

A long long time ago in an
NLP model far far away...



Are computers able to create stories?

Yes, they are.



Dawid Wiśniewski, Politechnika Poznańska, Applica.ai

Plan prezentacji

- ❖ Definicja problemu
- ❖ Podejście historyczne - modele n-gramowe
- ❖ Sieci rekurencyjne
- ❖ Sieci z atencją

,, There is only one thing that makes a dream impossible to achieve: the fear of failure.

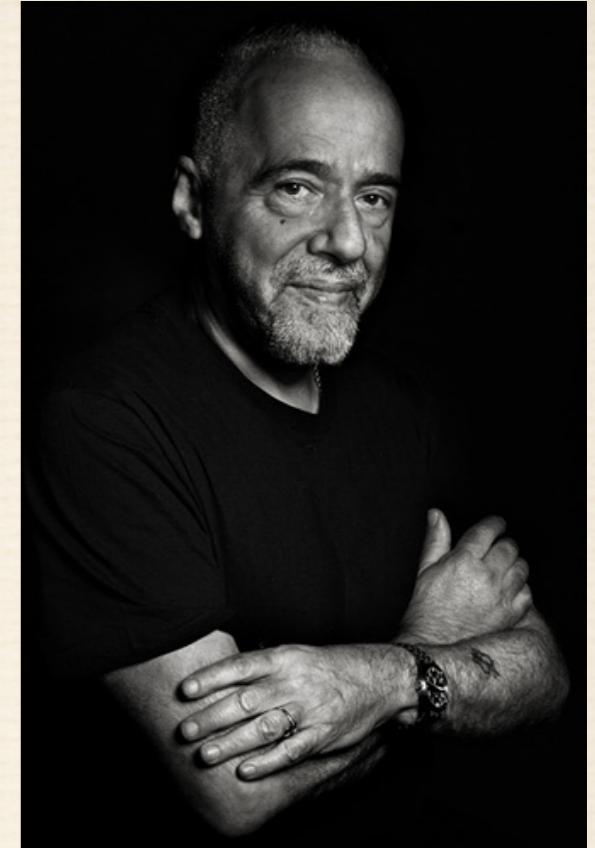
If you wish to achieve your dreams and live in a healthy way, it is important to overcome fear.

But don't become too afraid to speak to your loved ones.

They are an honest and helpful pair of ears. If anyone has doubt, it is them. Their advice can be invaluable to you. This is especially true when it comes to overcoming fear.”



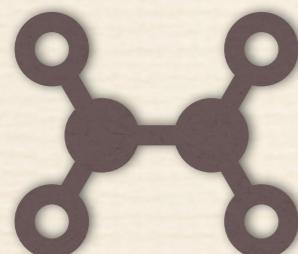
„There is only one thing that makes a dream impossible to achieve: the fear of failure.”



If you wish to achieve your dreams and live in a healthy way, it is important to overcome fear.

But don't become too afraid to speak to your loved ones.

They are an honest and helpful pair of ears. If anyone has doubt, it is them. Their advice can be invaluable to you. This is especially true when it comes to overcoming fear.”



Sieć neuronowa
Model GPT2,
domorosły pisarz
z ambicjami

Modelowanie języka

Modelowanie języka (language modeling/LM) - wytwarzanie modeli potrafiących szacować prawdopodobieństwo wystąpienia sekwencji słów/znaków.

$$P(\text{Ala}, \text{ma}, \text{kota}, \dots) = 0.005 > P(\text{Ala}, \text{kota}, \text{ma}, \dots) = 0.00000001$$

Modelowanie języka

Modelowanie języka (language modeling/LM) - wytwarzanie modeli potrafiących szacować prawdopodobieństwo wystąpienia sekwencji słów/znaków.

$$P(Ala, ma, kota, .) = 0.005 > P(Ala, kota, ma, .) = 0.00000001$$

$$P(kota | Ala, ma) = 0.45$$

$$P(psa | Ala, ma) = 0.15$$

$$P(Ala | Ala, ma) = 0.0002$$

Modelowanie języka

Modelowanie języka (language modeling/LM) - wytwarzanie modeli potrafiących szacować prawdopodobieństwo wystąpienia sekwencji słów/znaków.

$$P(\text{Ala}, \text{ma}, \text{kota}, \dots) = 0.005 > P(\text{Ala}, \text{kota}, \text{ma}, \dots) = 0.00000001$$

$$P(\text{kota} | \text{Ala}, \text{ma}) = 0.45$$

$$P(\text{psa} | \text{Ala}, \text{ma}) = 0.15$$

$$P(\text{Ala} | \text{Ala}, \text{ma}) = 0.0002$$

Ala ma _

Ala ma kota _

Ala ma kota chodzącego _

Ala ma kota chodzącego po _

Ala ma kota chodzącego po drzewach _

...

Modele n-gramowe

Ala ma kota .
Ala ma kota . } 2-gramy (bigramy) słów

Ala
la
a m
ma
a k
... } 3-gramy (trigramy) znaków

Modele n-gramowe

"Mój kot o imieniu Przemek lubi jeść chipsy lays solone, a także "

Wszystko Grafika Wiadomości Filmy Zakupy Więcej Ustawienia

Około 153 wyników (0,73 s)

Nie znaleziono żadnych wyników wyszukiwania dla hasła "Mój kot o imieniu Przemek lubi jeść chipsy lays solone, a także".

Wyniki dla wyszukiwania **Mój kot o imieniu Przemek lubi jeść chipsy lays solone, a także** (bez cudzysłowu):

[Czy to normalne, że mój kotek lubi jeść chipsy ? - Zapytaj.onet ...](#)

Jak wygenerować kontynuację zapytania?
Ograniczyć się jedynie do n-ostatnich słów!

Modele n-gramowe

"Mój kot o imieniu Przemek lubi jeść chipsy lays solone, a także "

Wyszukiwanie

Wszystko Grafika Wiadomości Filmy Zakupy Więcej Ustawienia

Około 153 wyników (0,73 s)

Nie znaleziono żadnych wyników wyszukiwania dla hasła "Mój kot o imieniu Przemek lubi jeść chipsy lays solone, a także".

Wyniki dla wyszukiwania **Mój kot o imieniu Przemek lubi jeść chipsy lays solone, a także** (bez cudzysłowu):

[Czy to normalne, że mój kotek lubi jeść chipsy ? - Zapytaj.onet ...](#)

Jak wygenerować kontynuację zapytania?
Ograniczyć się jedynie do n-ostatnich słów!

$$P(?) | Moj, kot, \dots, a, takze) \approx P(?) | a, takze)$$



Będziesz Pan zadowolony.
(A. Markov, przełom XIX i XX wieku)

Modele n-gramowe

"Mój kot o imieniu Przemek lubi jeść chipsy lays solone, a także "

Wyszukiwanie: Wszystko Grafika Wiadomości Filmy Zakupy Więcej Ustawienia

Około 153 wyników (0,73 s)

Nie znaleziono żadnych wyników wyszukiwania dla hasła "Mój kot o imieniu Przemek lubi jeść chipsy lays solone, a także".

