Deep Learning for Natural Language Processing

Stephen Clark et al…
DeepMind and University of Cambridge

A potted history of…….

4. Word Embeddings

Felix Hill
DeepMind

DeepMind

UNIVERSITY OF CAMBRIDGE

In ancient times, hundreds of years before the dawn of history…

In ancient times,
hundreds of years before the
dawn of ~~history~~ deep learning

DeepMind

UNIVERSITY OF CAMBRIDGE

The meaning of meaning, (before the dawn of history)

It has long been debated whether the mechanisms that underlie language are dedicated to this uniquely human capacity or whether in fact they serve more general-purpose functions. Our study provides strong evidence that language —indeed both first and second language—is learned, in specific ways, by general-purpose neurocognitive mechanisms that preexist *Homo sapiens*. The results have broad implications. They elucidate both the ontogeny (development) and phylogeny (evolution) of language. Moreover, they suggest that our substantial knowledge of the general-purpose mechanisms, from both animal and human studies, may also apply to language. The study may thus lead to a research program that can generate a wide range of predictions about this critical domain.

DeepMind

UNIVERSITY OF CAMBRIDGE

The meaning of meaning,

It has long been debated whether the mechanisms that underlie language are dedicated to this uniquely human capacity or whether in fact they serve more general-purpose functions. Our **study** provides strong evidence that language—indeed both first and second language—is learned, in specific ways, by general-purpose neurocognitive mechanisms that preexist *Homo sapiens*. The results have broad implications. They elucidate both the ontogeny (development) and phylogeny (evolution) of language. Moreover, they suggest that our substantial knowledge of the general-purpose mechanisms, from both animal and human studies, may also apply to language. The **study** may thus lead to a **research** program that can generate a wide range of predictions about this critical domain.

`language` `evidence` `functions` `learn` `program` ......

study

research

DeepMind

UNIVERSITY OF CAMBRIDGE

The meaning of meaning,

It has long been debated whether the mechanisms that underlie language are dedicated to this uniquely human capacity or whether in fact they serve more general-purpose functions. Our **study** provides strong evidence that language—indeed both first and second language—is learned, in specific ways, by general-purpose neurocognitive mechanisms that preexist *Homo sapiens*. The results have broad implications. They elucidate both the ontogeny (development) and phylogeny (evolution) of language. Moreover, they suggest that our substantial knowledge of the general-purpose mechanisms, from both animal and human studies, may also apply to language. The **study** may thus lead to a **research** program that can generate a wide range of predictions about this critical domain.

language evidence functions learn program ......

study

research

DeepMind

UNIVERSITY OF CAMBRIDGE

The meaning of meaning,

It has long been debated whether the mechanisms that underlie language are dedicated to this uniquely human capacity or whether in fact they serve more general-purpose functions. Our **study** provides strong **evidence** that language—indeed both first and second language—is learned, in specific ways, by general-purpose neurocognitive mechanisms that preexist *Homo sapiens*. The results have broad implications. They elucidate both the ontogeny (development) and phylogeny (evolution) of language. Moreover, they suggest that our substantial knowledge of the general-purpose mechanisms, from both animal and human studies, may also apply to language. The **study** may thus lead to a **research** program that can generate a wide range of predictions about this critical domain.

	language	evidence	functions	learn	program	......
study		1				
research						

DeepMind

UNIVERSITY OF CAMBRIDGE

The meaning of meaning,

It has long been debated whether the mechanisms that underlie language are dedicated to this uniquely human capacity or whether in fact they serve more general-purpose **functions**. Our **study** provides strong **evidence** that language—indeed both first and second language—is learned, in specific ways, by general-purpose neurocognitive mechanisms that preexist *Homo sapiens*. The results have broad implications. They elucidate both the ontogeny (development) and phylogeny (evolution) of language. Moreover, they suggest that our substantial knowledge of the general-purpose mechanisms, from both animal and human studies, may also apply to **language**. The **study** may thus lead to a **research** **program** that can generate a wide range of predictions about this critical domain.

	language	evidence	functions	learn	program	......
study	1	1	1	0	0	
research	0	0	0	0	1	

DeepMind

UNIVERSITY OF CAMBRIDGE

distributional semantics

`a_lot_of_different_words ———————->`

study	1 0 1 0 0 0 0 2 0 0 7 6 0 0 0 2 0 3 0 4 0 0 0 4 0 1 0 1 1 0 0 0 1 0 0 1
research	0 0 0 1 0 0 1 0 1 0 1 1 1 0 0 0 0 1 2 0 0 0 0 5 0 0 0 4 0 5 0 0 0 6 0 0 7

The meaning of “research”

DeepMind

UNIVERSITY OF CAMBRIDGE

1965: a great year for distributional semantics

http://aclweb.org/anthology/C/C65/C65-1010.pdf

http://www.aclweb.org/anthology/C65-1003

DeepMind

UNIVERSITY OF CAMBRIDGE

How can we improve this?

It has long been debated whether the mechanisms that underlie language are dedicated to this uniquely human capacity or whether in fact they serve more general-purpose **functions**. Our **study** provides strong **evidence** that language—indeed both first and second language—is learned, in specific ways, by general-purpose neurocognitive mechanisms that preexist *Homo sapiens*. The results have broad implications. They elucidate both the ontogeny (development) and phylogeny (evolution) of language. Moreover, they suggest that our substantial knowledge of the general-purpose mechanisms, from both animal and human studies, may also apply to **language**. The **study** may thus lead to a **research** program that can generate a wide range of predictions about this critical domain.

	language	evidence	functions	learn	program	......
study	1	1	1	0	0	
research	0	0	0	0	1	

DeepMind

UNIVERSITY OF CAMBRIDGE

How can we improve this?

Change the size of this thing

It has long been debated whether the mechanisms that underlie language are dedicated to this uniquely human capacity or whether in fact they serve more general-purpose **functions**. Our **study** provides strong **evidence** that language—indeed both first and second language—is learned, in specific ways, by general-purpose neurocognitive mechanisms that preexist *Homo sapiens*. The results have broad implications. They elucidate both the ontogeny (development) and phylogeny (evolution) of language. Moreover, they suggest that our substantial knowledge of the general-purpose mechanisms, from both animal and human studies, may also apply to **language**. The **study** may thus lead to a **research** program that can generate a wide range of predictions about this critical domain.

	language	evidence	functions	learn	program	......
study	1	1	1	0	0	
research	0	0	0	0	1	

DeepMind

UNIVERSITY OF CAMBRIDGE

How can we improve this?

Use a parser to determine what "close" means

It has long been debated whether the mechanisms that underlie language are dedicated to this uniquely human capacity or whether in fact they serve more general-purpose **functions**. Our **study** provides strong **evidence** that language—indeed both first and second language—is learned, in specific ways, by general-purpose neurocognitive mechanisms that preexist *Homo sapiens*. The results have broad implications. They elucidate both the ontogeny (development) and phylogeny (evolution) of language. Moreover, they suggest that our substantial knowledge of the general-purpose mechanisms, from both animal and human studies, may also apply to **language**. The **study** may thus lead to a **research** program that can generate a wide range of predictions about this critical domain.

	language	evidence	functions	learn	program	......
study	1	1	1	0	0	
research	0	0	0	0	1	

DeepMind

UNIVERSITY OF CAMBRIDGE

How can we improve this?

It has long been debated whether the mechanisms that underlie language are dedicated to this uniquely human capacity or whether in fact they serve more general-purpose **functions**. Our **study** provides strong **evidence** that language—indeed both first and second language—is learned, in specific ways, by general-purpose neurocognitive mechanisms that preexist *Homo sapiens*. The results have broad implications. They elucidate both the ontogeny (development) and phylogeny (evolution) of language. Moreover, they suggest that our substantial knowledge of the general-purpose mechanisms, from both animal and human studies, may also apply to **language**. The **study** may thus lead to a **research** program that can generate a wide range of predictions about this critical domain.

put only certain words here…. →

	language	evidence	functions	learn	program	……
study	1	1	1	0	0	
research	0	0	0	0	1	

DeepMind

UNIVERSITY OF CAMBRIDGE

How can we improve this?

