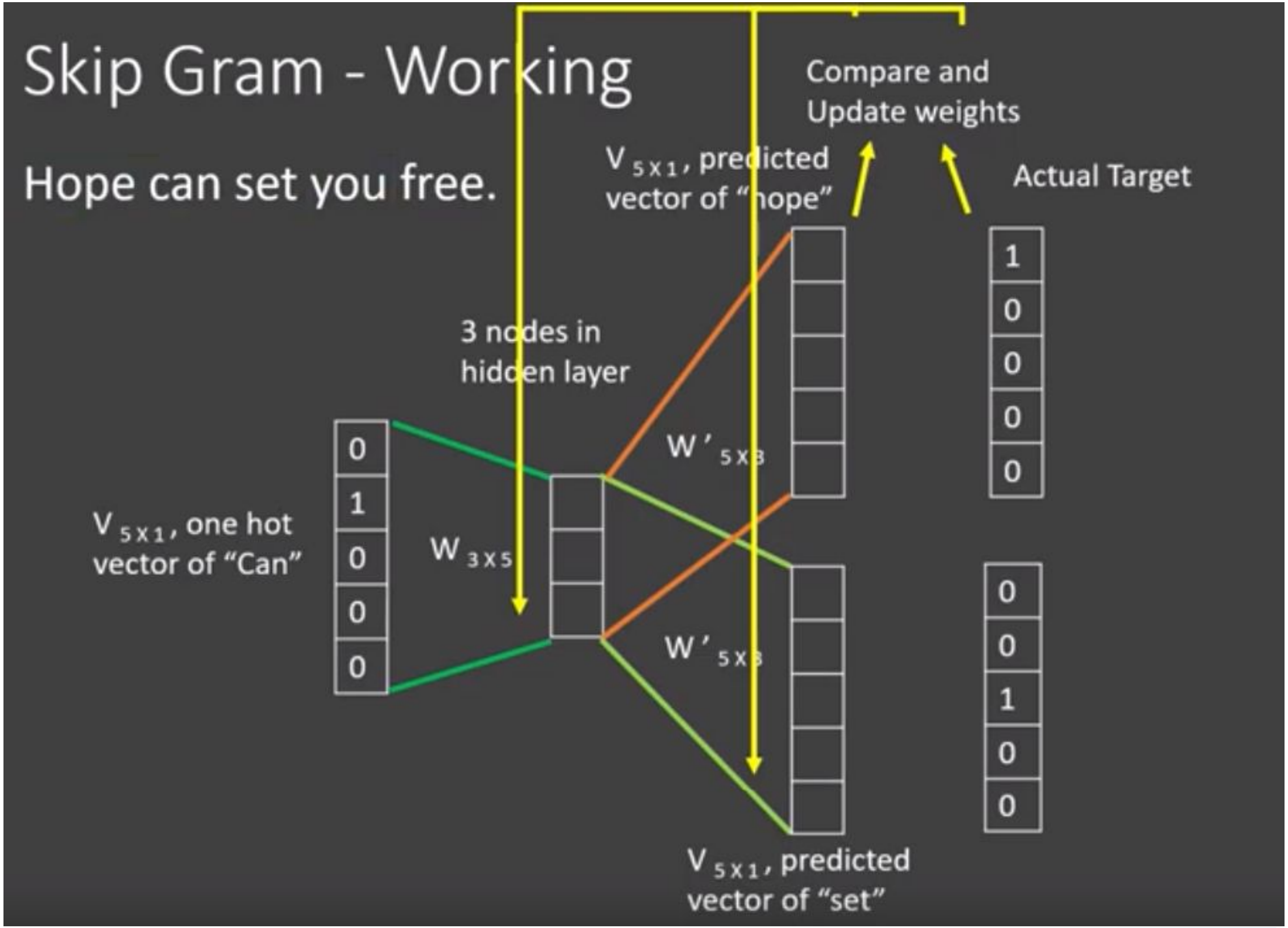
Predict context from words



FRIENDS

I	went	out	with	my	🟥	to	the	movies

https://www.youtube.com/watch?v=UqRCEmrv1gQ



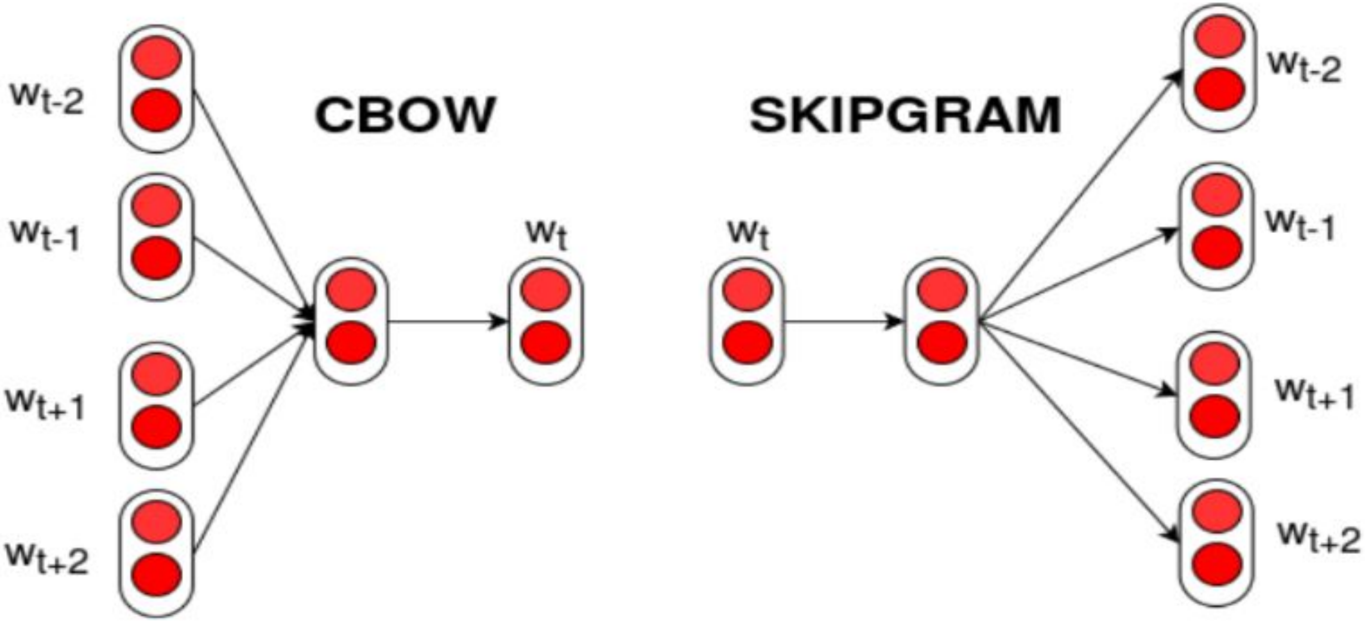
Skip Gram - Working

Hope can set you free.

Compare and Update weights
V 5x1, predicted vector of “hope”
Actual Target
1
0
0
0
0
3 nodes in hidden layer
W ’ 5x3
V 5x1, one hot vector of “Can”
0
1
0
0
0
W 3x5
W ’ 5x3
0
0
1
0
0
V 5x1, predicted vector of “set”