Wyniki dla wyszukiwania **Mój kot o imieniu Przemek lubi jeść chipsy lays solone, a także** (bez cudzysłowu):

[Czy to normalne, że mój kotek lubi jeść chipsy ? - Zapytaj.onet ...](#)

Jak wygenerować kontynuację zapytania?
Ograniczyć się jedynie do n-ostatnich słów!

"a także "

Wyszukiwanie: Wszystko Wiadomości

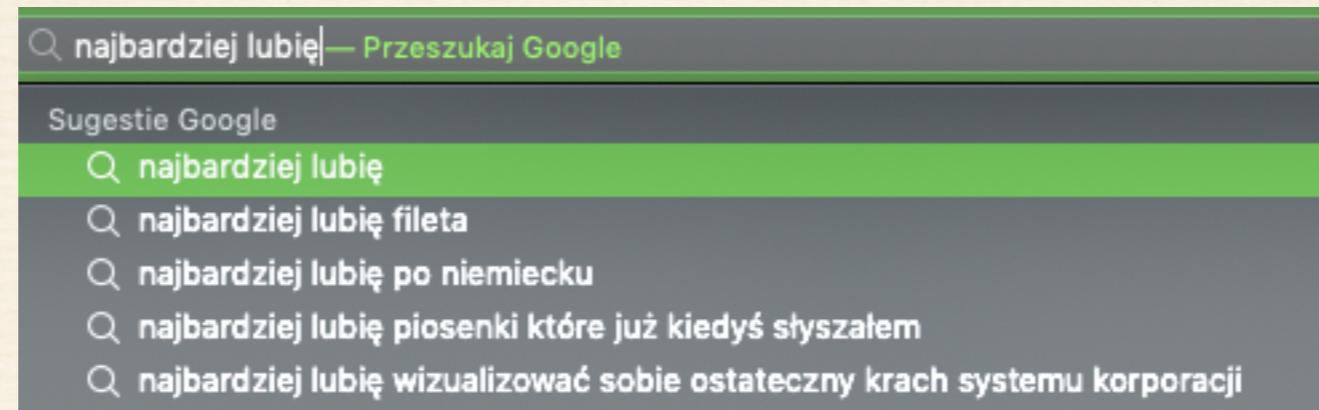
Około 115 000 000 wyników (0,67 s)



$$P(?) | \text{Moj, kot, ..., a, takze}) \approx P(?) | a, takze)$$

Będziesz Pan zadowolony.
(A. Markov, przełom XIX i XX wieku)

Modele n-gramowe



Najbardziej lubię kolor zielony.

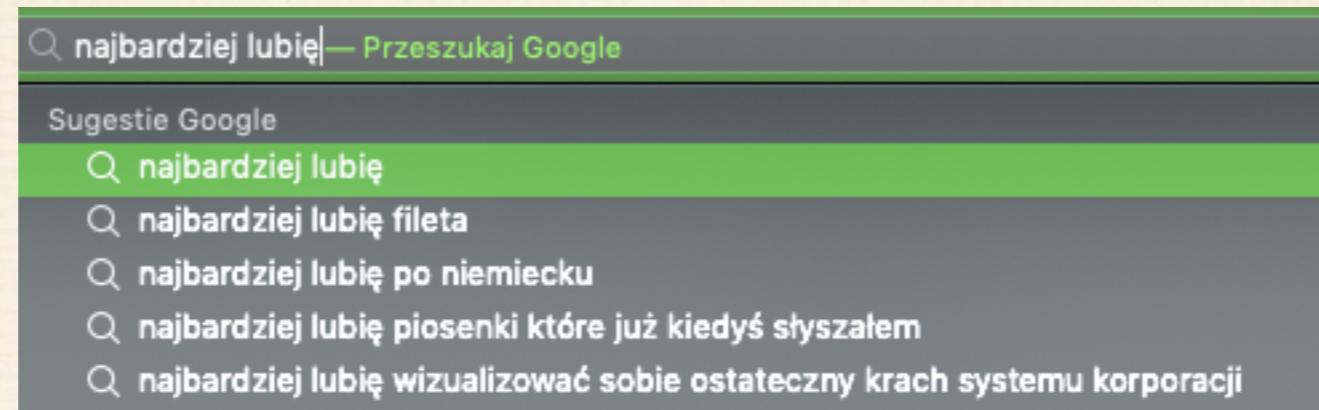
W makijażu najbardziej lubię podkreślać oczy!

Najbardziej lubię tańczyć do disco polo.

W pracy najbardziej lubię piątek i wypłatę.

Najbardziej lubię kolor zachodzącego słońca.

Modele n-gramowe

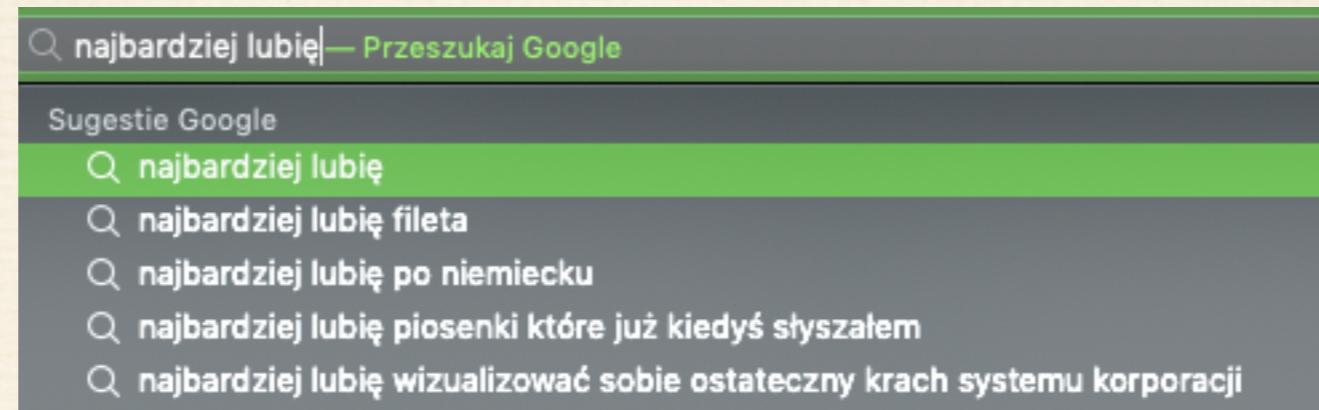


Najbardziej lubię kolor zielony.
W makijażu najbardziej lubię podkreślać oczy!
Najbardziej lubię tańczyć do disco polo.
W pracy najbardziej lubię piątek i wypłatę.
Najbardziej lubię kolor zachodzącego słońca.

