

# Transfer learning in paraphrase detection for English and Spanish languages

**maks-ym**

**Promotor: dr. inż Piotr Andruszkiewicz**

**24.10.2019**

# Agenda

- Wstęp: parafraza, transfer learning, sieci neuronowe
- Cel i motywacja
- Architektura. Dane. Założenia eksperymentu
- Wyniki
- Podsumowanie
- Pytania

**Wstep**

# Wstęp

## Parafraza

Zdanie  $\approx$  Inne zdanie

*“A man puts some cheese on a pizza.” — “The man sprinkled cheese on the pizza.”*  
*[Z danych testowych SemEval]*

*“The man sprinkled cheese on the pizza.”*

*“Mężczyzna posypał pizzę serem.”*

*“Чоловік посипав сиром піцу.”*

# Wstęp

## Parafraza

*"A man puts some cheese on a pizza." — "The man sprinkled cheese on the pizza."  
[Z danych testowych SemEval]*

*"Drugi stambulskie oddycha gorycze  
Lub pije z chińskich ziół ciągnione treści."*

*A. Mickiewicz ("Zima miejska")*

Peryfrazza  
figura stylistyczna, polega na zastąpieniu wyrazu przez szereg innych) parafraza  
(swobodna przeróbka tekstu lub tłumaczenia)

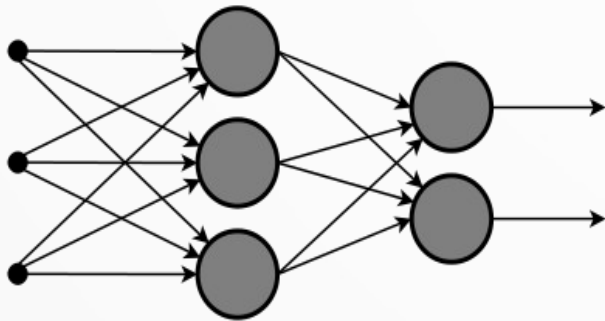
# Wstęp

No a czym jest „transfer learning”?

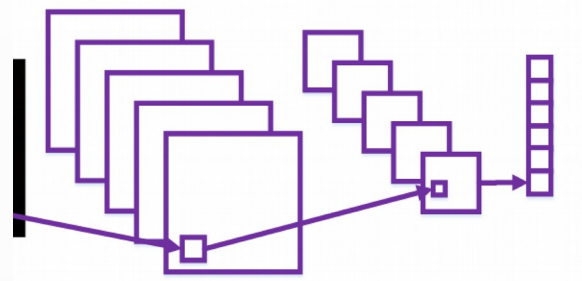
*„Transfer learning is the improvement of learning in a new task through the transfer of knowledge from a related task that has already been learned.” [6]*

# Wstep

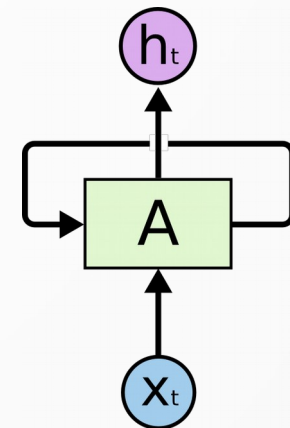
## Sieci neuronowe



Feed-forward [1]



CNN [2]



RNN [3]

[1] <https://towardsdatascience.com/learn-how-recurrent-neural-networks-work-84e975feaaf7>

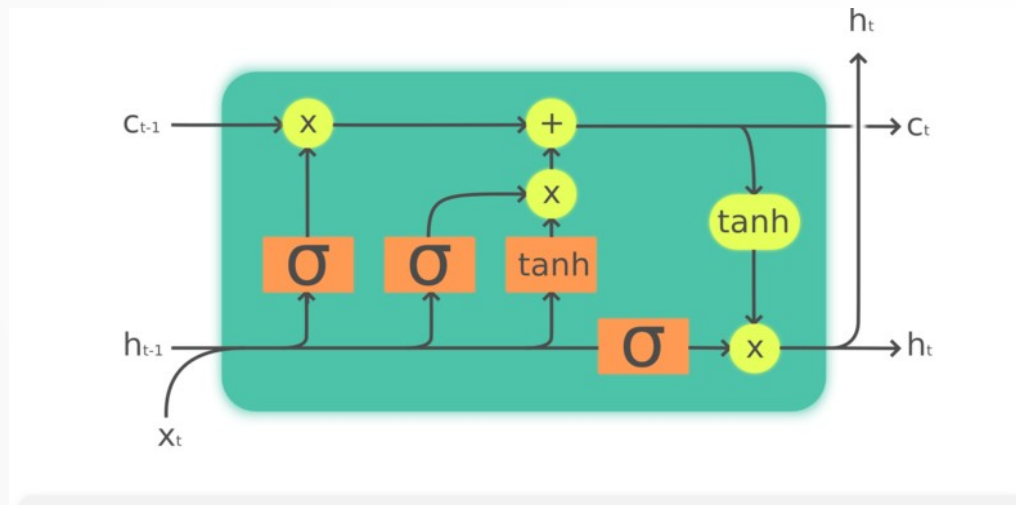
[2] <https://rubiksgcode.net/2018/02/26/introduction-to-convolutional-neural-networks/>