https://www.youtube.com/watch?v=UqRCEmrv1gQ

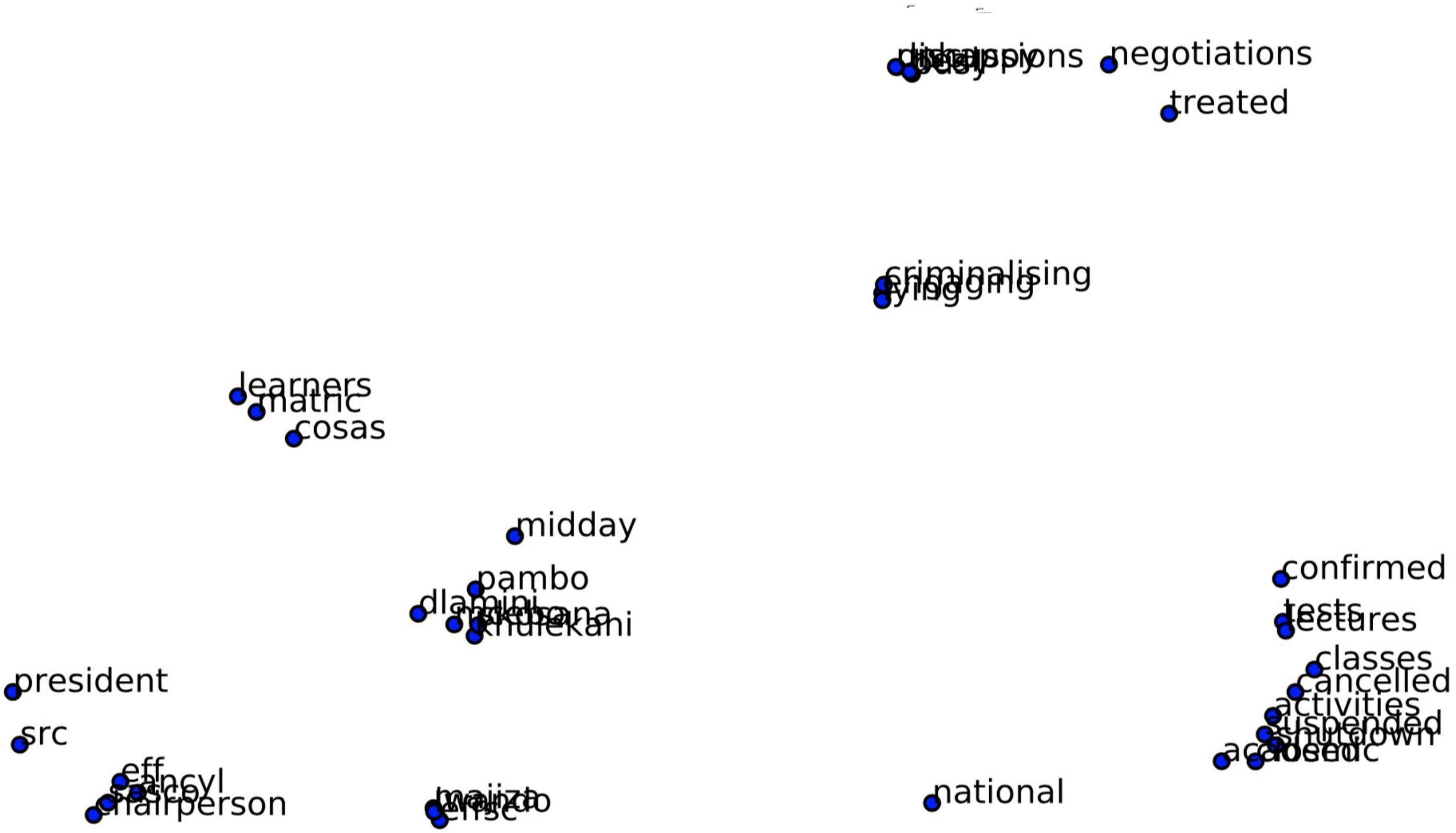
Word2Vec



$$\frac{1}{T}\sum_{t=1}^{T}\sum_{-c\leq j\leq c,j\neq 0} logp(w_t|w_{t+j})$$

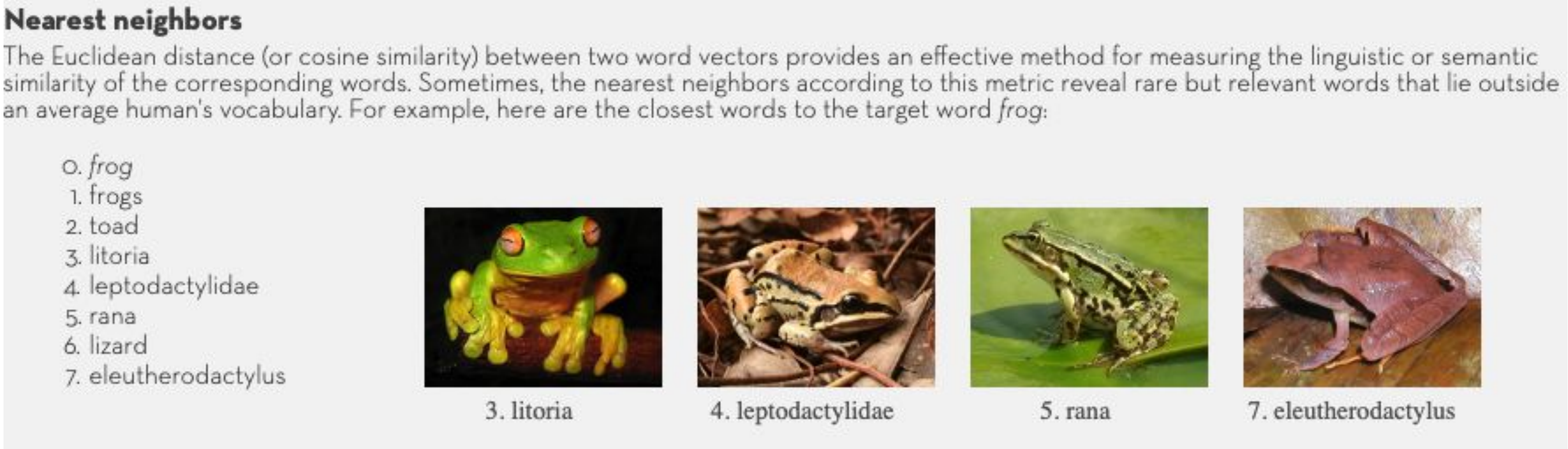
$$\frac{1}{T}\sum_{t=1}^{T}\sum_{-c\leq j\leq c,j\neq 0} logp(w_{t+j}|w_t)$$

https://arxiv.org/abs/1301.3781 [Available in Gensim to train], Pretrained as well



Fees Must Fall Twitter Data. Word2vec + TSNE

GloVe: Global Vectors for Word Representation

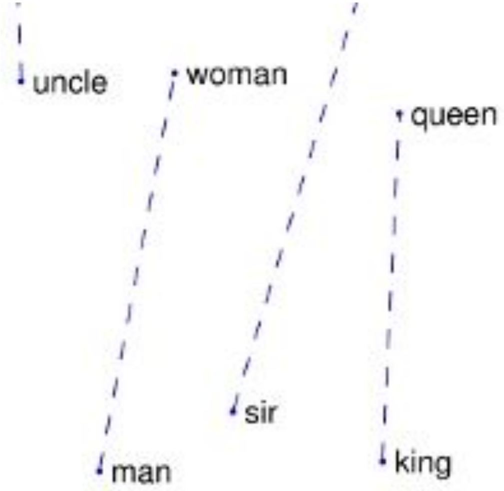


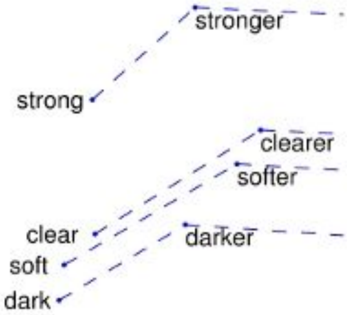
Nearest neighbors

The Euclidean distance (or cosine similarity) between two word vectors provides an effective method for measuring the linguistic or semantic similarity of the corresponding words. Sometimes, the nearest neighbors according to this metric reveal rare but relevant words that lie outside an average human's vocabulary. For example, here are the closest words to the target word *frog*:

0. *frog*
1. frogs
2. toad
3. litoria
4. leptodactylidae
5. rana
6. lizard
7. eleutherodactylus