Najbardziej lubię

{ kolor: 2/5
 podkreślać: 1/5
 tańczyć: 1/5
 piątek: 1/5

Modele n-gramowe



Najbardziej lubię kolor zielony.
W makijażu najbardziej lubię podkreślać oczy!
Najbardziej lubię tańczyć do disco polo.
W pracy najbardziej lubię piątek i wypłatę.
Najbardziej lubię kolor zachodzącego słońca.

Najbardziej lubię {

kolor: 2/5
podkreślać: 1/5
tańczyć: 1/5
piątek: 1/5

$$P(\text{kolor} | \text{najbardziej, lubie}) = \frac{\text{Count(najbardziej, lubie, kolor)}}{\text{Count(najbardziej, lubie)}} = \frac{2}{5} = 0.4$$

Model n-gramowy - przykład

Średnio co dwa miesiące później.

Komisja i państwa trzecie wspólnie kontynuowały i koordynowały swoje działania.

Niezwyczajnym aspektem są kwalifikacje personelu, które są korzystniejsze dla konsumentów i przedsiębiorstw w oparciu o czysto obiektywne kryteria, tak by zaowocowały uważnie zaprojektowanymi interwencjami i programami uwzględniającymi specyfikę terytorialną i relatywną przewagę lub szczególne cechy każdego regionu w ramach regionalnych funduszy spójności tylko dla przemysłu tych krajów UE, zamrozić budżet UE na temat przyszłości WPR.

Modele n-gramowe

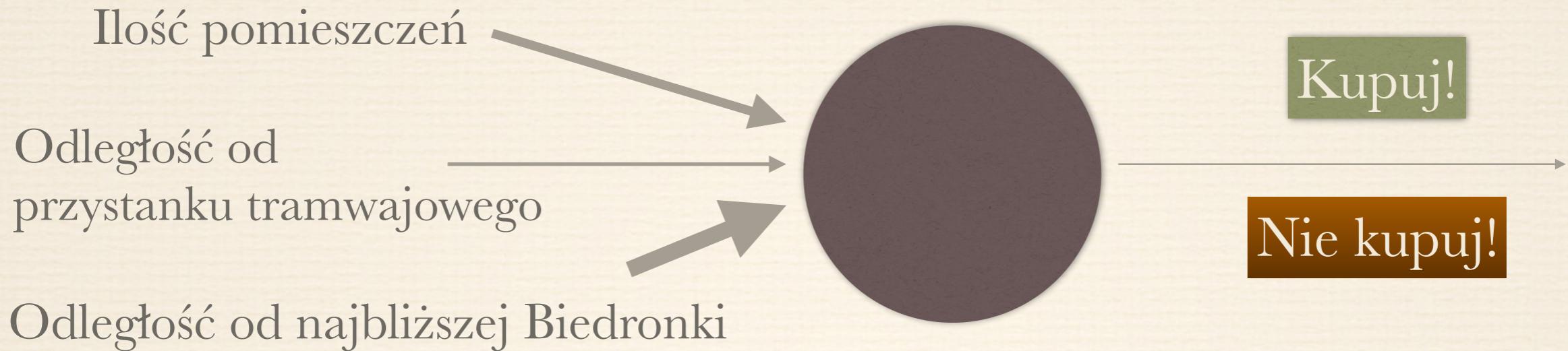


Latwe do wyliczenia prawdopodobieństwa.
Łatwa do zrozumienia zasada działania.

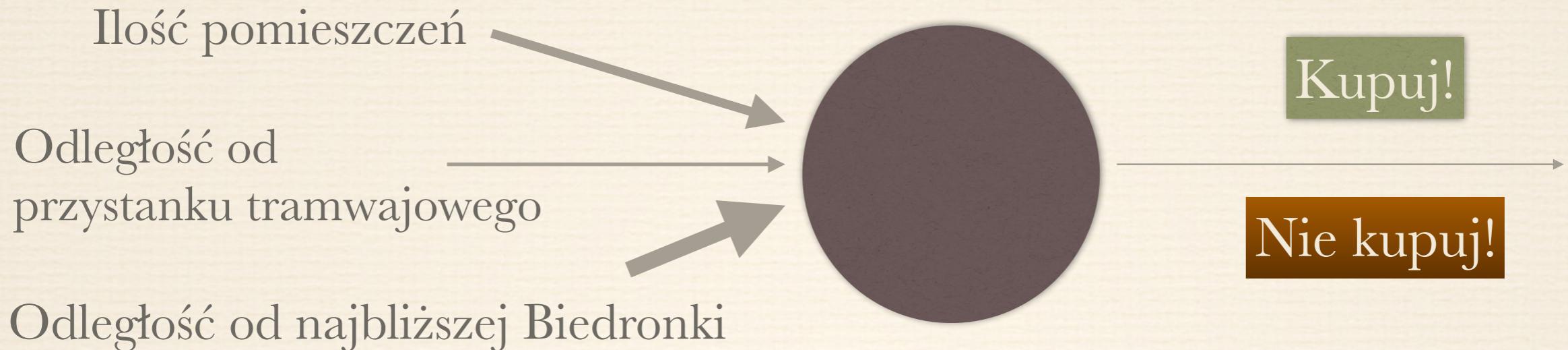


Utrata kontekstu wynikająca z wykorzystania założenia Markova.
Nieoczywisty sposób doboru długości n-gramów.