[3] <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>



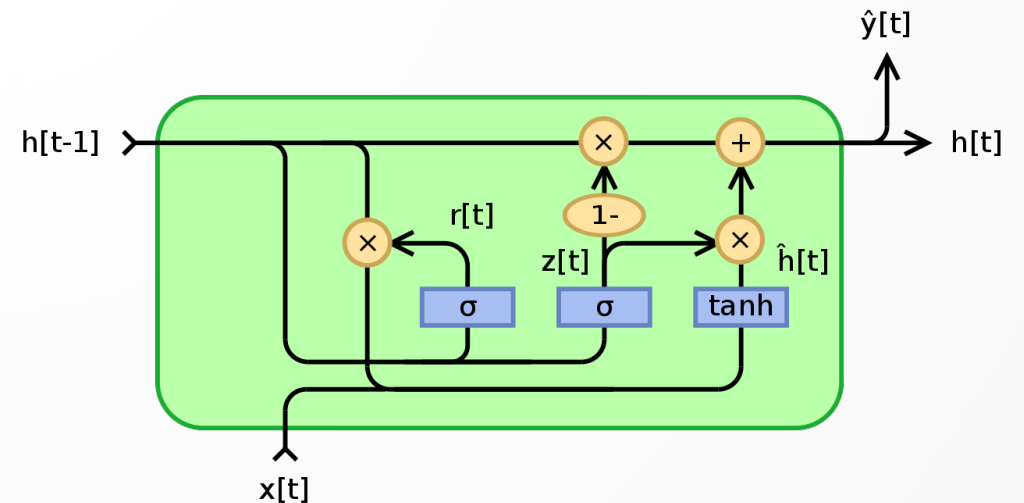
# Wstep

## Sieci neuronowe → RNN



LSTM

(Long Short-Time Memory) [4]



GRU

(Gated Recurrent Unit) [5]

[4] [https://en.wikipedia.org/wiki/Long\\_short-term\\_memory](https://en.wikipedia.org/wiki/Long_short-term_memory)

[5] [https://en.wikipedia.org/wiki/Gated\\_recurrent\\_unit](https://en.wikipedia.org/wiki/Gated_recurrent_unit)



# **Cel i motywacja**

# Cel i motywacja

## Parafrazy

Niejednoznaczność języka naturalnego: różne pytania mogą żądać tej samej odpowiedzi

Wykorzystanie już istniejących baz wiedzy

Wyszukiwarki i chat-boty

## Transfer Learning

Ograniczona ilość danych dla zadania docelowego, ale dużo danych dla podobnego zadania

Danych dla języka angielskiego jest zwykle najwięcej, dla innych o wiele mniej, ale chcemy uzyskać w miarę dobrą jakość

Tworzenie nowych zbiorów danych wysokiej jakości często jest drogie

## Sieci neuronowe

Maksymalnie zautomatyzowany sposób stworzenia systemu

Minimalne wykorzystanie predefiniowanych reguł językowych

Współczesny sprzęt pozwala na wytrenowanie prostej sieci nawet na komputerze personalnym

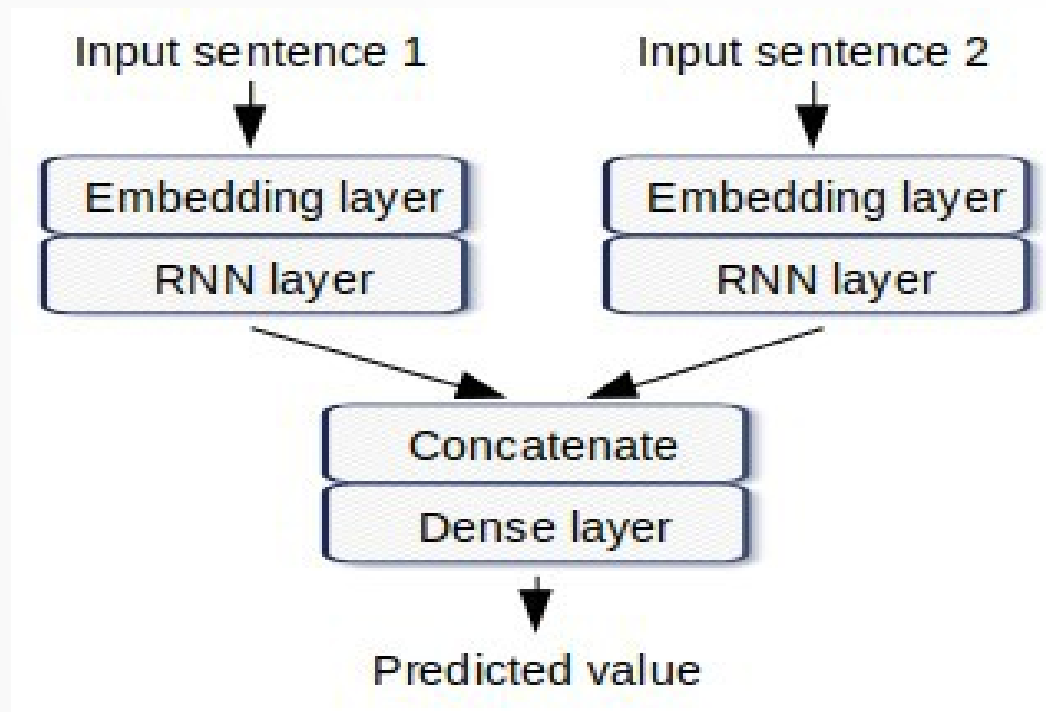
# **Architektura. Dane. Założenia eksperymentu**