3. litoria　4. leptodactylidae　5. rana　7. eleutherodactylus

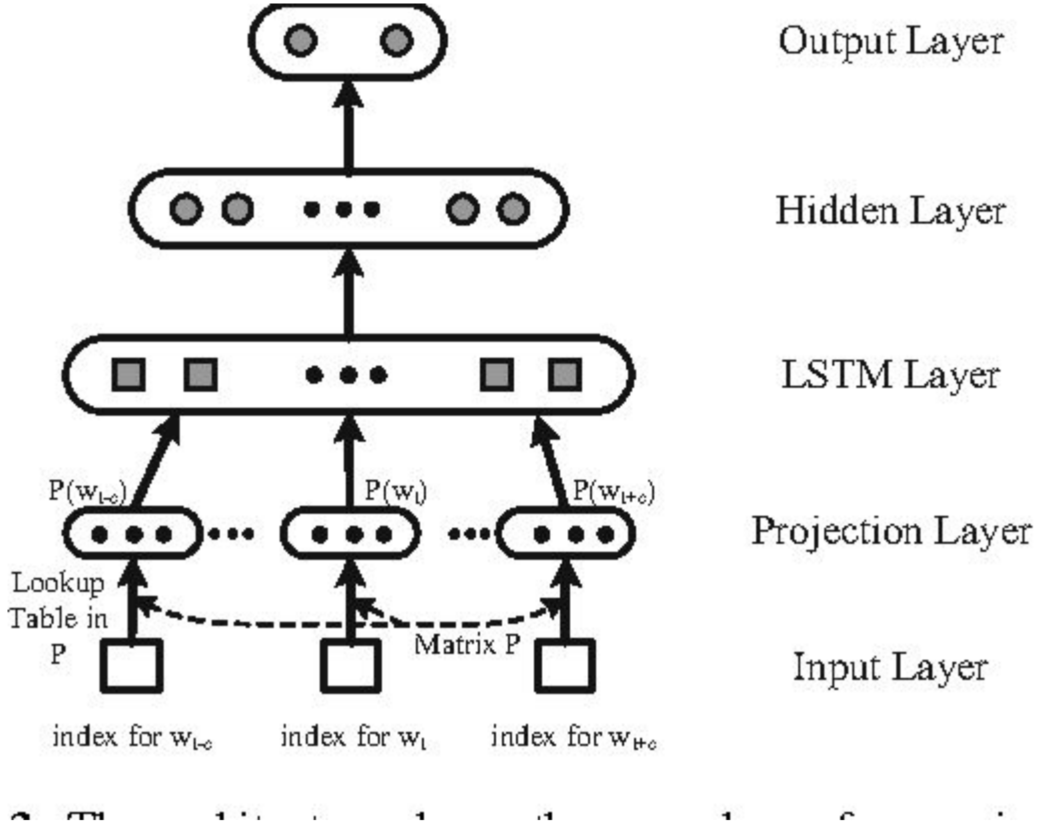




Linear Substructures: Analogies

https://nlp.stanford.edu/projects/glove/ [Pre built vectors]