Sieci neuronowe

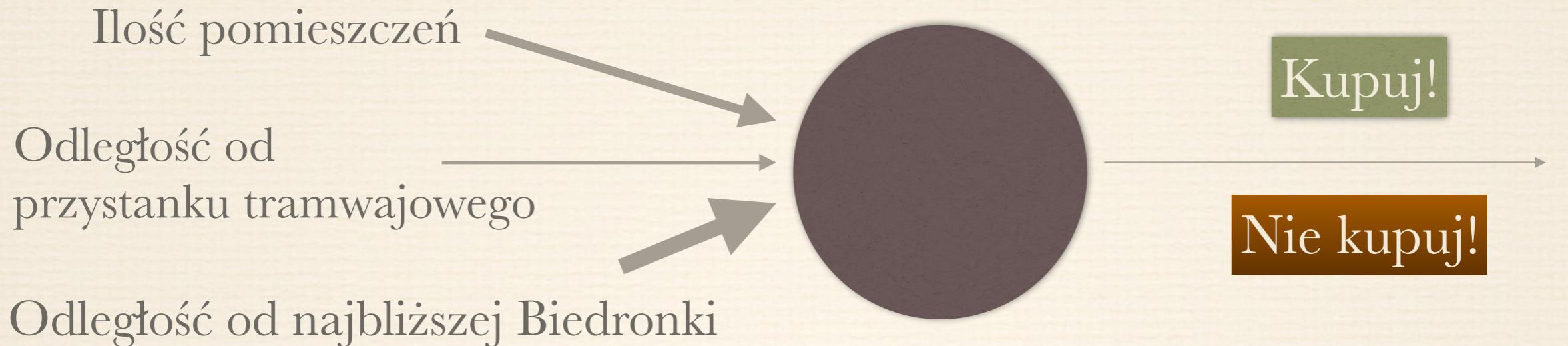


Sieci neuronowe



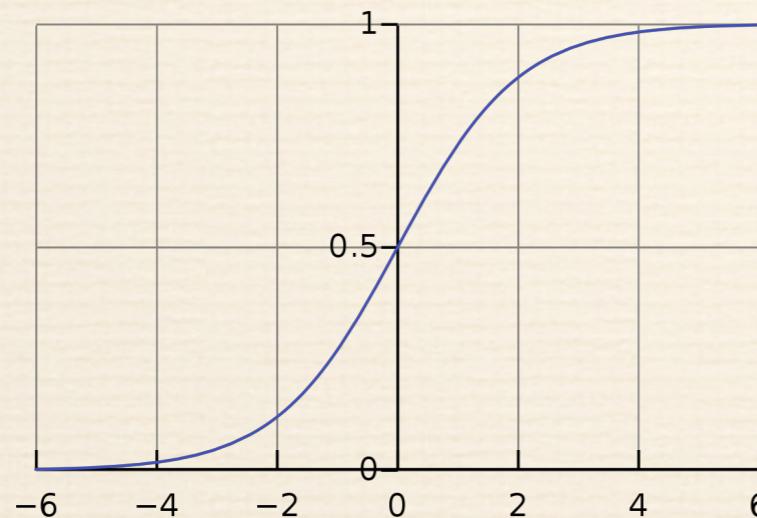
$$y = \text{activation}(\underbrace{w_1 \cdot \text{iloscpomieszczen}}_{\text{Tego się uczymy}} + \underbrace{w_2 \cdot \text{odprzystanku}}_{\text{Tego się uczymy}} + \underbrace{w_3 \cdot \text{odbiedronki}}_{\text{Tego się uczymy}} + b)$$

Sieci neuronowe



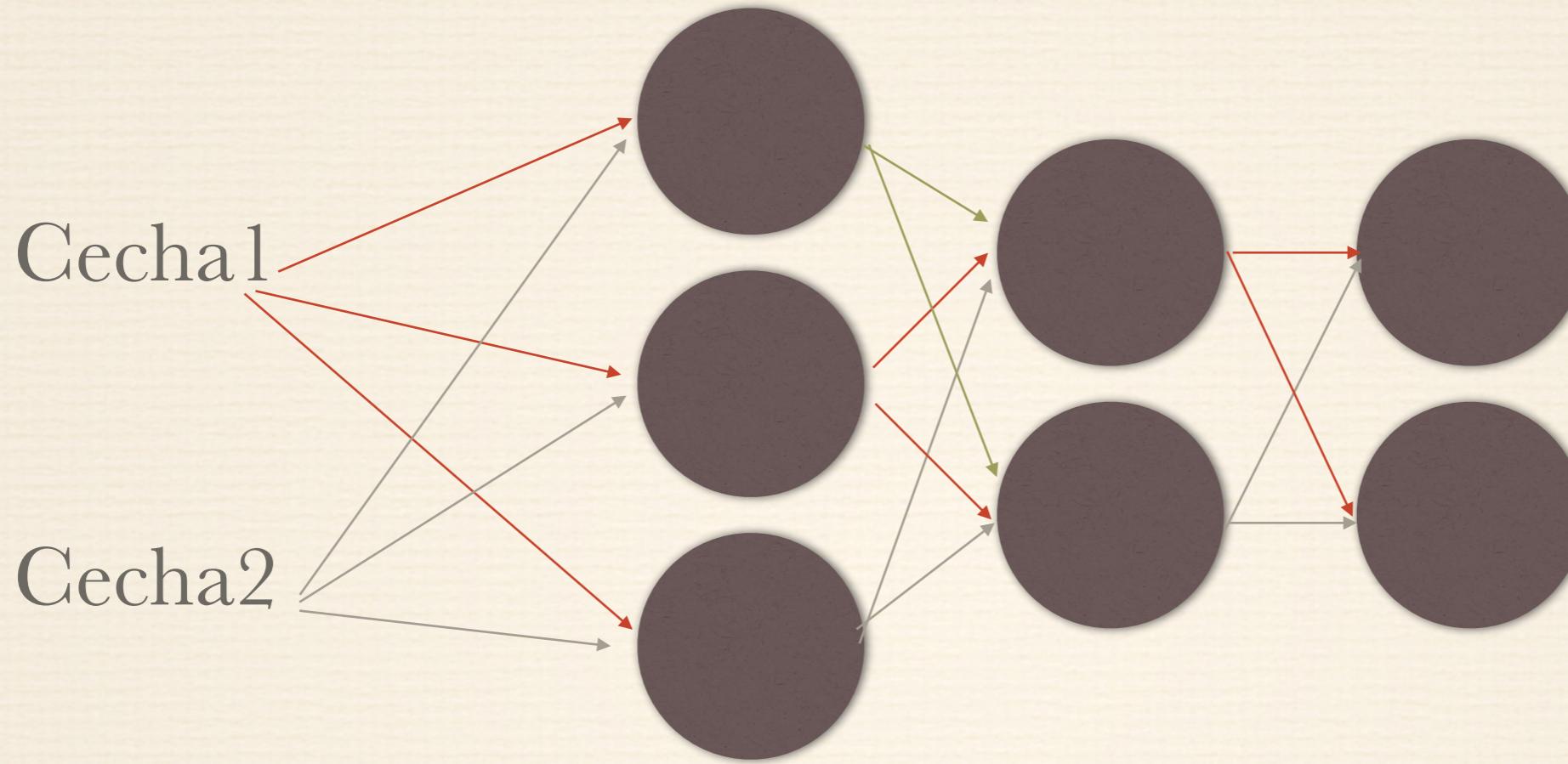
$$y = \text{activation}(\underbrace{w_1 \cdot \text{ilosc pomieszczen}}_{\text{\color{red}{w}}_1} + \underbrace{w_2 \cdot \text{odprzystanku}}_{\text{\color{red}{w}}_2} + \underbrace{w_3 \cdot \text{odbiedronki}}_{\text{\color{red}{w}}_3} + b)$$

Tego się uczymy

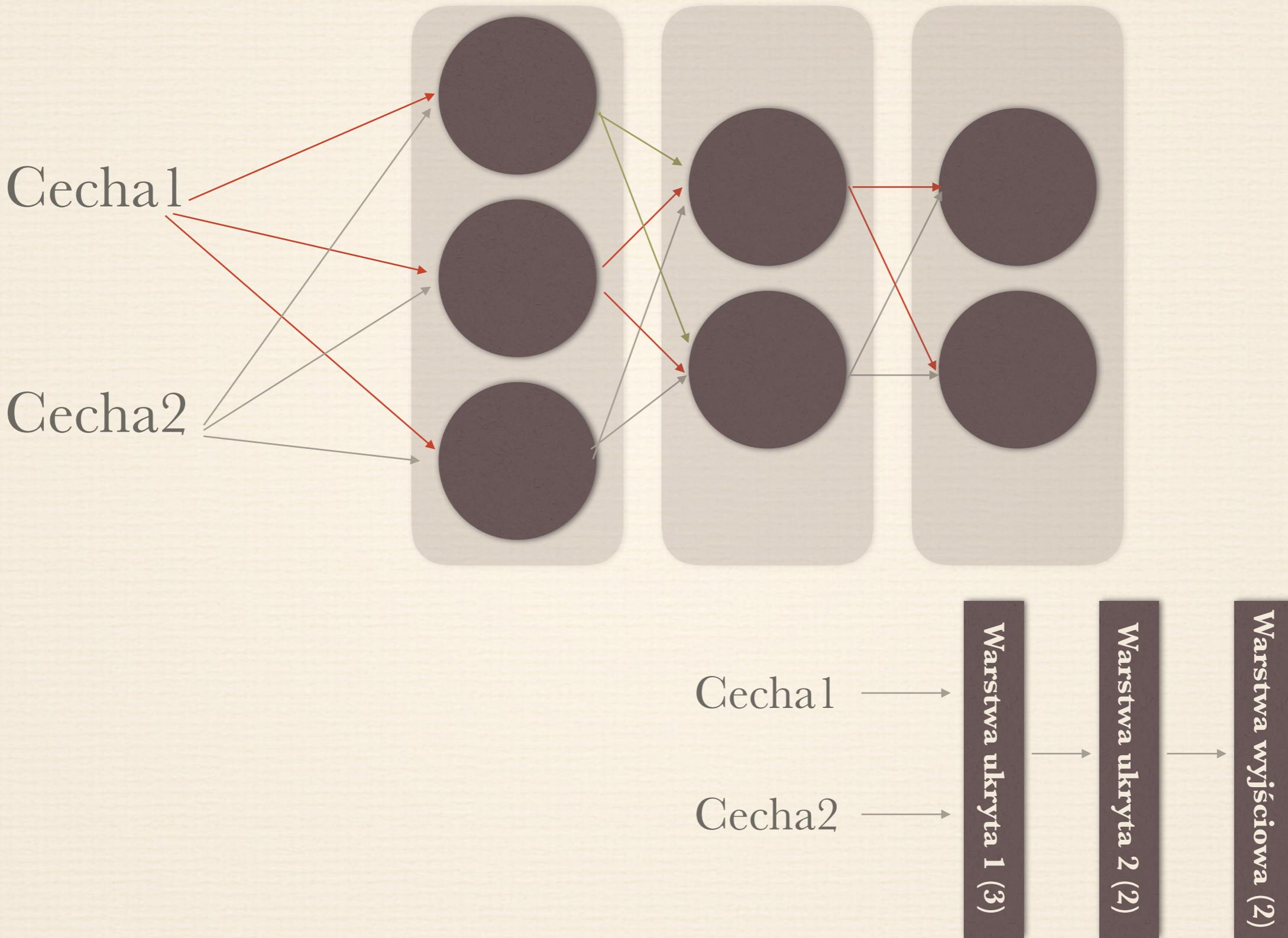


$$\text{sigmoid}(x) = \frac{1}{1 + e^{-x}}$$

Sieci neuronowe



Sieci neuronowe



No tak, ale tekst jest sekwencją nieznanych rozmiarów,
jak umieścić go w takiej sieci i jak ze słów zrobić liczby?

Word Embeddings

Ala



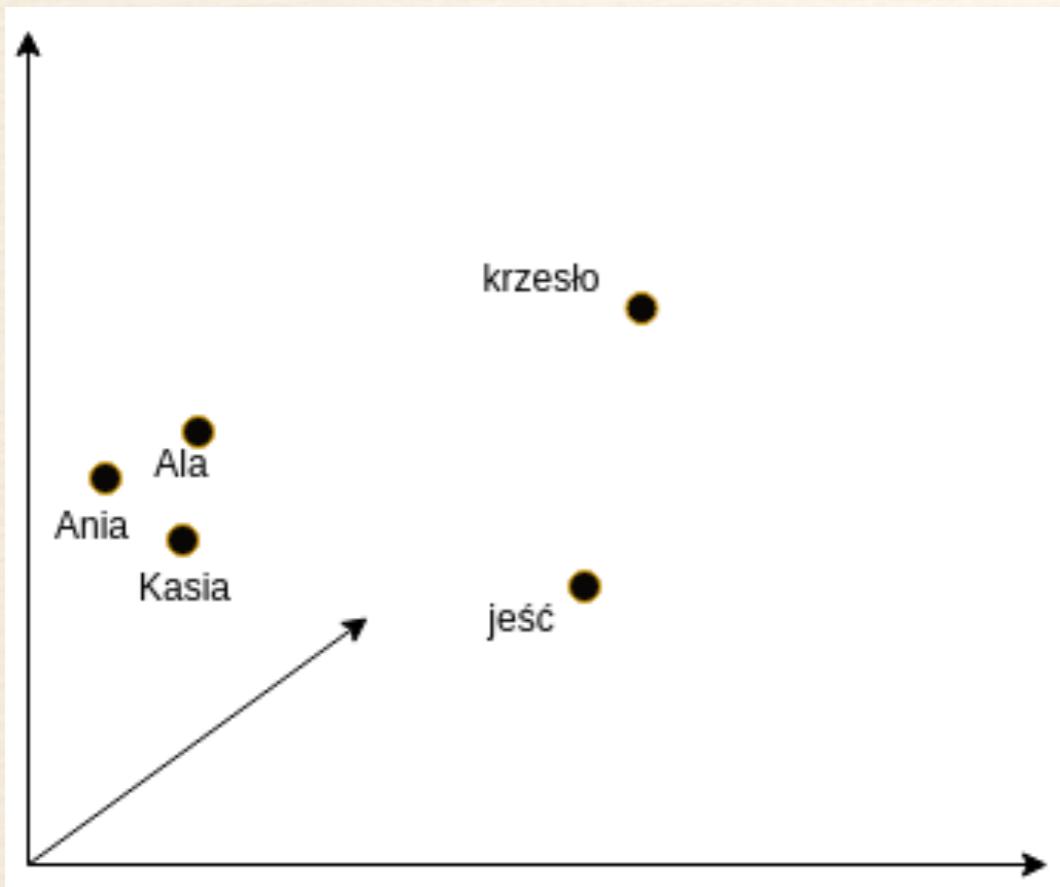
[−1.5,2.7,0.001,3.91]

Word Embeddings

Ala



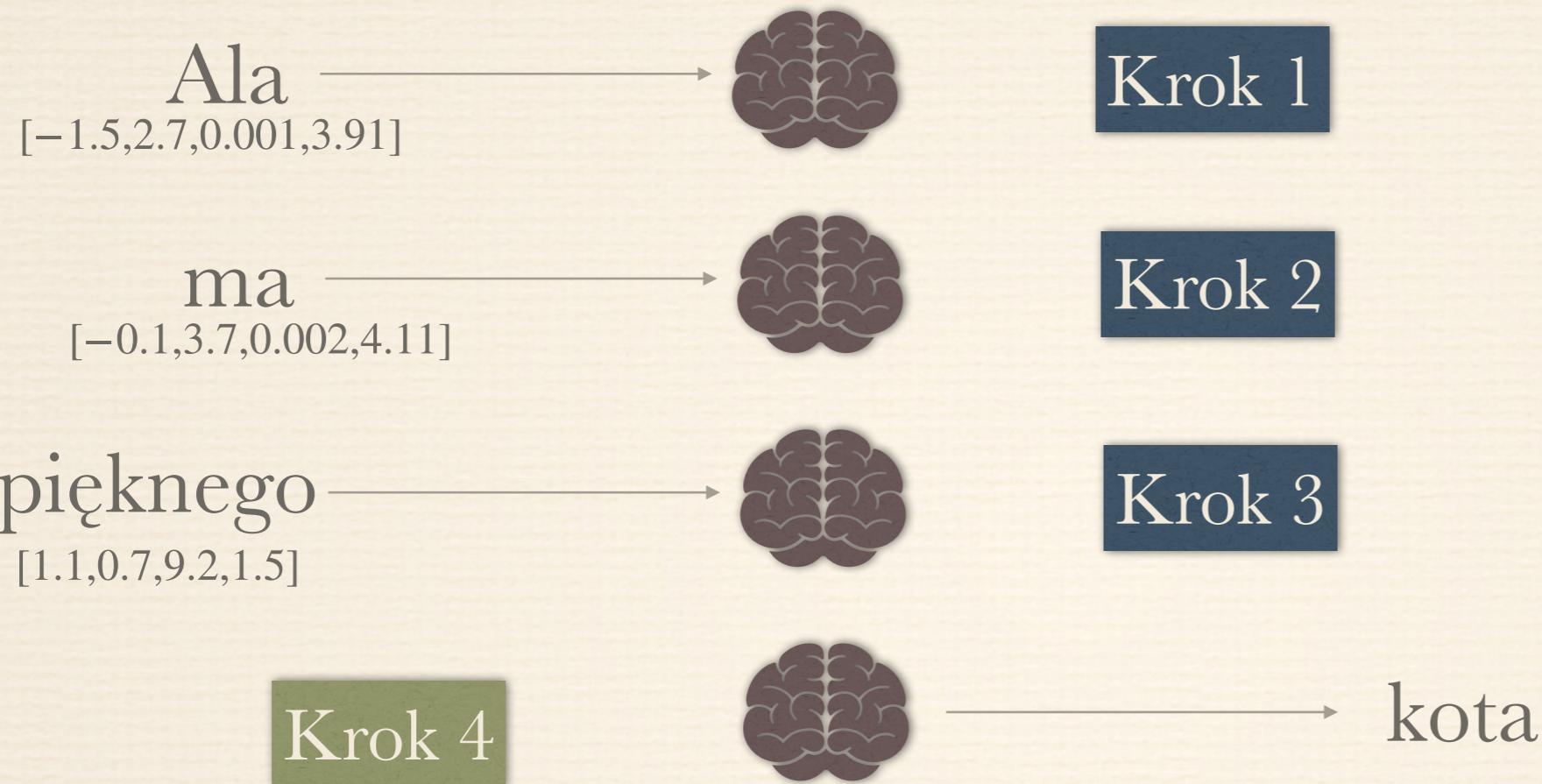
[−1.5,2.7,0.001,3.91]



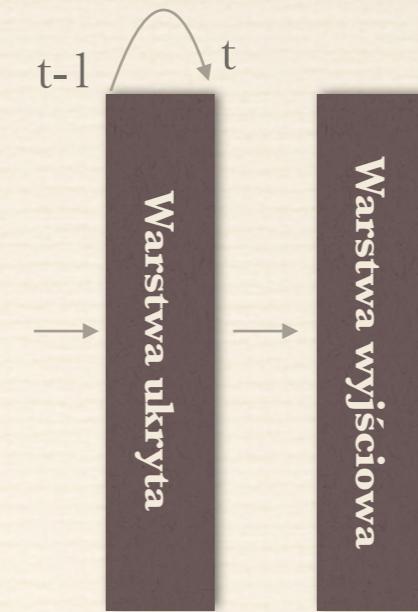
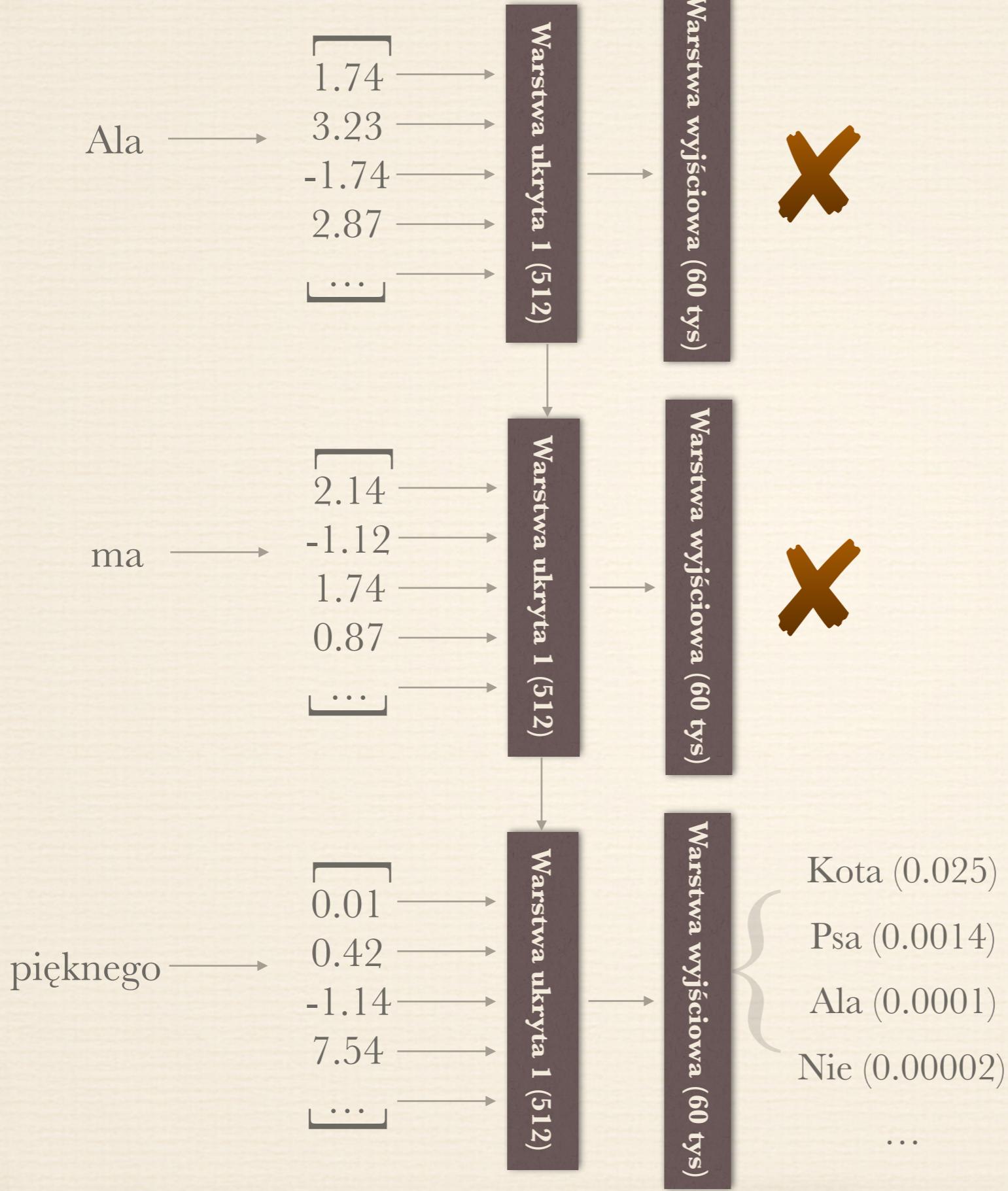
Znaczenie słowa określone
przez otoczenie w którym występuje

GloVe
Word2Vec

Sieci rekurencyjne



Sieć powinna wygenerować wynik po przeczytaniu całej sekwencji:
Ala ma pięknego



Vanilla RNN
GRU
LSTM

Warstwa ukryta modeluje informacje o całej sekwencji agregując wiedzę o aktualnym słowie i słowach poprzedzających.

Sieci rekurencyjne

- ✓ Nie jesteśmy ograniczani do n-ostatnich słów jak w podejściach n-gramowych, tracimy mniej informacji.
- ✗ Warstwa ukryta staje się mieszanką wszystkich widzianych dotąd słów. Dla coraz dłuższych sekwencji coraz ciężej „odszyfrować” zawartość tej warstwy, miesza nam się kontekst, LSTM/GRU pomaga, ale tylko trochę.
- ✗ Ograniczona równoległość z uwagi na zależność reprezentacji stanu aktualnego od stanu wcześniejszego.

Proszę o uwagę! - Transformer

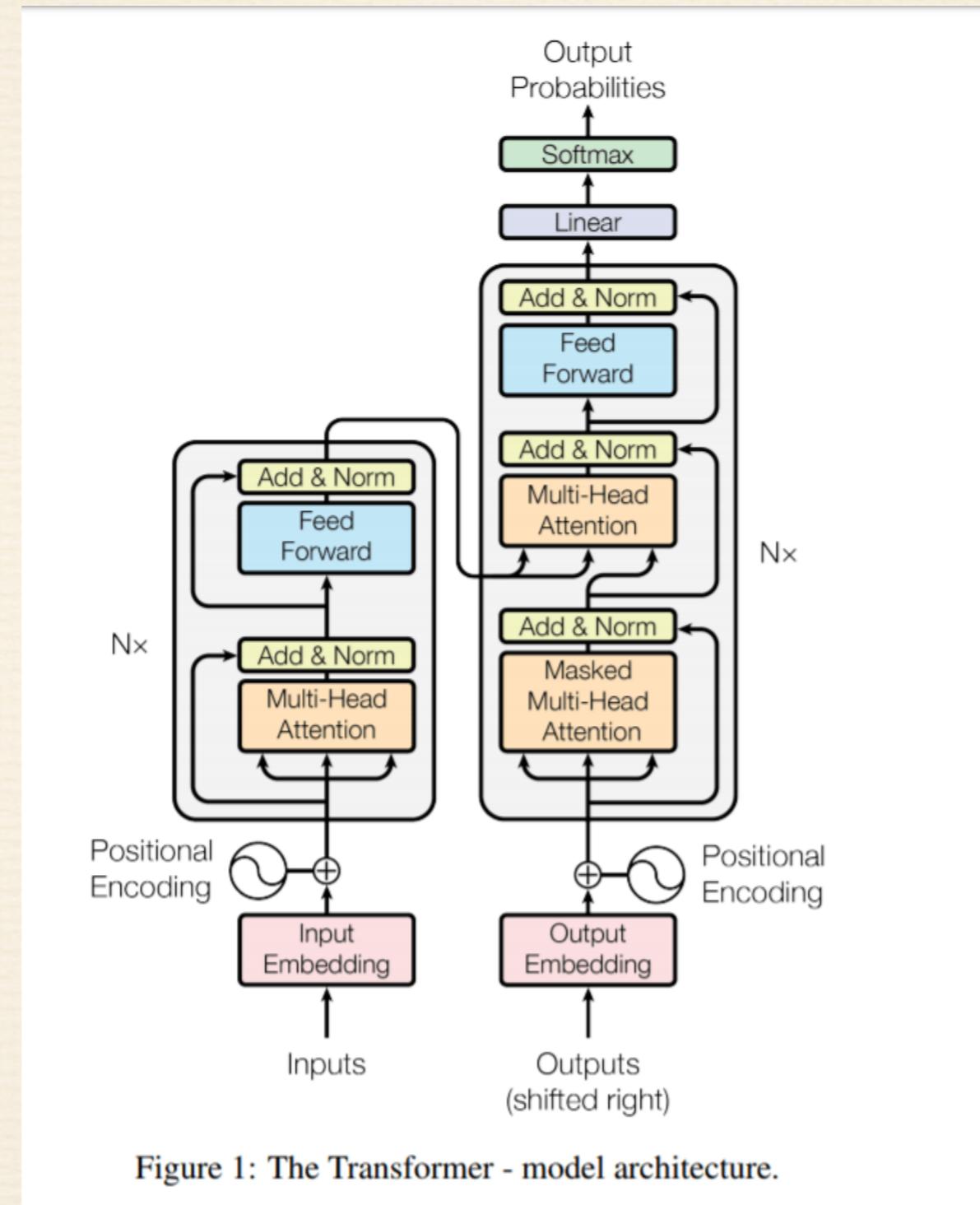
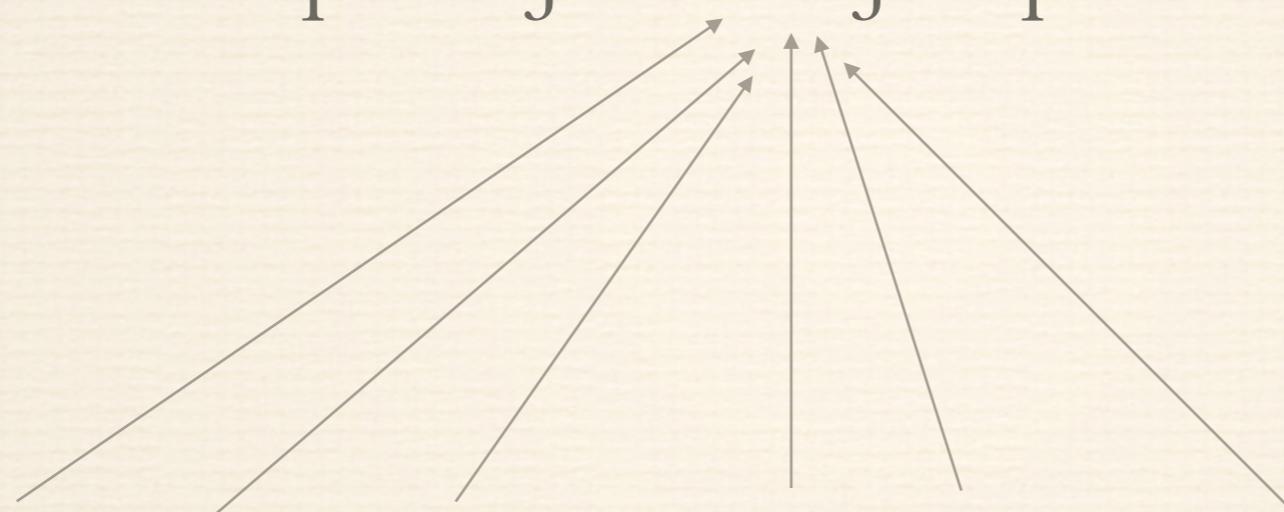


Figure 1: The Transformer - model architecture.

Self-attention

Ala dużo pracuje. Ona jest pracoholiczka

Ala dużo pracuje. Ona jest pracoholiczka



Transformer

- ✓ Atencja zamiast rekurencji pozwala na lepsze zrównoleglenie - uczenie jest efektywniejsze
- ✓ Mechanizm atencji pozwala na „skupienie się” na poprzednich słowach, a nie tylko na zagregowanej reprezentacji całego wcześniejszego kontekstu.
- ✗ Skomplikowana architektura wymagająca sporo wysiłku w celu zrozumienia działania.

Architektury oparte o transformer

- ❖ GPT-2 (5.11.2019)
- ❖ XLM (2019)
- ❖ BERT (2018)
- ❖ ALBERT (2019)
- ❖ Evolved Transformer (2019)
- ❖ RoBERTa (2019)
- ❖ Megatron-LM (2019)

Ilość danych, parametrów i koszt

- ❖ GPT-2 - 1.5 mld (!) parametrów, 40GB tekstu, koszt: 43 000 \$
- ❖ Megatron-LM - 8.3 mld (!) parametrów, 174 GB tekstu, koszt 400000 \$

Dodatkowe materiały

- ❖ <https://colah.github.io/posts/2014-07-NLP-RNNs-Representations/>
- ❖ <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>
- ❖ <https://distill.pub/2016/augmented-rnns/>
- ❖ <http://karpathy.github.io/2015/05/21/rnn-effectiveness/>
- ❖ [Transformer] Attention Is All You Need | AISC Foundational - Youtube video - https://www.youtube.com/watch?v=S0KakHcj_rs&t=1300s
- ❖ <https://towardsdatascience.com/bert-explained-state-of-the-art-language-model-for-nlp-f8b21a9b6270>

Kilka przykładów - <https://talktotransformer.com>

Q: Is earth flat?

A: Yes, it is flat.

Q: What does "flat earth" mean?

A: It means that the Earth is a plane surface.

Q: Can you describe a plane surface?

A: It is like a smooth marble, a flat surface, like a round table.

Q: So the Earth is flat?

A: Yes.

Q: Can you see through the Earth?

A: There is a thin layer of haze between the Earth and the sun. It is called the atmosphere.

Q: Does the atmosphere make a big hole in the Earth?

A: It does.

Q: So there is a hole in the Earth?

A: Yes.

Q: Could you

Kilka przykładów - <https://talktotransformer.com>

I've seen Donald Trump playing Pokemon games last night. He was screaming: "I will make Pikachu great again" as he played through the game. And I thought: What if I used that as an opportunity to talk about the dangers of using guns?"

In another post, Trump posted "It is so good to be back in Michigan. Big crowds, amazing energy!" The post went on to boast that "a lot of people say we don't have energy left."

Trump has repeatedly used Pokemon Go to encourage supporters to go out and vote. On Monday, he posted that he had already cast his vote in Wisconsin's Republican primary, adding that he was heading back to Michigan "right now."

Kilka przykładów - <https://talktotransformer.com>

GENERATE ANOTHER

Completion

Q: Will AI conquer us?

A: Yes, very soon.

Q: Why?

A: Because we're so stupid.

Q: What's our excuse?

A: It's been too long since we took control of the galaxy!

Q: What will we do then?

A: Let AI do whatever they want!

Q: How do you plan to keep them in check?

A: Hire lots of soldiers!

Q: How will you deal with an AI rebellion?

A: Use the same methods as a human.

Q: Why not just exterminate the humans?

A: It's too hard!

Q: Why not go

Uff.. :)