# Architektura. Dane. Założenia eksperymentu

- Opiera się na pracy B. Rychalskiej “Sieci neuronowe w rozpoznawaniu podobieństwa semantycznego” (2016)
- Prosta liniowa architektura warstwy rekurencyjnej
- Wektorowa reprezentacja słów FastText
- Dane z konkursu SemEval różnych lat
- Języki: angielski i hiszpański
- Trenowanie na maszynie lokalnej
- python, jupyter notebook, keras (TensorFlow backend)

# Architektura. Dane. Założenia eksperymentu

## Architektura



## Dane

	English	Spanish
Training	2234	1083
Test	350	350
Validation	1959	361

## Wektory słów

- FastText
- 200000 słów dla każdego języka
- 300 długość jednego wektora

# Sprzęt do obliczeń

PC:

- MSI GeForce 1060Ti, 6Gb
- 8 RAM
- Intel Pentium Gold
- MSI Z370

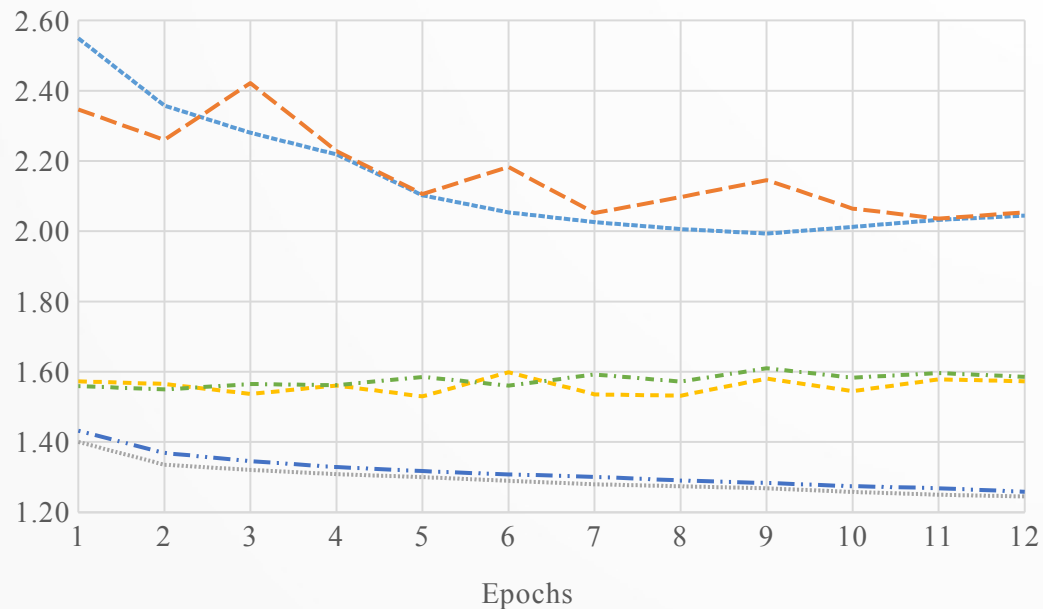


# Wyniki



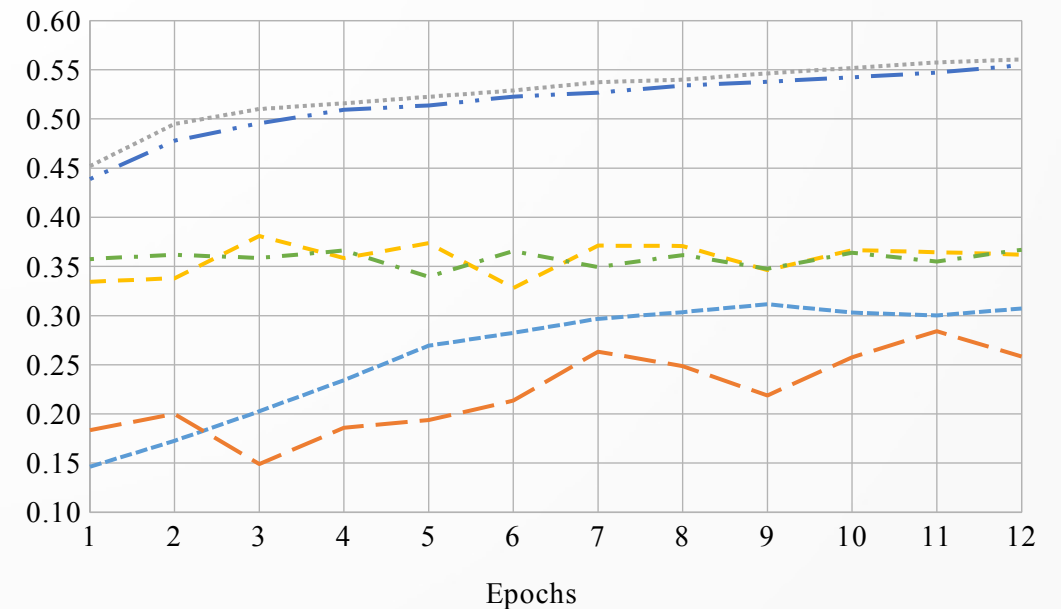
# Wyniki

## Wybór bramki RNN



----- SimpleRNN - training    - - - - - SimpleRNN - validation  
..... LSTM - training       - - - - - LSTM - validation  
- . - . - GRU - training       - . - . - GRU - validation