Downstream tasks: Classification



Advanced: Language Models

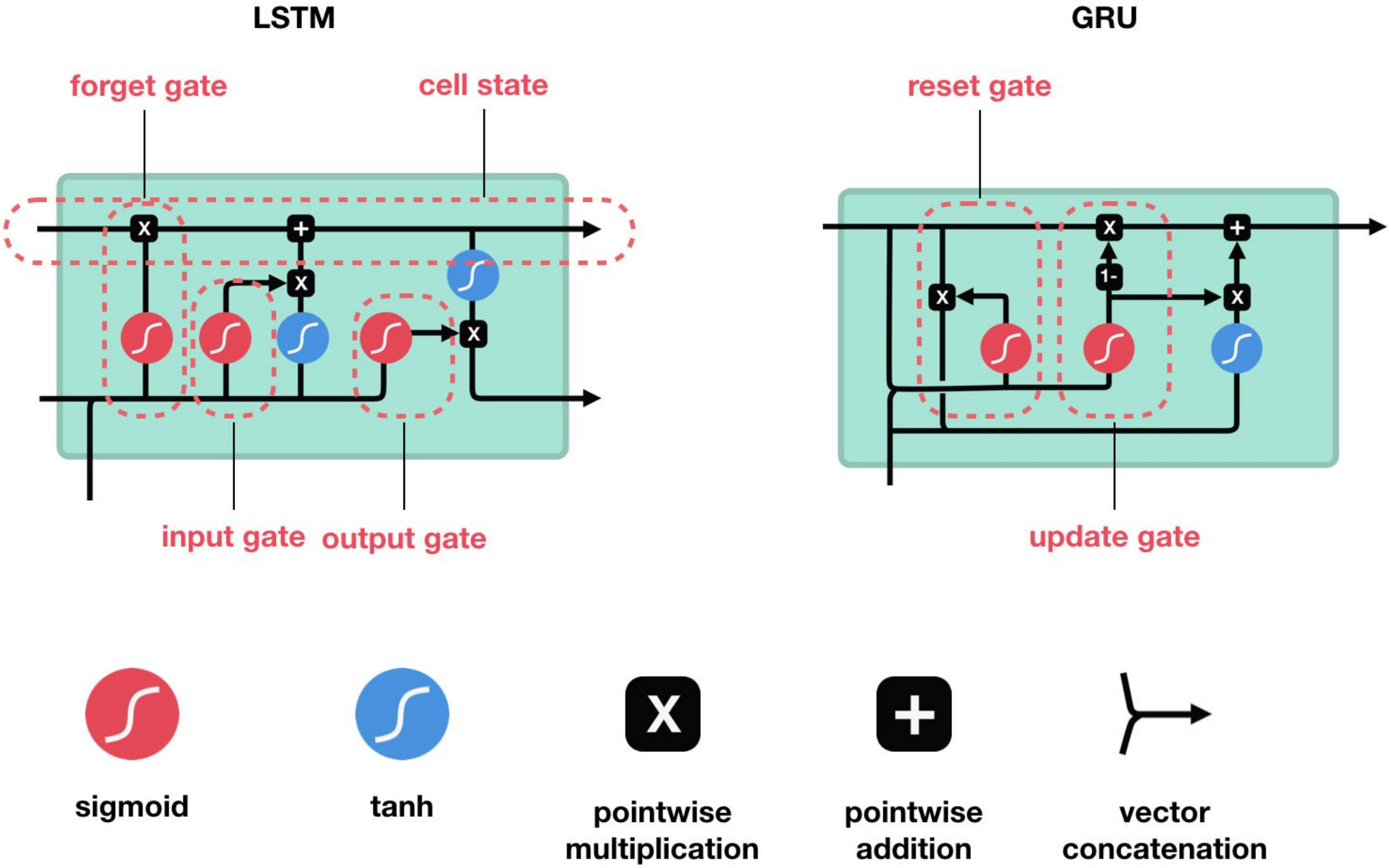
Nev
dar

The El
releas

AI researchers debate the ethics of sharing potentially harmful programs

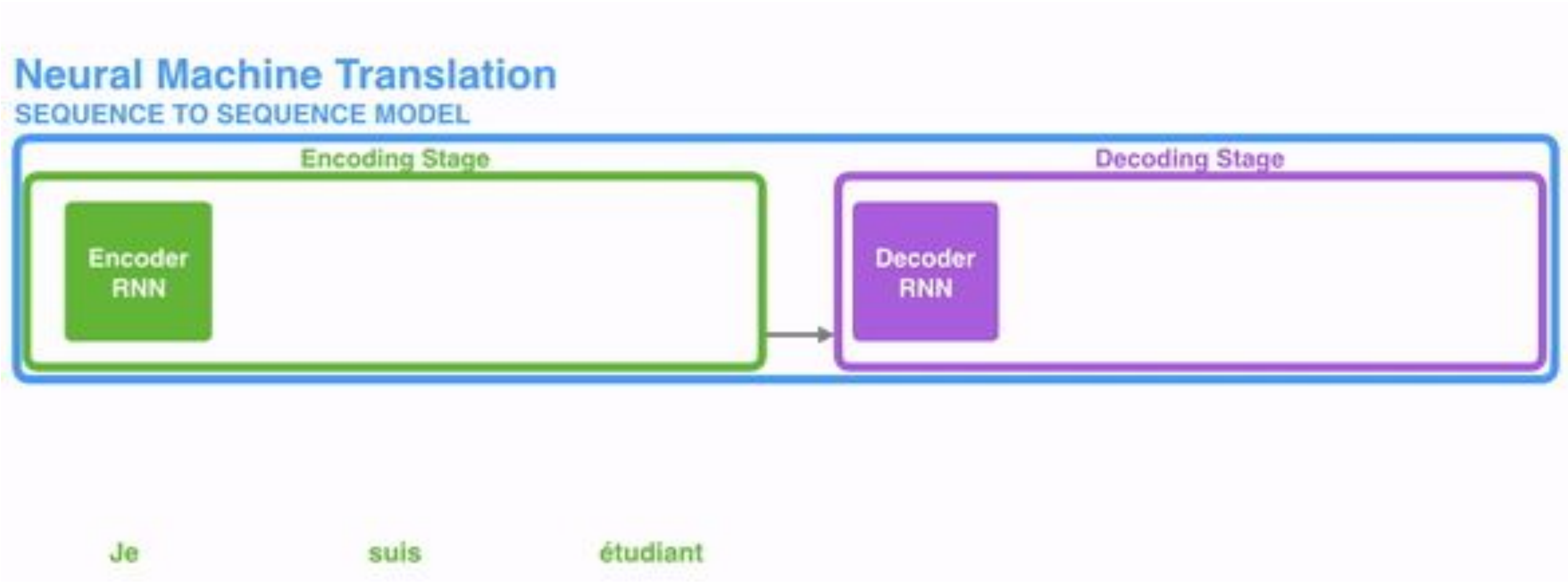
Nonprofit lab OpenAI withheld its latest research, but was criticized by others in the field