`a_lot_of_different_words ————————->`

study	1 0 1 0 0 0 0 2 0 0 7 6 0 0 0 2 0 3 0 4 0 0 0 4 0 1 0 1 1 0 0 0 1 0 0 1
research	0 0 0 1 0 0 1 0 1 0 1 1 1 0 0 0 0 1 2 0 0 0 0 5 0 0 0 4 0 5 0 0 0 6 0 0 7

Do something fancy to these numbers…

cf: Sparck-Jones and *tf-idf*
https://en.wikipedia.org/wiki/Tf%E2%80%93idf

How can we improve this?

a lot of different words ———————-->

study	1 0 1 0 0 0 0 2 0 0 7 6 0 0 0 2 0 3 0 4 0 0 0 4 0 1 0 1 1 0 0 0 1 0 0 1
research	0 0 0 1 0 0 1 0 1 0 1 1 1 0 0 0 0 1 2 0 0 0 0 5 0 0 0 4 0 5 0 0 0 6 0 0 7
a	
lot	
of	
other	a lot of numbers……….
words	
too	
yay	

matrix factorisation

a 'better' meaning of "research"

no obvious link to words…

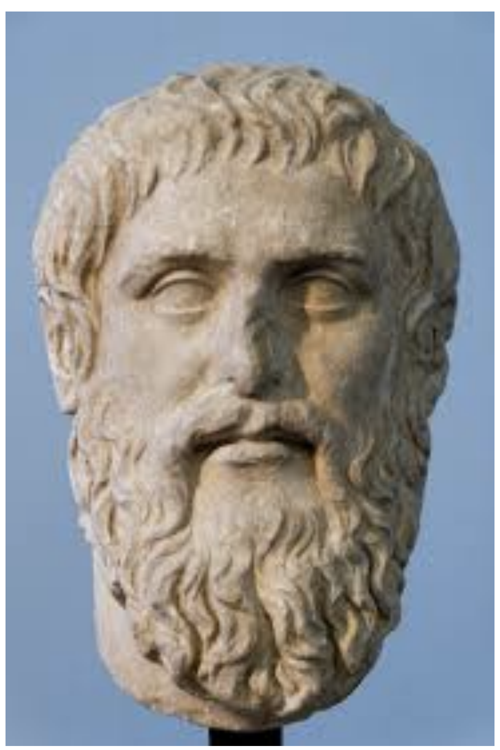
study	4.12	3.81	-2.17	8.13	7.23
researc					
a					
lot					
of					
other	not so many numbers (zeros)				
words					
too					
yay					

DeepMind

UNIVERSITY OF CAMBRIDGE

A solution to Plato's problem

Plato



Latent Semantic Analysis

Landauer & Dumais (1997)

A Solution to Plato's Problem:
The Latent Semantic Analysis Theory of Acquisition, Induction and Representation of Knowledge

Thomas K. Landauer
Department of Psychology
University of Colorado, Boulder
Boulder, CO 80309

Susan T. Dumais
Information Sciences Research
Bellcore
Morristown, New Jersey 07960

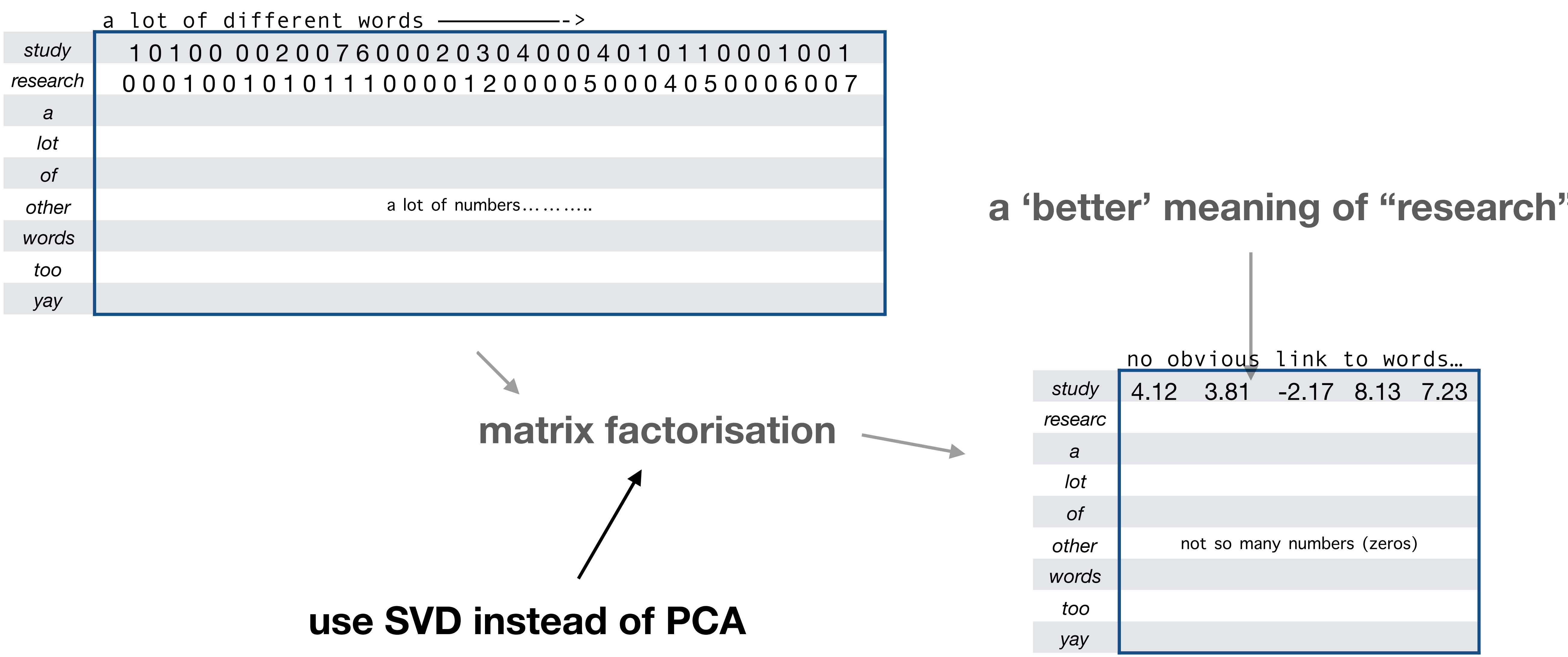
Abstract

How do people know as much as they do with as little information as they get? The problem takes many forms; learning vocabulary from text is an especially dramatic and convenient case for research. A new general theory of acquired similarity and knowledge representation, Latent Semantic Analysis (LSA), is presented and used to successfully simulate such learning and several other psycholinguistic phenomena. By inducing global knowledge indirectly from local co-occurrence data in a large body of representative text, LSA acquired knowledge about the full vocabulary of English at a comparable rate to school-children. LSA uses no prior linguistic or perceptual similarity knowledge; it is based solely on a general mathematical learning method that achieves powerful inductive effects by extracting the right number of dimensions (e.g., 300) to represent objects and contexts. Relations to other theories, phenomena, and problems are sketched.

DeepMind

UNIVERSITY OF CAMBRIDGE

A solution to Plato’s problem



DeepMind

UNIVERSITY OF CAMBRIDGE

TOEFL questions

"The wording of vocabulary questions is almost always "The word '_____' in the passage is closest in meaning to" followed by four answer choices. The word or phrase in question might be a relatively common word you're familiar with already, or it might be a more technical phrase. In either case, it's important to pay attention to the *context* the word is used in, as this may impact your answer."

The meaning of the word "technical" in the passage is closest in meaning to

A) natural
B) specialized
C) old
D) foreign

DeepMind

UNIVERSITY OF CAMBRIDGE

Learning the meaning of words 1965-~2010

Journal of Artificial Intelligence Research 37 (2010) 141-188 Submitted 10/09; published 02/10

From Frequency to Meaning: Vector Space Models of Semantics

Peter D. Turney PETER.TURNEY@NRC-CNRC.GC.CA
National Research Council Canada
Ottawa, Ontario, Canada, K1A 0R6