Funkcja  
kosztu

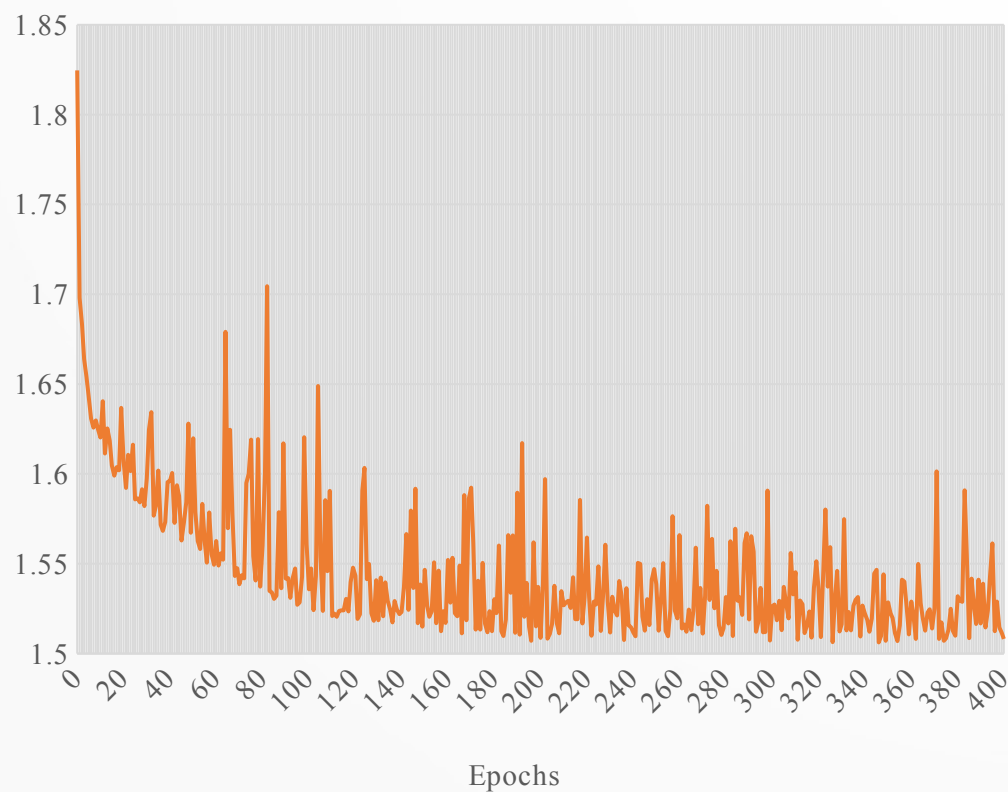


----- SimpleRNN - training    - - - - - SimpleRNN - validation  
..... LSTM - training       - - - - - LSTM - validation  
- . - . - GRU - training       - . - . - GRU - validation

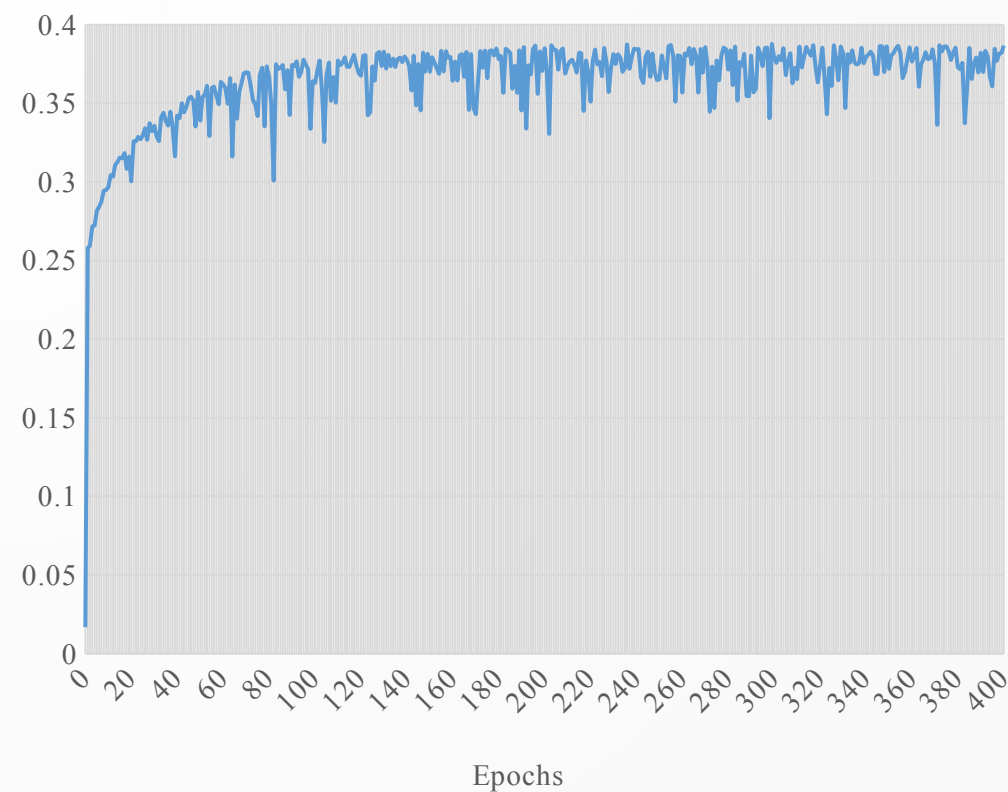
Współczynnik korelacji  
Pearsona

# Wyniki

## Wybór liczby iteracji



Funkcja  
kosztu

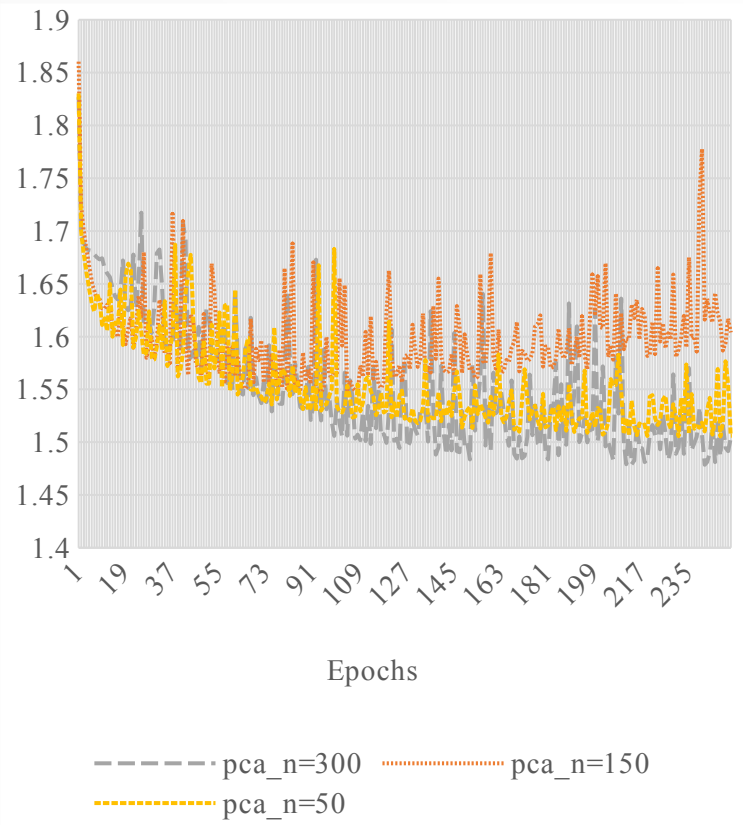


Współczynnik korelacji  
Pearsona

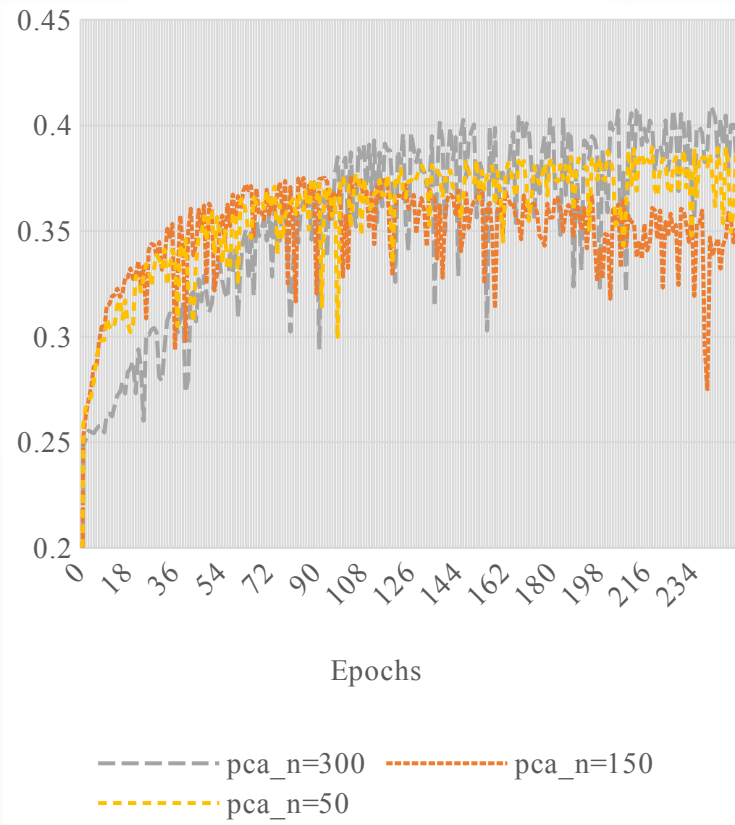
\* zbiór walidacyjny

# Wyniki

## Wybór liczby głównych składowych (PCA)



Funkcja kosztu

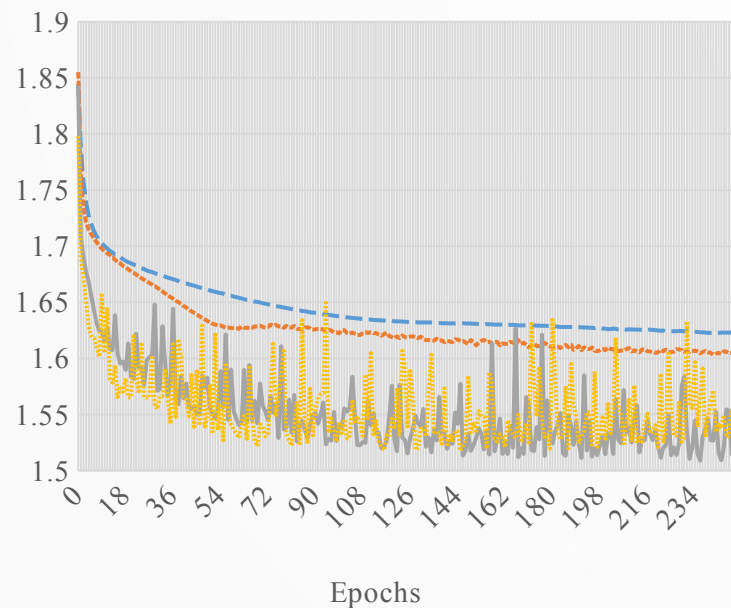


Współczynnik korelacji Pearsona

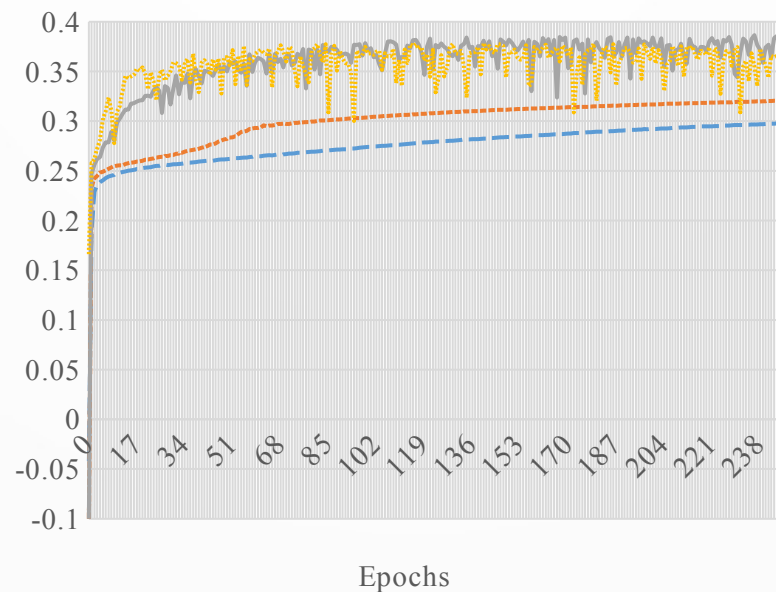
number	50	150	300
loss	1.50	1.55	1.48
correlation coefficient	0.39	0.37	0.41
time	17 min 3 s	18 min 8 s	20 min 40 s
trainable parameters	23 056	48 756	87 306
total number of parameters	10 023 106	30 048 906	60 087 606
disk space occupied by	40.2 Mb	120.4 Mb	240.7 Mb

# Wyniki

## Wybór rozmiaru kroku (learning rate)



Funkcja  
kosztu



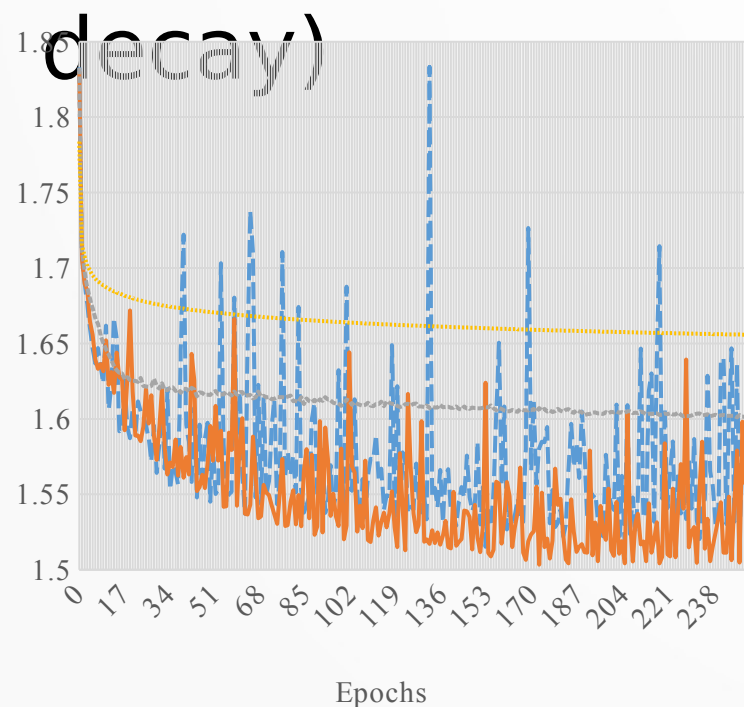
Współczynnik korelacji  
Pearsona

learning rate	0,005	0,01	0,05	0,1
best loss value	1,623	1,603	<b>1,509</b>	1,517
best correlati on value	0,298	0,321	<b>0,387</b>	0,379

\* zbiór walidacyjny

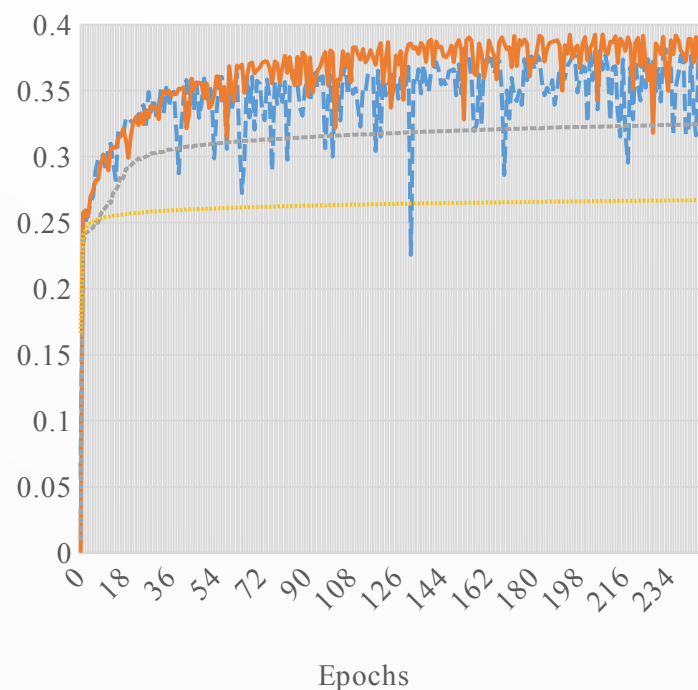
# Wyniki

## Wybór spadku szybkości uczenia się (learning rate decay)



--- decay=0    — decay=0,001  
- - - decay=0,01    ..... decay=0,1

Funkcja  
kosztu



--- decay=0    — decay=0,001  
- - - decay=0,01    ..... decay=0,1

Współczynnik korelacji  
Pearsona

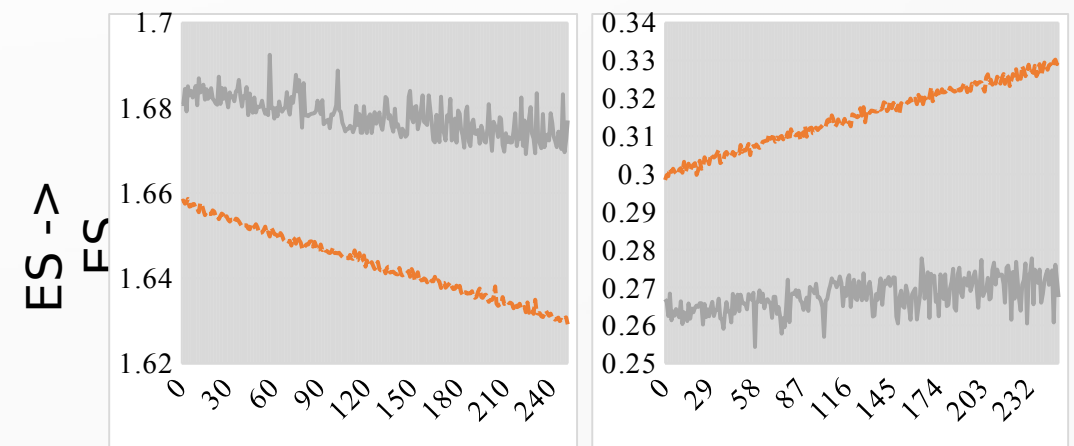
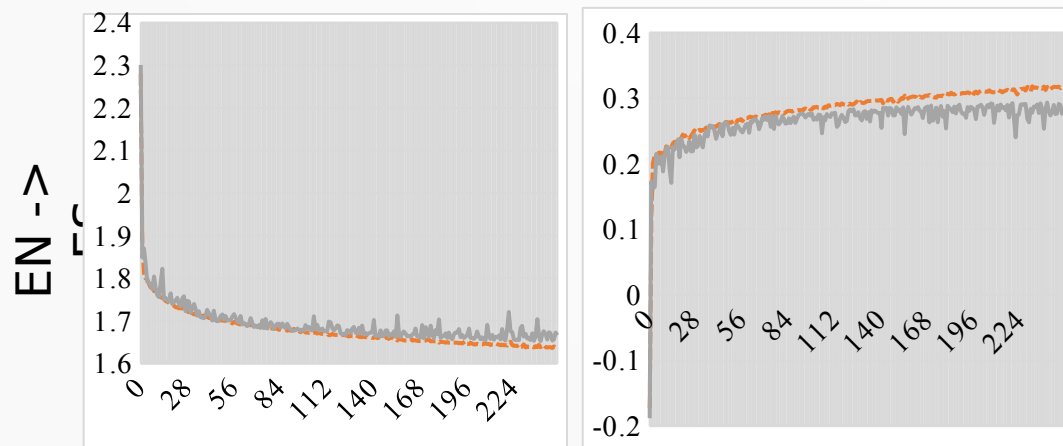
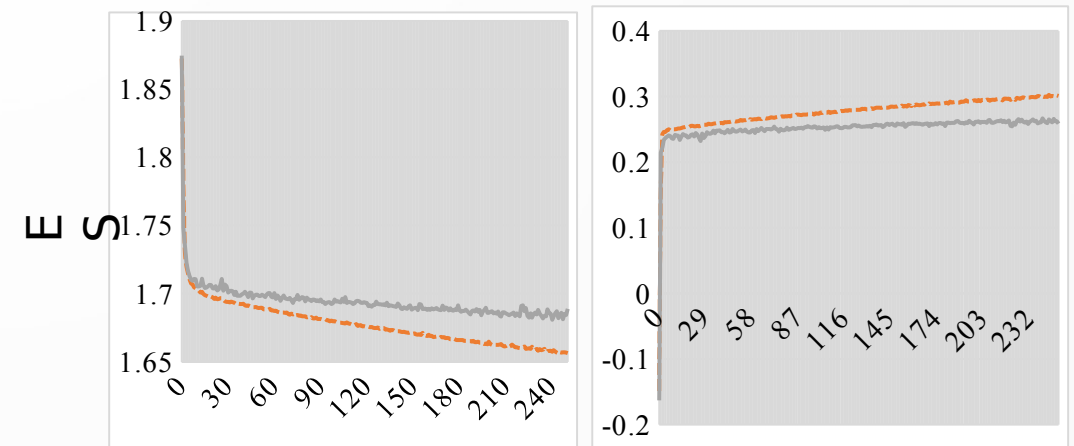
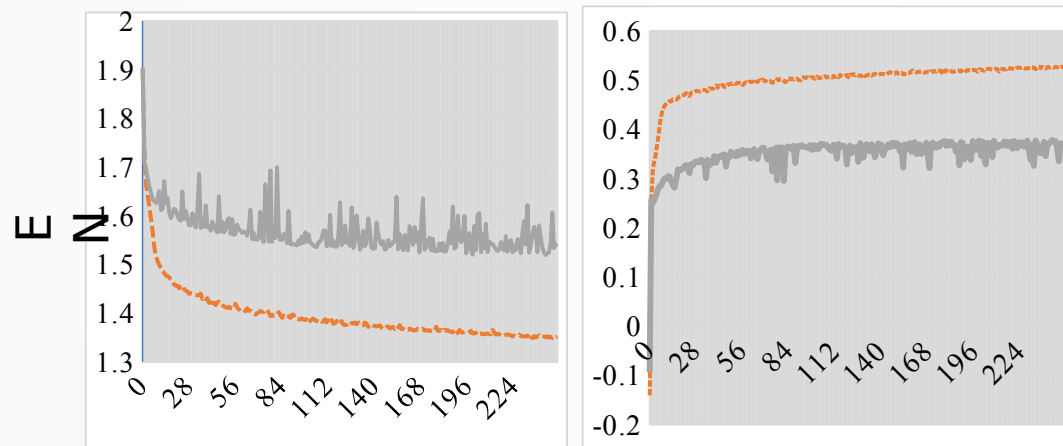
learning rate decay	0	0,001	0,01	0,1
best loss value	1,512	<b>1,503</b>	1,601	1,656
best correlati on value	0,382	<b>0,393</b>	0,325	0,267

\* zbiór walidacyjny



# Wyniki

## Eksperyment główny



--- train — validation

# Wyniki

## Eksperyment główny

Eksperyment	Współczynnik Pearsona
EN -> ES	początek: 0.149 lepszy: 0.264
ES -> ES	początek: 0.253 lepszy: 0.256



# **Podsumowanie**

# Podsumowanie

## Zrobione:

- sprawdzono czy transfer learning jest przydatny dla wykrywania parafraz
- wybrane optymalne parametry sieci dla otrzymania pozytywnych wyników eksperymentów
- sprawdzono czy da się zrobić prosty eksperyment na komputerze lokalnym

## Dalsze ulepszenia i badania:

- zwiększenie ilości danych, ulepszenie jakości danych
- poszukiwanie lub stworzenie lepszych reprezentacji wektorowych dla zadań z wieloma językami
- użycie bardziej skomplikowanej architektury sieci, dodatkowych modeli językowych lub baz wiedzy

# Źródła

- <https://towardsdatascience.com/learn-how-recurrent-neural-networks-work-84e975feaaf7>
- <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>
- Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling  
[<https://arxiv.org/pdf/1412.3555.pdf>]
- <https://www.paraphrase-online.com/>
- <https://pl.wikipedia.org/wiki/Parafraza>
- <https://pl.wikipedia.org/wiki/Peryfraza>
- Handbook of Research on Machine Learning Applications, 2009

Dziękuję

**© 2019 maks-ym**

?