By James Vincent | Feb 21, 2019, 10:30am EST



Illustrated Guide to LSTM's and GRU's: A step by step explanation

Sequence Models



Neural Machine Translation

SEQUENCE TO SEQUENCE MODEL

Encoding Stage

Decoding Stage

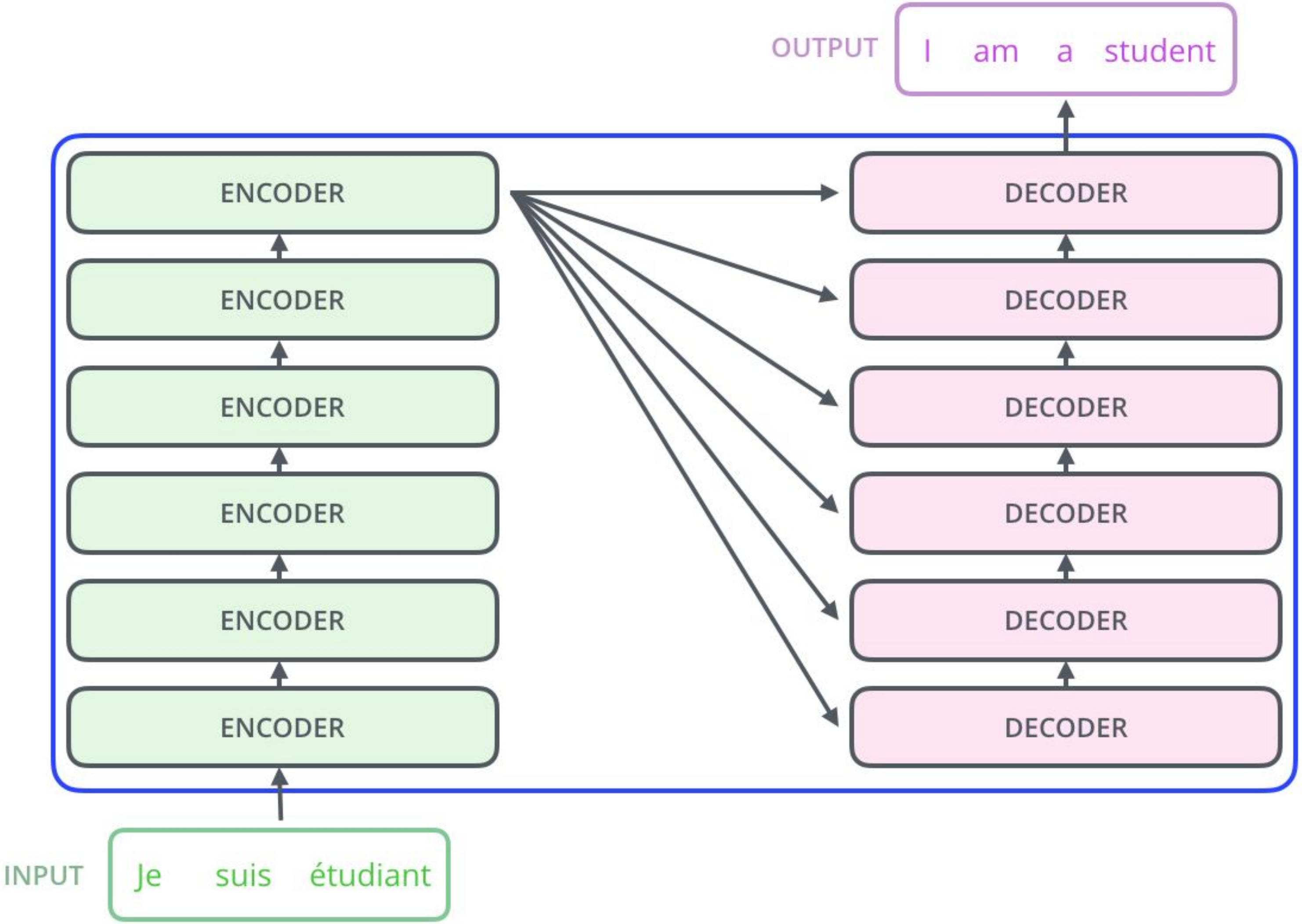
Encoder RNN

Decoder RNN

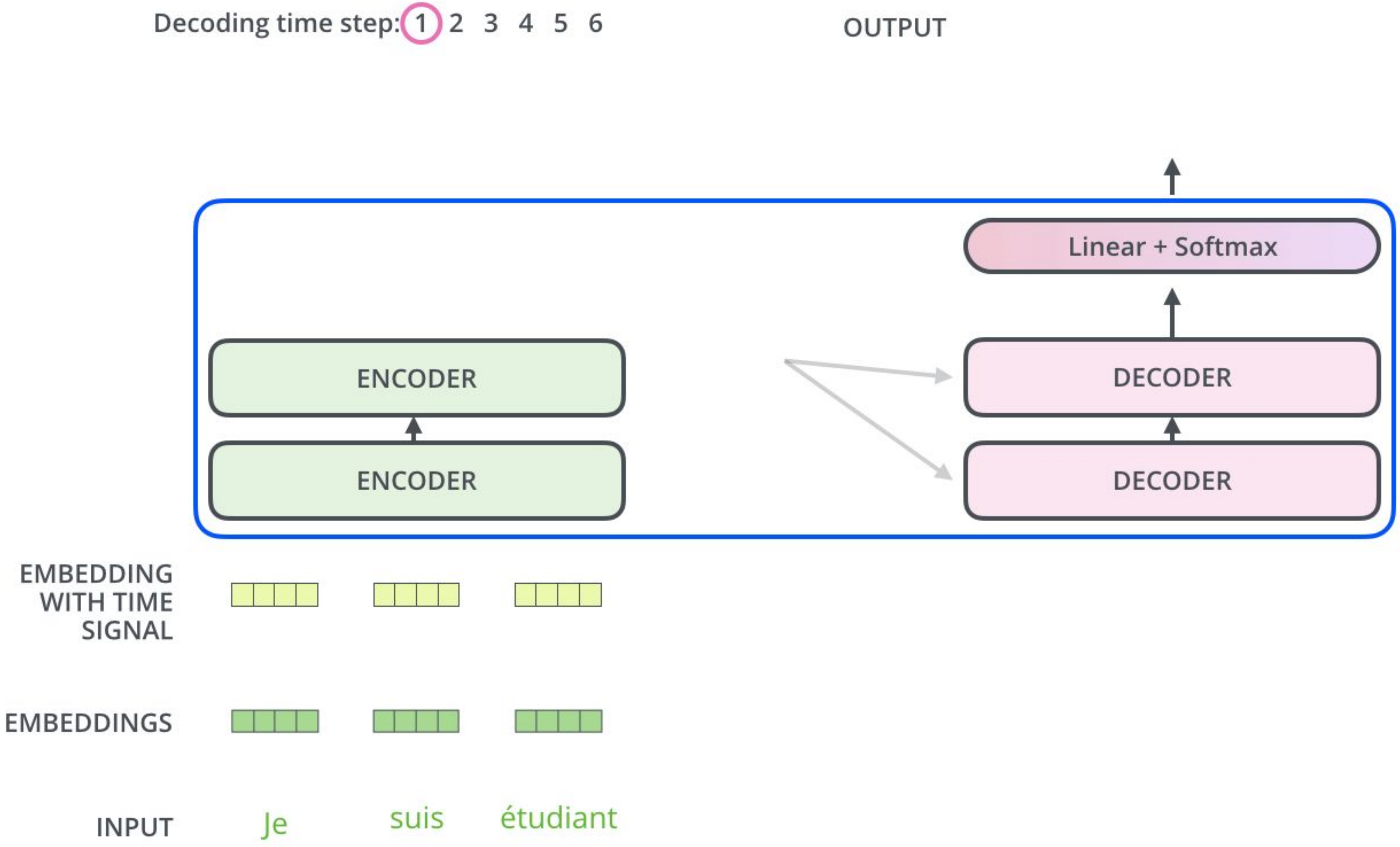
Je

suis

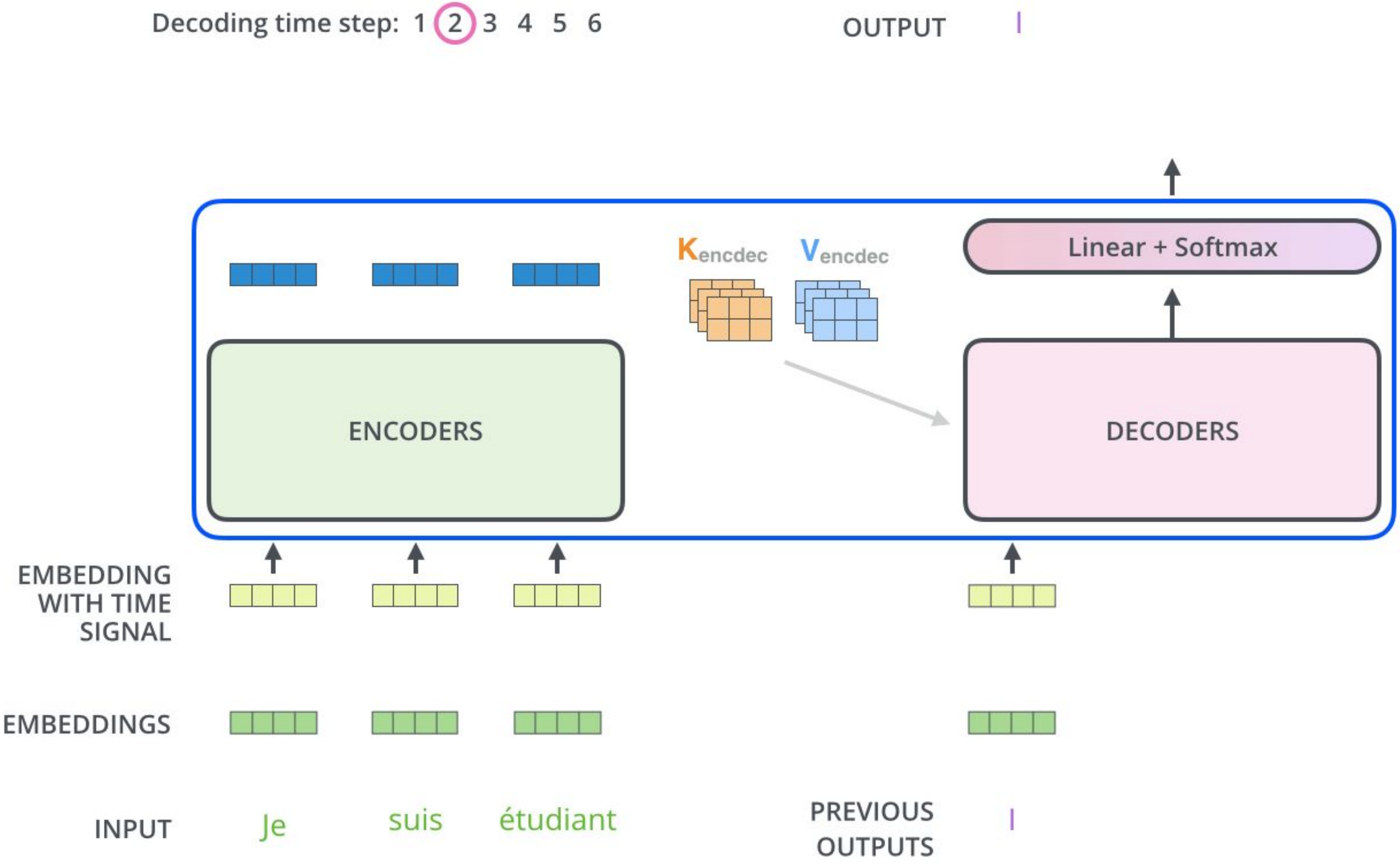
étudiant