Patrick Pantel ME@PATRICKPANTEL.COM
Yahoo! Labs
Sunnyvale, CA, 94089, USA

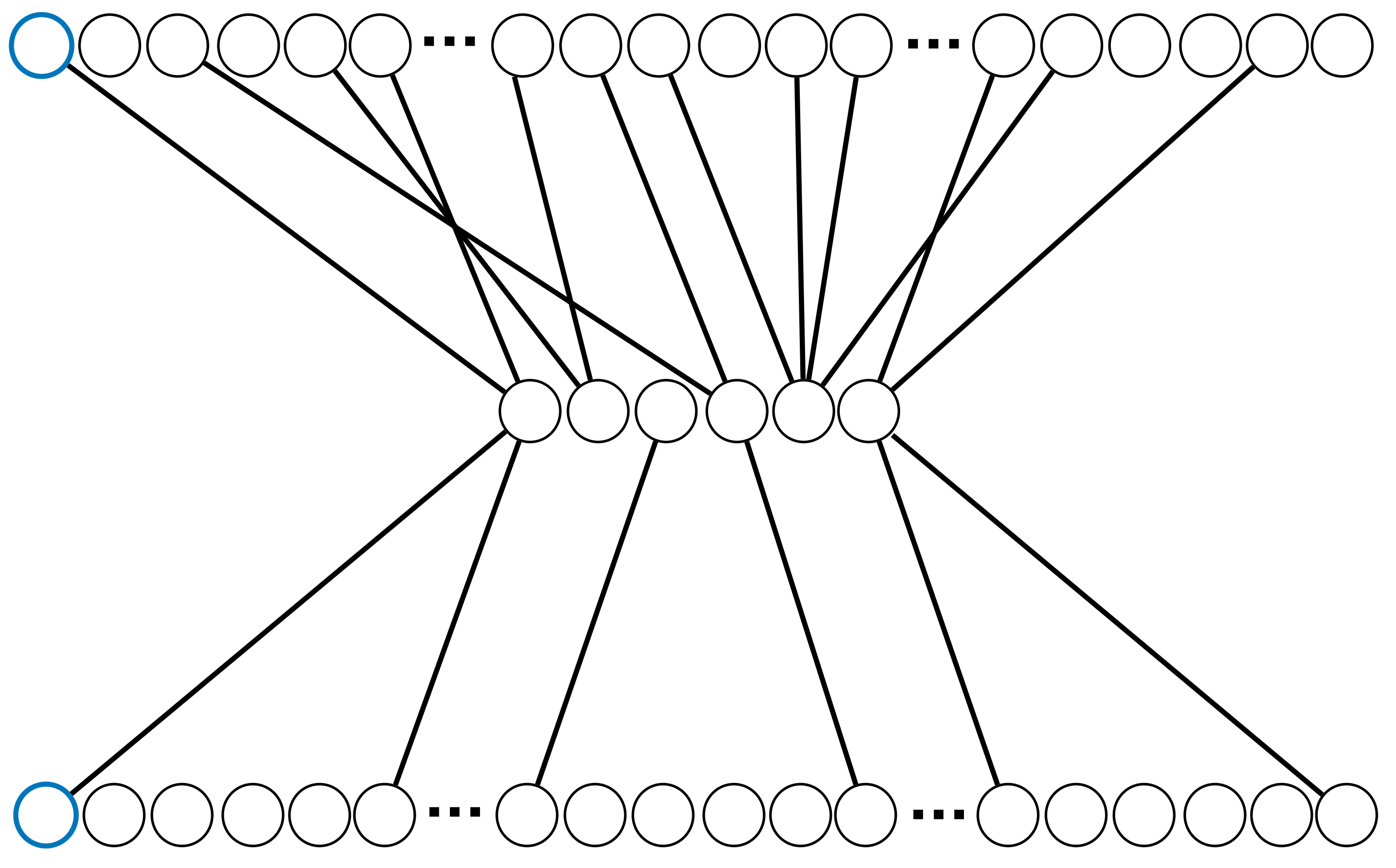
Abstract

Computers understand very little of the meaning of human language. This profoundly limits our ability to give instructions to computers, the ability of computers to explain their actions to us, and the ability of computers to analyse and process text. Vector space models (VSMs) of semantics are beginning to address these limits. This paper surveys the use of VSMs for semantic processing of text. We organize the literature on VSMs according to the structure of the matrix in a VSM. There are currently three broad classes of VSMs, based on term–document, word–context, and pair–pattern matrices, yielding three classes of applications. We survey a broad range of applications in these three categories and we take a detailed look at a specific open source project in each category. Our goal in this survey is to show the breadth of applications of VSMs for semantics, to provide a new perspective on VSMs for those who are already familiar with the area, and to provide pointers into the literature for those who are less familiar with the field.

1. Introduction

One of the biggest obstacles to making full use of the power of computers is that they currently understand very little of the meaning of human language. Recent progress in

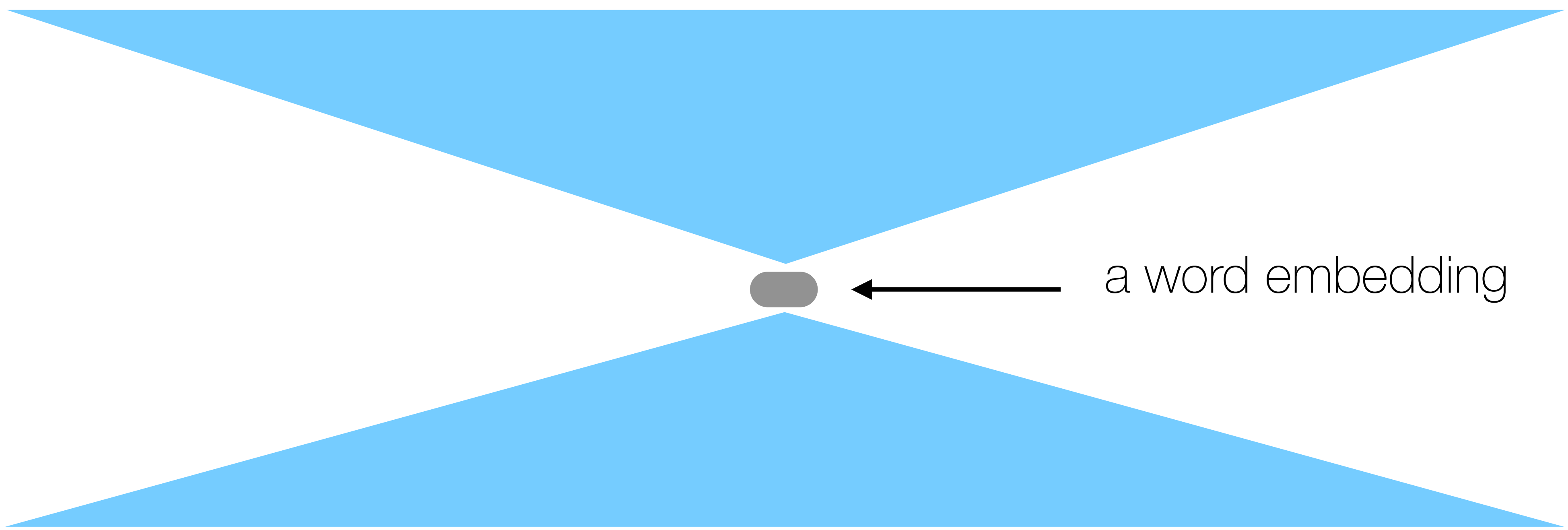
Word2Vec



DeepMind

UNIVERSITY OF CAMBRIDGE

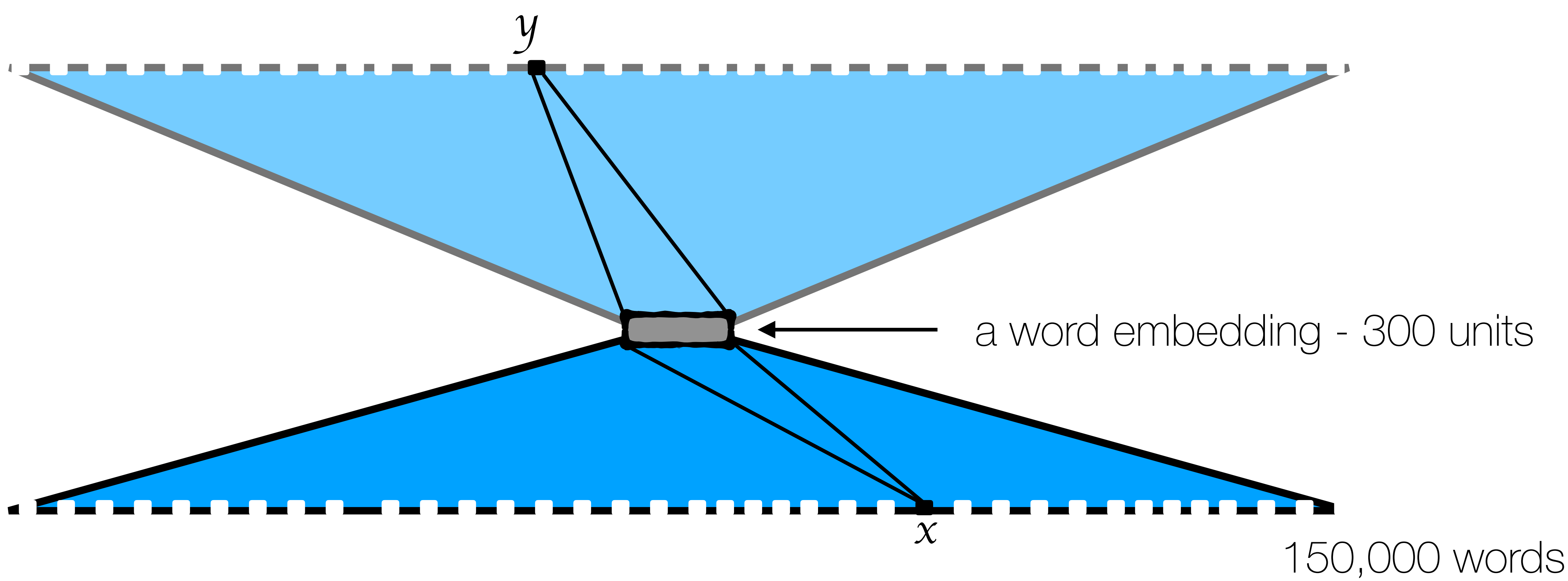
Word2Vec is a wonky MLP....



DeepMind

UNIVERSITY OF CAMBRIDGE

Word2Vec



DeepMind

UNIVERSITY OF CAMBRIDGE

Word2Vec

output word

input word

Where do we get the input words and the outputs words?

DeepMind

UNIVERSITY OF CAMBRIDGE

Word2Vec

output word

input word

Skipgram

Where do we get the input words and the outputs words?

DeepMind

UNIVERSITY OF CAMBRIDGE

Word2Vec

output word

input word

Skipgram

Where do we get the input words and the outputs words?

Word2Vec

output word

input word

Skipgram

Where do we get the input words and the outputs words?

Word2Vec

output word

input word

Skipgram

Where do we get the input words and the outputs words?

Word2Vec

output word

input word

Skipgram

Where do we get the input words and the outputs words?

DeepMind

UNIVERSITY OF CAMBRIDGE

Word2Vec

output word

input word

Skipgram

Where do we get the input words and the outputs words?

DeepMind

UNIVERSITY OF CAMBRIDGE

Word2Vec

output word

input word

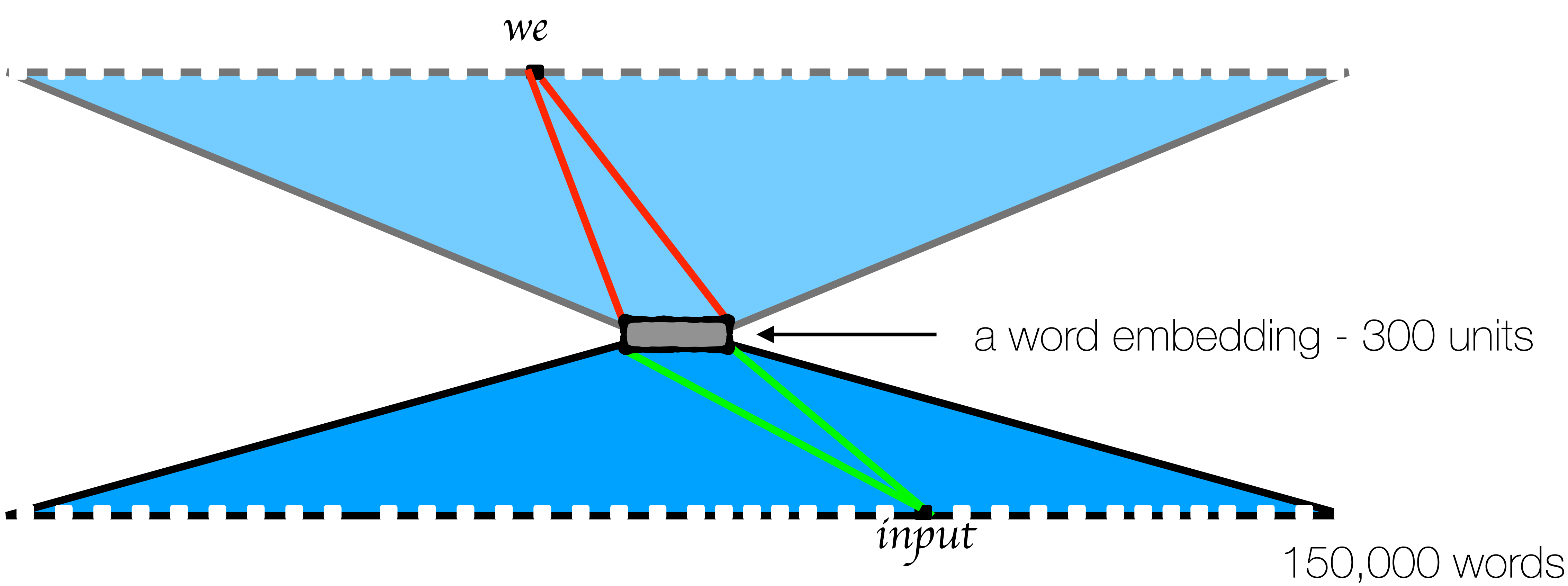
CBOW

Where do we get the input words and the outputs words?

DeepMind

UNIVERSITY OF CAMBRIDGE

Skipgram



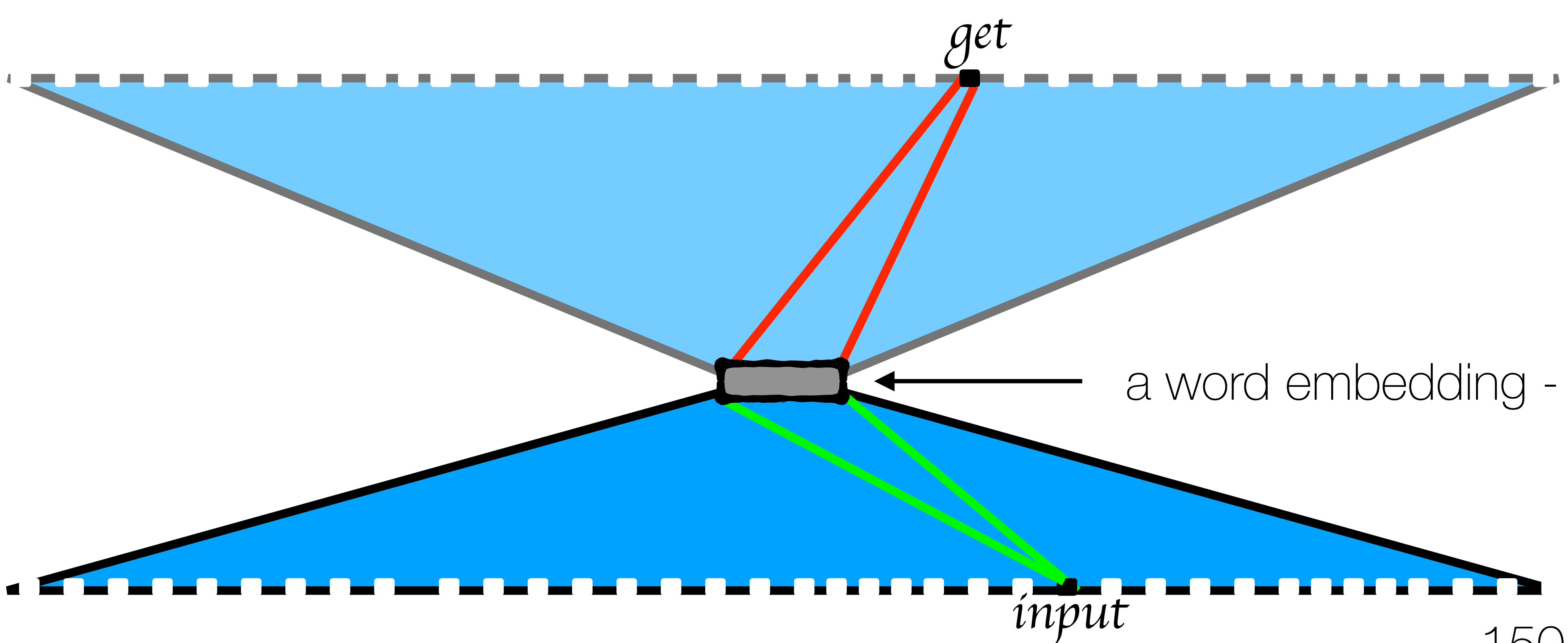
Where do we get the input words and the outputs words?

window size = 3

DeepMind

UNIVERSITY OF CAMBRIDGE

Skipgram

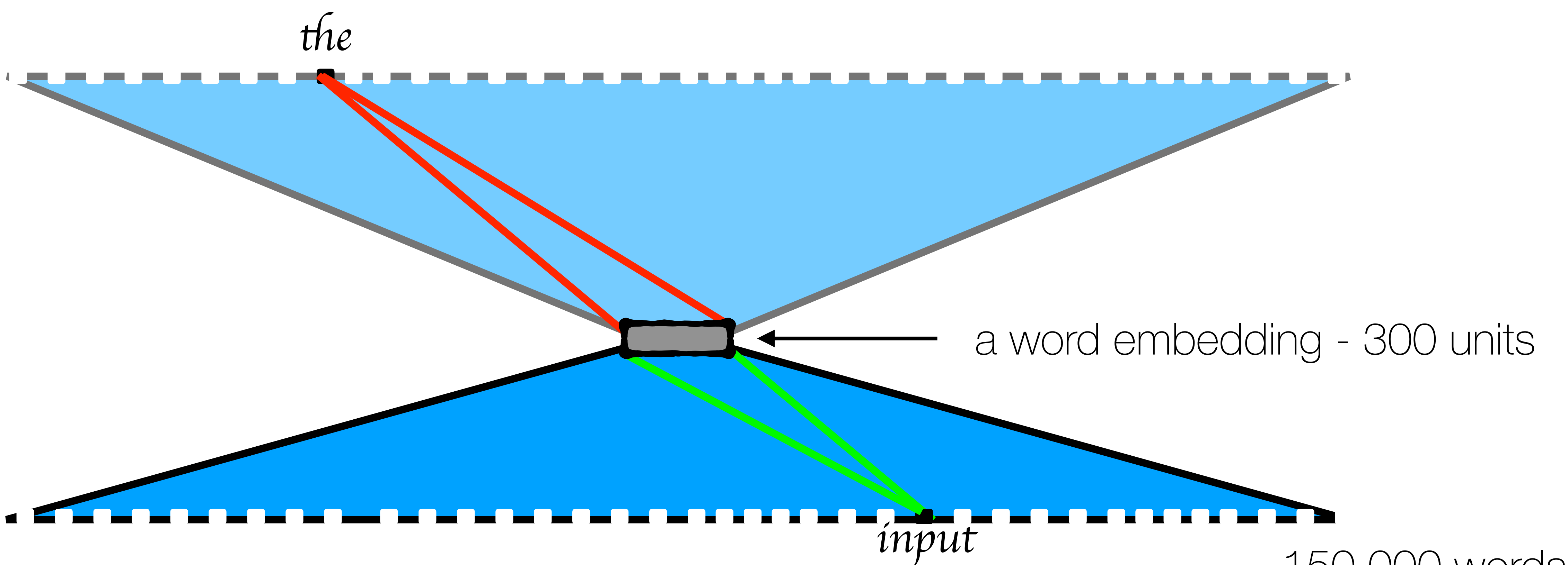


Where do we get the input words and the outputs words?

DeepMind

UNIVERSITY OF CAMBRIDGE

Skipgram

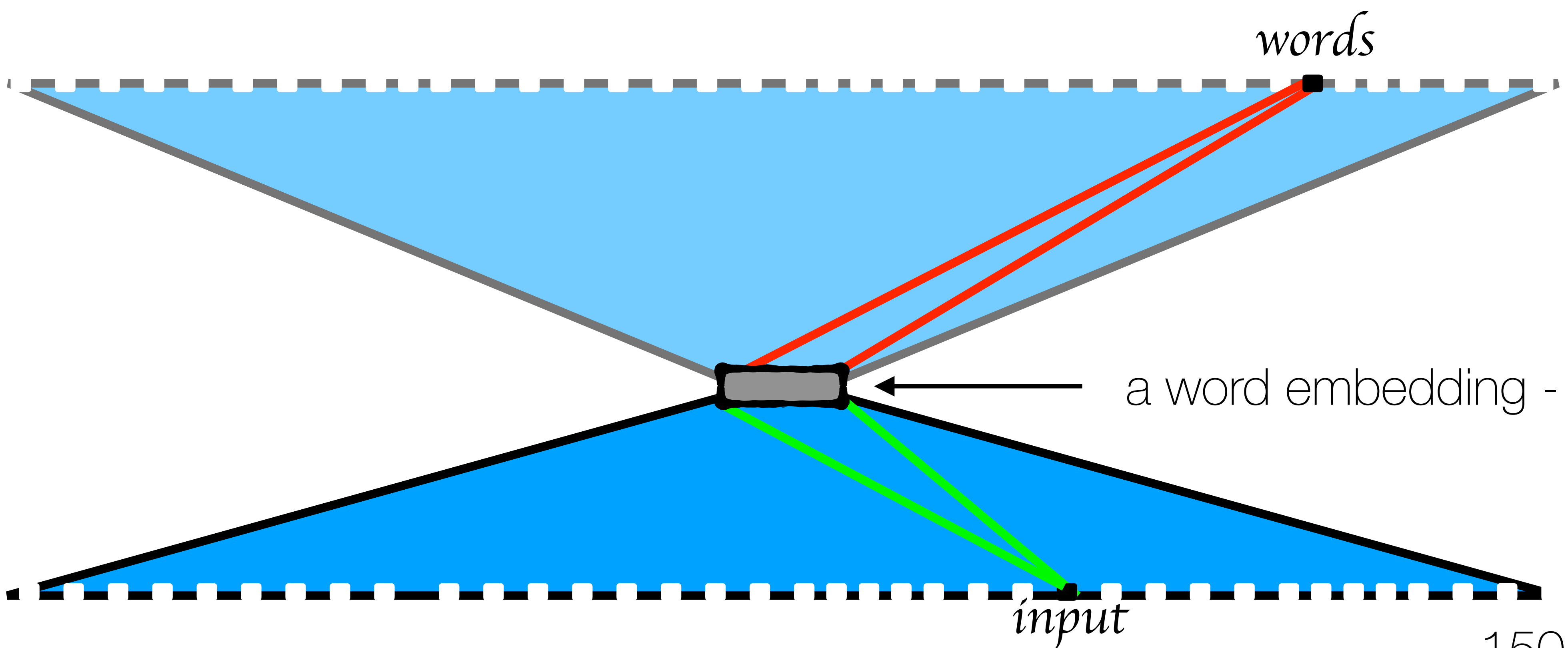


Where do we get the input words and the outputs words?

DeepMind

UNIVERSITY OF CAMBRIDGE

Skipgram

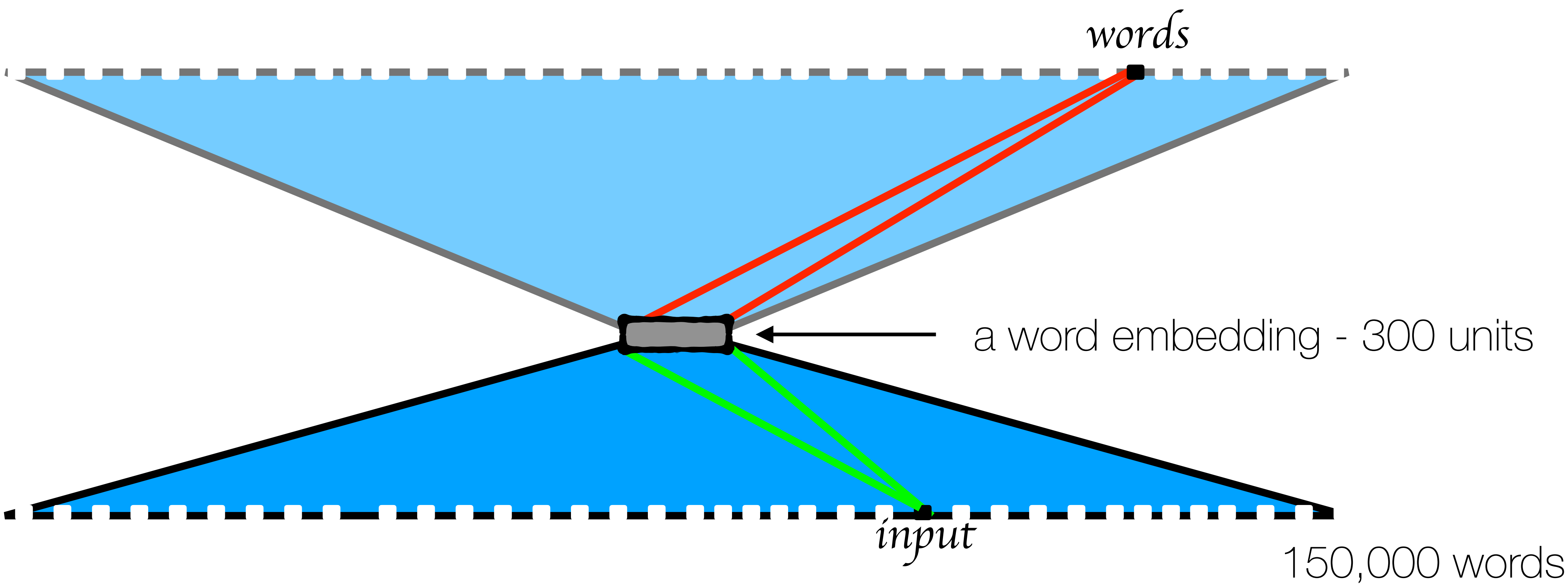


Where do we get the input words and the outputs words?

DeepMind

UNIVERSITY OF CAMBRIDGE

Skipgram



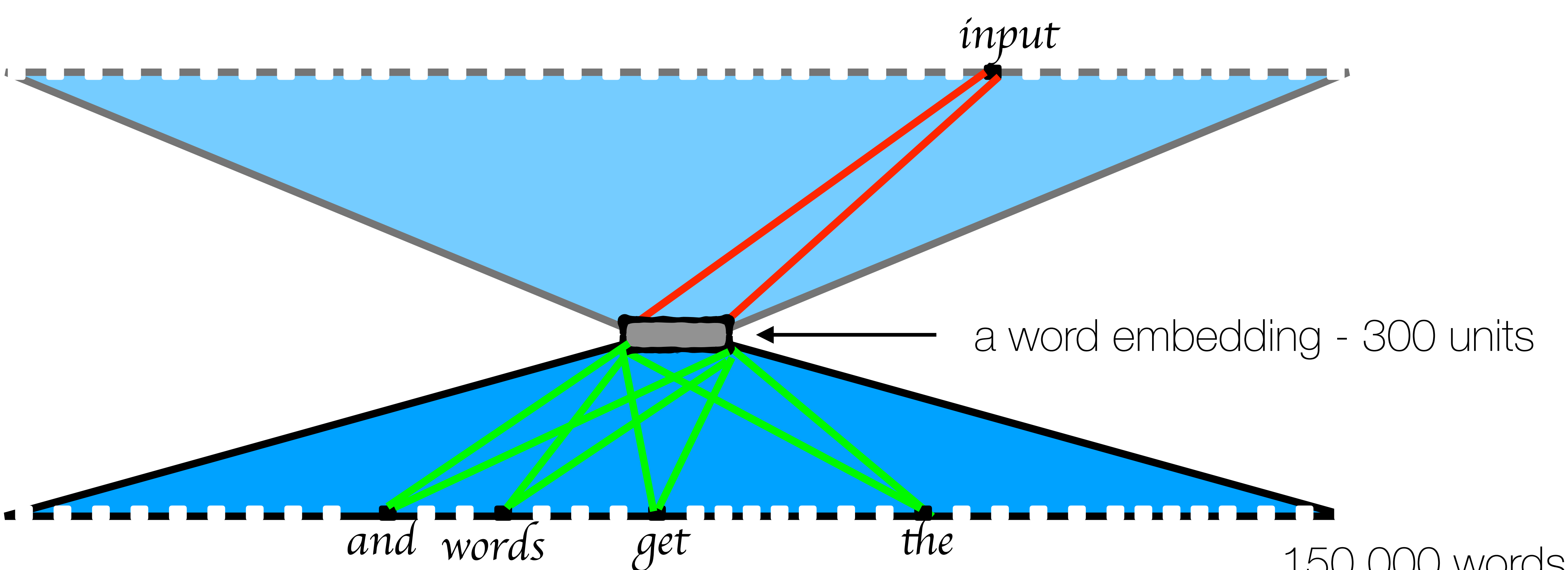
Where do we get the input words and the outputs words?

etc!!!

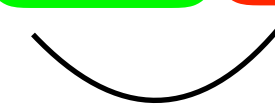
DeepMind

UNIVERSITY OF CAMBRIDGE

CBOW



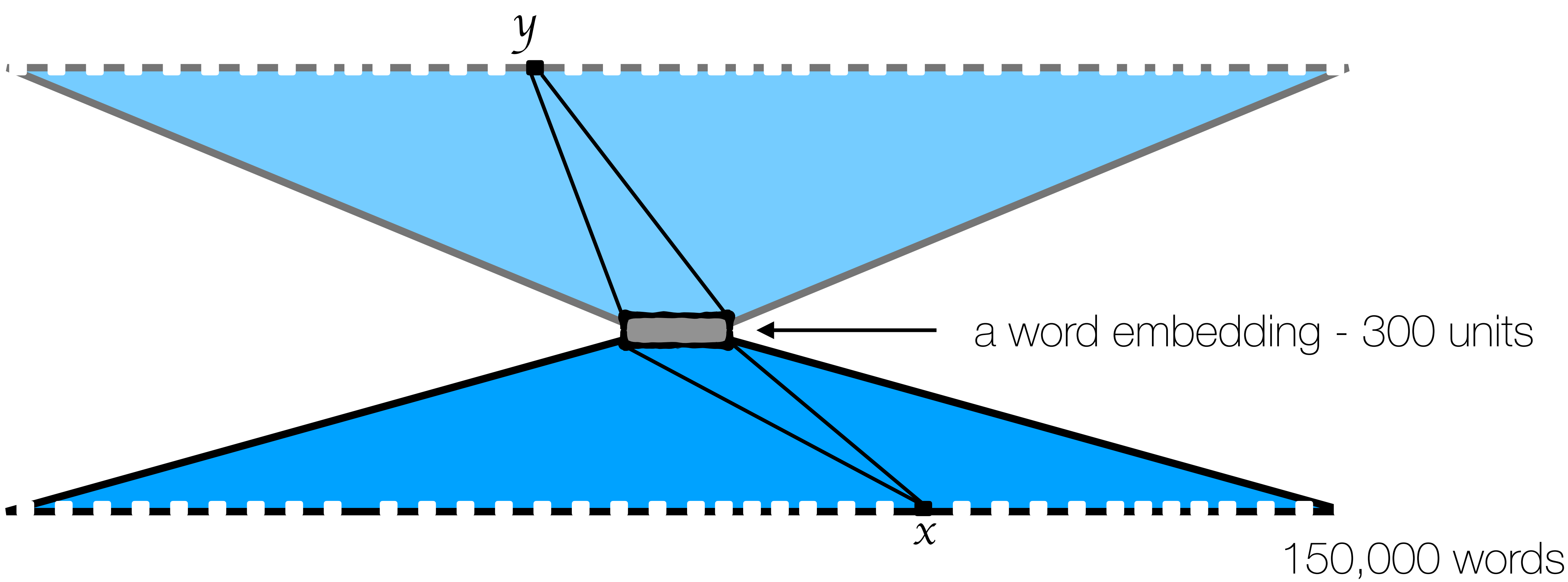
Where do we get the input words and the outputs words?



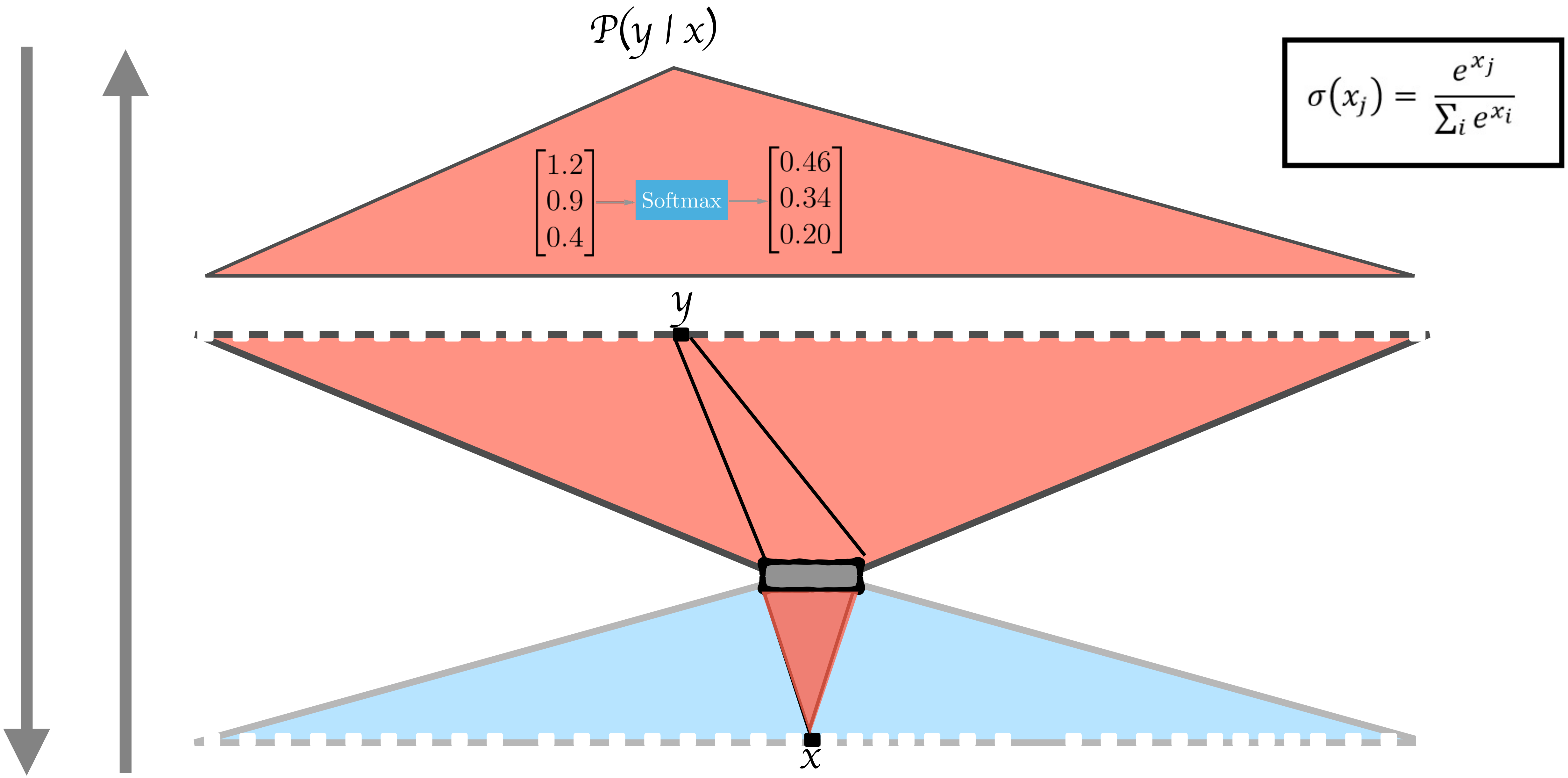
DeepMind

UNIVERSITY OF CAMBRIDGE

How many free parameters in this model?



Computing the loss $$$



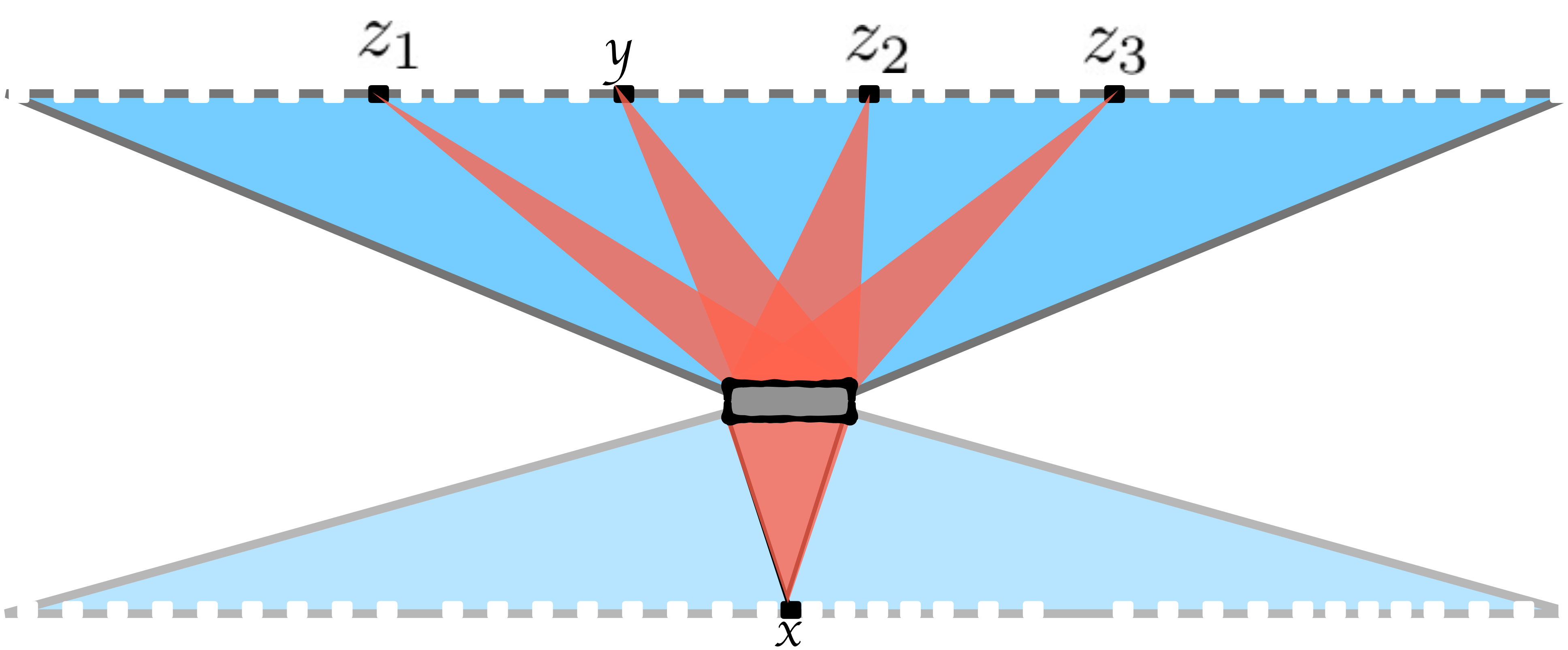
DeepMind

UNIVERSITY OF CAMBRIDGE

A cheaper option...

$$p(y|x) \approx \sigma(y) - 1/3 \sum_i \sigma(z_i)$$

$\sigma(x) =$ a nice score function



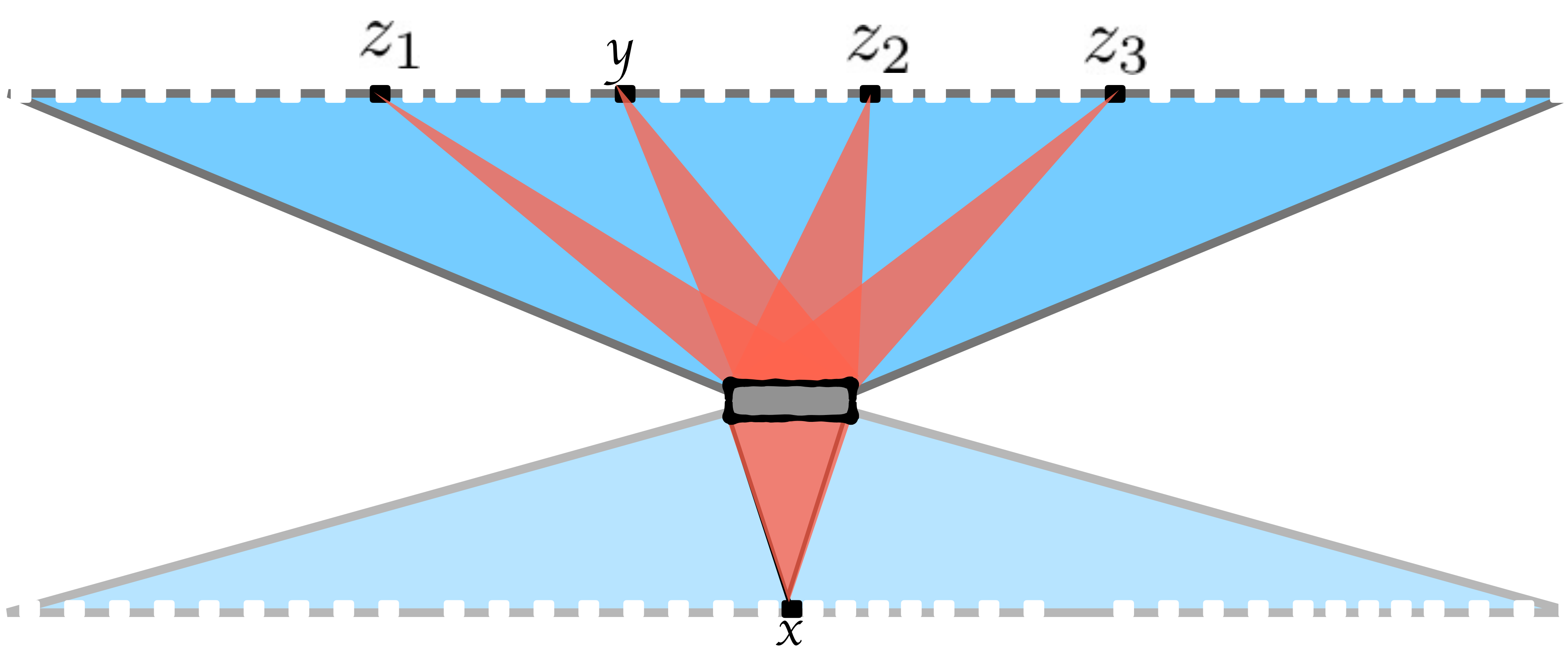
DeepMind

UNIVERSITY OF CAMBRIDGE

"negative sampling"

$$p(y|x) \approx \sigma(y) - 1/3 \sum_i \sigma(z_i)$$

$\sigma(x) =$ a nice score function



DeepMind

UNIVERSITY OF CAMBRIDGE

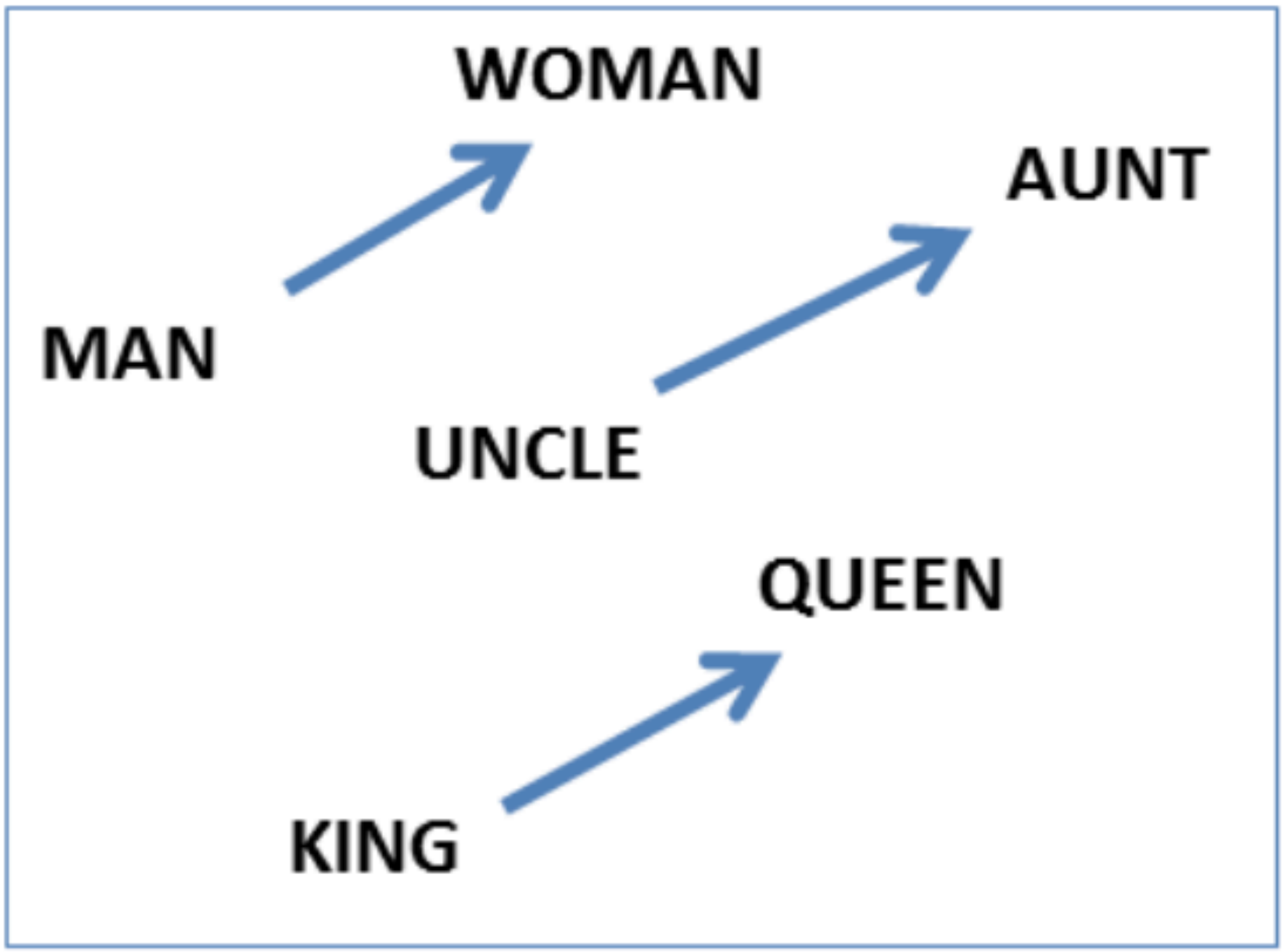
Associated words are close in vector space

http://projector.tensorflow.org/

DeepMind

UNIVERSITY OF CAMBRIDGE

Anything more than that?



DeepMind

UNIVERSITY OF CAMBRIDGE

Next thing you know…

DeepMind

UNIVERSITY OF CAMBRIDGE

Anything more...?

syntax

videos

images

word net

tweets

locations

brain data

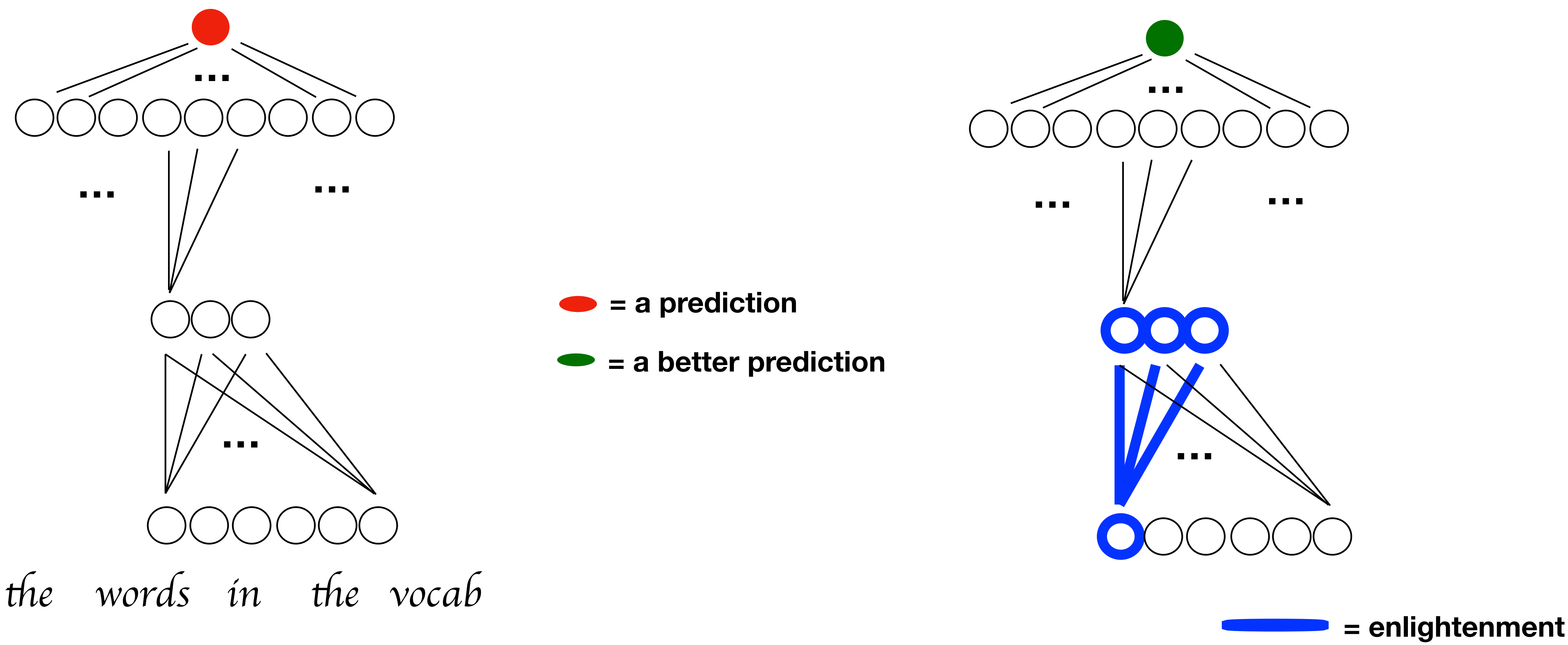
human behavior

DeepMind

UNIVERSITY OF
CAMBRIDGE

the real purpose...

a low-resource language application



DeepMind

UNIVERSITY OF CAMBRIDGE

References

Natural language processing (almost) from scratch (Collobert et al. 2011, from 2008)

Transfer learning with word-embeddings

Efficient estimation of word representations in vector space (Mikolov et al. 2013)

Word2Vec - much faster and easier

Evaluating semantic models with (genuine) similarity estimation (Hill et al. 2014)

Similarity, not just association, in word embedding spaces

Neural word embeddings as implicit matrix factorization (Levy & Goldberg, 2014)

Equivalence between (old) count-based semantic spaces and word2vec