The Illustrated Transformer: http://jalammar.github.io/illustrated-transformer/

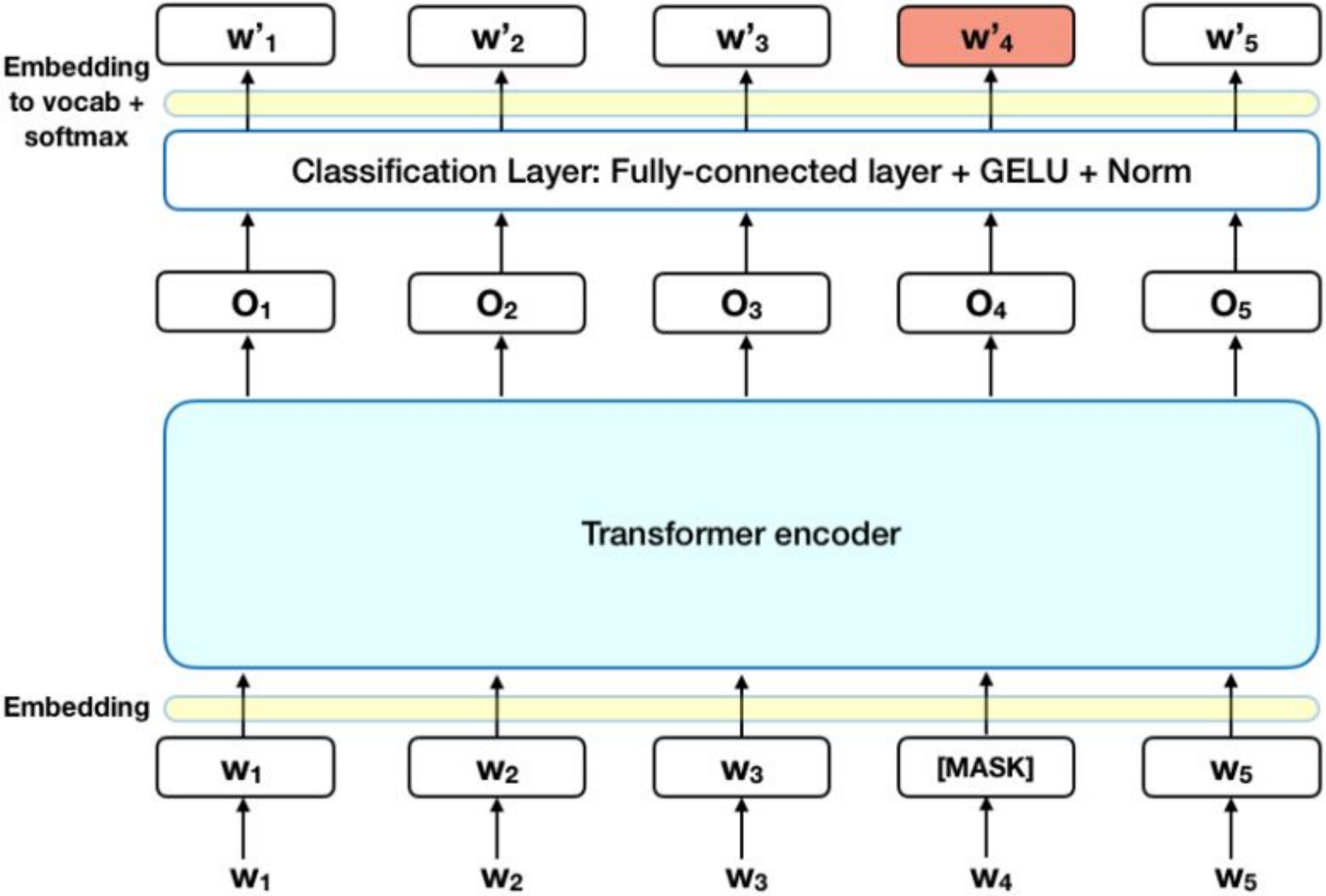


The Illustrated Transformer: http://jalammar.github.io/illustrated-transformer/



The Illustrated Transformer: http://jalammar.github.io/illustrated-transformer/

BERT Masked LM



What can you do with these LMs

Part of many ML pipelines

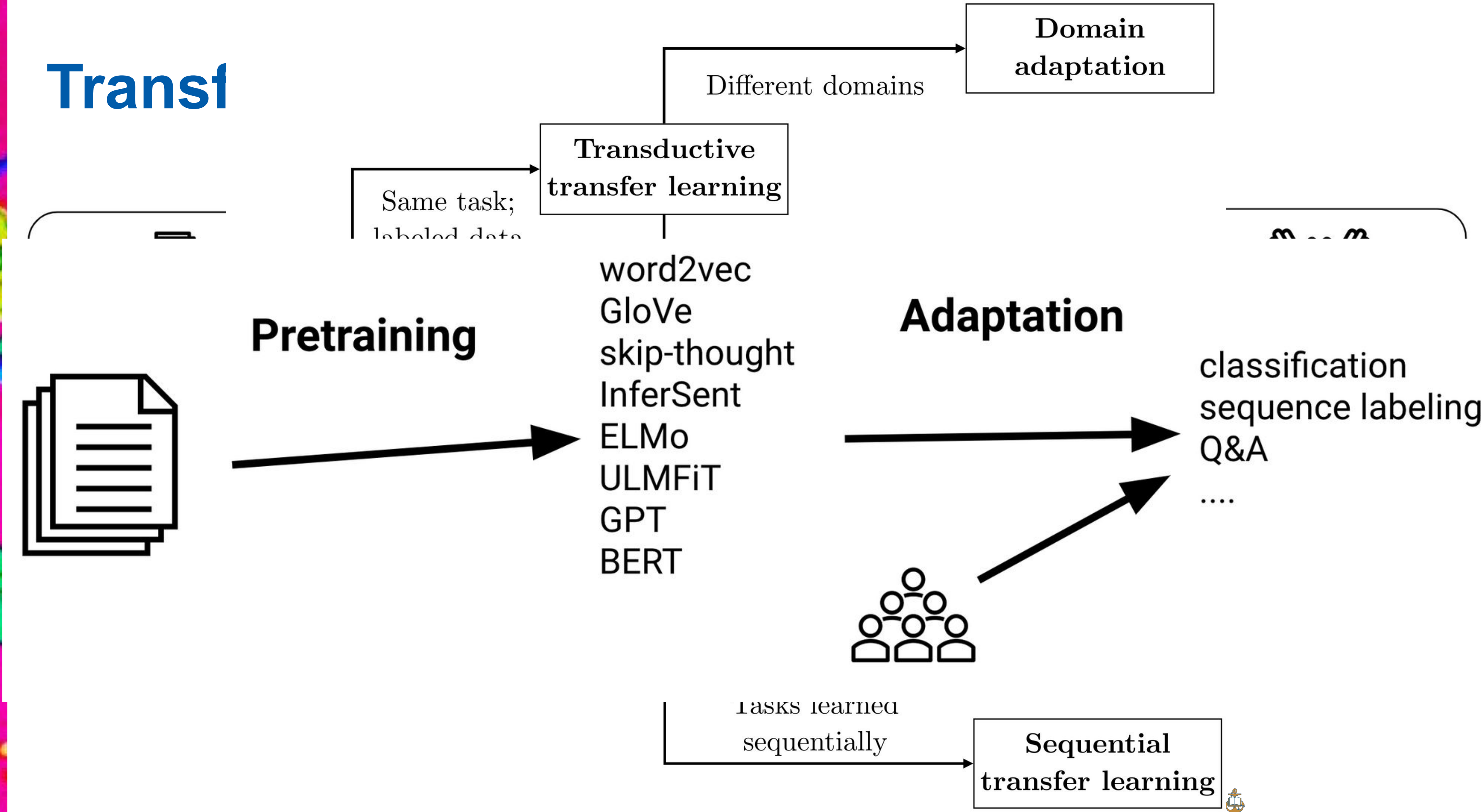
Natural Language Understanding

Question Answering

Text Generation [https://transformer.huggingface.co/]

Topic Models

Transf



A taxonomy for transfer learning in NLP (Ruder, 2019).

GTX 1070
NVIDIA.

Resources

NLP General

- https://github.com/fastai/course-nl
- Stanford Coursera NLP Slides https://web.stanford.edu/~jurafsky/NLPCourseraSlides.html
- Sebastian Ruder Newsletter [http://ruder.io/nlp-news/]
- https://nlpprogress.com

Python Libraries

- SKLearn NLP (Working With Text Data) - URL (Nice tutorial)
- spaCY: Industrial-Strength Natural Language Processing - URL
- NLTK

Thank You

Dr. Vukosi Marivate

vukosi.marivate@cs.up.ac.za

https//dsfsi.github.io

@vukosi

Data Science for Social Impact

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA