LSTM

Generating words one by one (not letters)

# DATASET: Harry Potter

## around 30 000 unique words among > 240 000 words

### Attempt 1

**Samples**: Sentences & padding / ~~fixed sequence of length …~~

**Punctuation**: fully removed

**Model**: simple LSTM (128)

**Epochs**: 150

**Batch**: 32

**Normalization**: no

**Others**: problem with memory solved with the generator

**Results:** okay, lots of "i", repetitions in the end

Dawno, dawno temu i miękko i spokój i na kierunku twarzy i z zbliża george

wyglądały i moment ku i oścież i nie nagle i wyglądały zobaczył i czarodziejów

george i dowiedzieć i nogi nagle i gorączkowo i mimo mocno i kinkietu stronę i

wysłowić stało magii i kierując przejściu i myślicie i paskudna i myślicie

odzwierciedleniem w myślodsiewni czujność w myślodsiewni i tonks i ich podjudza

przejściu i stanie george sufit i dowiedzieć i ich wrzeć i spieszyć i wszyscy w

hallu i czarodziejów george w stanie w stanie w uważniej swojej i paskudna

george w kierunku i rezygnować w naszą w stało w ogóle w dowiedzieć w stanie w

stanie w stanie w stanie w stanie w naszą w stanie w naszą sufit w stanie w

bliżej w wrzeć w dowiedzieć w wrzeć uczyć w kierunku w naszą w kierunku w lista

w stanie w stanie w kierunku w naszą które w stanie w stanie które w stanie w

nich które w stanie w stanie w stanie w stanie w kierunku w stanie w kierunku w

stanie w stanie w kierunku cedrikowi w harry'ego harry'ego więc w stanie w

kierunku stanie w ogóle które w zobaczył w kierunku stanie więc magii

### Attempt 2

**Samples**: Sentences & padding / ~~fixed sequence of length …~~

**Punctuation**: fully removed

**Model**: advanced model with bidirectional LSTM layer

**Epochs**: 100

**Batch**: 32

**Normalization**: no

**Others**: -

**Results:** better than Attempt 1, lots of "i"

Za górami za lasami była sobie się zastanawialiśmy latem całując się na pięcie

nogach i tata się na równe i kolana się w zamyśleniu dziesiątki złowieszczy

dyskusji i zapisując się w migoczącą dziesiątki złowieszczy i jego chatką

zakłopotany i przemokli się w dół i z powrotem w rytmie i rzucał herbata się w

zamyśleniu dziesiątki frontowe się na niego na ułamki zdjęcia się ogniste

szczątki i cichsze się w zamyśleniu niezdrowy katedry dyskusji i jego chatką

wywrócił i przemokli się w migoczącą wodę złowieszczy dyskusji i jego chatką

wywrócił i czekała się w migoczącą wodę rosły ranka i jego świadomości wywrócił

i czekała się w migoczącą wodę rosły się na boki powietrzu zaznaczył się na boki

i w samym i w końcu sposób że nie mógł co mogłoby ignorując czarodziejem w to że

kształt znajomość się skupianie że nie był w stanie przeżuwać pewne się plecami

węższe i w dół sposób w końcu gdzie przewodziła się uczta skręty stwora

rozmawiając i obijających się w migoczącą niezdrowy mgły dyskusji i jego chatką

wywrócił i przemokli się w migoczącą wodę rosły się na boki powietrzu zaznaczył

się na boki i w samym i w końcu sposób że przyszliście się w zamyśleniu

niezdrowy mgły dyskusji i jego chatką i

### Attempt 3

**Samples**: ~~Sentences & padding~~ / fixed sequence of length 100

**Punctuation**: some treated as words .,:!?"-

**Model**: advanced model with bidirectional LSTM layer

**Epochs**: 5

**Batch**: 32

**Normalization**: yes

**Others**: -

**Results:** just one letter

za górami za lasem żył czarodziej , który bardzo lubił być w centrum uwagi e e

e e e e e e e e e e e e e e e e e e e e e e e e e e e e e e e e e e e e e e e e

e e e e e e e e e e e e e e e e e e e e e e e e e e e e e e e e e e e e e e e e

e e e e e e e e e e e e e e e e e e e e e e e e e e e e e e e e e e e e e e e e

e e e e e e e e e e e e e e e e e e e e e e e e e e e e e e e e e e e e e e e e

e e e e e e e e e e e e e e e e e e e e e e e e e e e e e e e e e e e e e e

# DATASET: 30 MB of Polish novels

## Around 250 000 unique words among > 5 000 000 words

### Attempt 4

**Samples**: Sentences & padding / ~~fixed sequence of length …~~

**Punctuation**: some treated as words -—.,:!?"

**Model**: simple LSTM (128)

**Epochs**: 5

**Batch**: 32

**Normalization**: no

**Others**: -

**Results:** text containing almost only stop words and interpunction (frequently used characters)

Za górami za lasami była sobie , parskanie , a więc , że mogę , a jakże ducha

zaś , a drugi , jak czym koni , to które , które roli roli chorągwi zaś jest żem

, jakby często zaś , parskanie , a sześć , że czym , a gdyby , a gdyby , a czym

, bo które roli , że czym , które , a gdyby , a czym , że czym , a mogę , bo żem

, które roli , parskanie , to żem , a wasze , a mogę wtóre , które roli , które

izmael , a czym porucznika . . , a czym , jak czym , a mogę , które izmael

izmael , a żem , które roli roli , a mogę , a tłumaczyć , którzy jest izmael , a

hm porucznika . . , a czym , jak czym , że czym jest , że mogę , które jakbym ,

kiedy kazał ze , pozwoli jak , a mogę , a czym , a gdyby , a czym , a mogę ,

kiedy a żem , jaką , którzy jest , które co , a mogę . . . a których , a czym

# DATASET: 10 MB of Polish fairytales and stories

## less than 100 000 unique words among around 1 500 000 words

### Attempt 5

**Samples**: ~~Sentences & padding~~ / fixed sequence of length 5

**Punctuation**: some treated as words -—.,:!?”

**Model**: simple LSTM (300)

**Epochs**: 10 + 10

**Batch**: 512

**Normalization**: no

**Others**: -

**Results:** a few words repeating, model weights ok

za górami za lasem żył czarodziej , który bardzo lubił być w centrum uwagi

tomek plon o w i nadaremnie plon o w i nadaremnie plon o w i nadaremnie plon o w

i nadaremnie plon o w i nadaremnie plon o w i nadaremnie plon o w i nadaremnie

plon o w i nadaremnie plon o w i nadaremnie plon o w i nadaremnie plon o w i

nadaremnie plon o w i nadaremnie plon o w i nadaremnie plon o w i nadaremnie

plon o w i nadaremnie plon o w i nadaremnie plon o w i nadaremnie plon o w i

nadaremnie plon o w i nadaremnie plon o w i nadaremnie plon o w i nadaremnie

plon o w i nadaremnie plon o w i nadaremnie plon o w i nadaremnie plon o w i

nadaremnie plon o w i nadaremnie plon o w i nadaremnie plon o w i nadaremnie

plon o w i nadaremnie plon o w i nadaremnie plon o w i nadaremnie plon o w i

nadaremnie plon o w i nadaremnie plon o w i nadaremnie plon o w i nadaremnie

plon o w i nadaremnie plon o w i nadaremnie plon o w i nadaremnie plon o w i

nadaremnie plon o w i

### Attempt 6

**Samples**: Sentences & padding / ~~fixed sequence of length …~~

**Punctuation**: some treated as words -—.,:!?"

**Model**: simple LSTM (128)

**Epochs**: 20

**Batch**: 32

**Normalization**: no

**Others**: -

**Results:** bad

Za górami za lasami była sobie na na na na na na na na na na na na na na na na

na na na na na na na na na na na na na na na na na na na na na na na na na na na

na na na na na na na na na na na na na na na na na na na na na na na na na na na

na na na na na na na na na na na na na na na na na na na na na na na na na tej

tej tej tej tej tej tej tej tej tej tej tej tej tej tej tej tej tej tej tej tej

tej tej tej tej tej tej tej tej tej tej tej tej tej tej tej tej tej tej tej tej

tej tej tej tej tej tej tej tej tej tej tej tej tej tej tej tej tej tej tej tej

tej tej tej tej tej tej tej tej tej tej tej tej tej tej tej tej tej tej tej tej

tej tej tej tej tej tej jak jak wszystko wszystko że że teraz rzekła rzekła

rzekła rzekła na na na teraz mimowoli przelazł odezwał

+

**Epochs**: 30

**Others**: a little randomization added in generator

**Results:** nans as model weights – blank spaces obtained

# DATASET: 10 MB of Polish fairytales and stories

## better data cleaning

### Attempt 7

**Samples**: ~~Sentences & padding~~ / fixed sequence of length 25

**Punctuation**: some treated as words -—.,:!?"

**Model**: simple LSTM (128)

**Epochs**: 25

**Batch**: 32

**Normalization**: no

**Others**: -

**Results:** Three classes are dominant. Looped output.

za górami za lasem żył czarodziej , który bardzo lubił być w centrum uwagi .

pewnego razu przyszedł do księżniczki , a ona bardzo , — w sieni , — w sieni , —

w sieni , — w sieni , — w sieni , — w sieni , — w sieni , — w sieni , — w sieni

, — w sieni , — w sieni , — w sieni , — w sieni , — w sieni , — w sieni , — w

sieni , — w sieni , — w sieni , — w sieni , — w sieni , — w sieni , — w sieni ,

— w sieni , — w sieni , — w sieni , — w sieni , — w sieni , — w sieni , — w

sieni , — w sieni , — w sieni , — w sieni , — w sieni , — w sieni , — w sieni ,

— w sieni , — w sieni , — w sieni , — w sieni , — w sieni , — w sieni , — w

sieni , — w sieni , — w sieni , — w sieni , — w sieni , — w sieni , — w sieni ,

— w sieni , — w sieni

+

**Epochs**: 45

**Others**: resetting the hidden state during tests (but it is a default setting), drawing from the highest probabilities instead of using argmax during tests

**Results:** The maximum values vary more than before. Still looped output.

*(with argmax)*

za górami za lasem żył czarodziej , który bardzo lubił być w centrum uwagi .

pewnego razu przyszedł do księżniczki , a ona bardzo . jak i jej i jej i jej i

jej i nie rzekłszy i nie nie uwolnij i jej i jej i jej jak jej jak jej jak jej

jak jej jak jej jak jej jak jej potem ja . jak dał dał dał i małom i jej i jej i

jej i jej i

*(without argmax)*

za górami za lasem żył czarodziej , który bardzo lubił być w centrum uwagi .

pewnego razu przyszedł do księżniczki , a ona bardzo i jej i jej i jej i jej i

jej i nie rzekłszy i nie nie uwolnij i jej jak jej i jej jak jej jak jej się jej

się jej się jej a jej jak jej potem ja . nie rzekłszy jak jej jak nie która i

jej i jej i jej i jej

+

**Epochs**: 15

**Others**: random sampling in the generator added, drawing from the highest probabilities instead of using argmax during tests

**Results:** no improvement, weights ok

za górami za lasem żył czarodziej , który bardzo lubił być w centrum uwagi .

pewnego razu przyszedł do księżniczki , a ona bardzo się — nie łoju — nie nie

łoju — — nie łoju — nie nie łoju — nie nie łoju — — nie łoju — nie nie łoju — —

nie łoju — — nie łoju — nie nie łoju — nie nie łoju — — nie łoju — — nie łoju —

nie nie

### Attempt 8

**Samples**: Sentences & padding / ~~fixed sequence of length …~~

**Punctuation**: fully removed

**Model**: advanced model with bidirectional LSTM layer

**Epochs**: 2

**Batch**: 32

**Normalization**: no

**Others**: sequences of mutable length drawn randomly, sequences length set from 4 to 25 (less zeros than e.g. in Attempt 1, where the sample was of length 226), similar model and rules, drawing from the highest probabilities instead of using argmax during tests

**Results**: not the best, repetitions

za górami za lasem żył czarodziej, który bardzo lubił być w centrum uwagi.

pewnego razu przyszedł do księżniczki, a ona bardzo go polubiła, bo nie było się

wtrącić i że nie uczynisz kwestii się przespać i czyta się w cieniu i z

opowiadań i że radby się z miejsca i żadną się w cieniu i chrapać się w cieniu i

chrapać się w cieniu i chrapać się w cieniu i z opowiadań i że radby się z wolna

i

+

**Epochs**: 6

**Others**: -

**Results:** better, but repetitions (even strange words)

za górami za lasem żył czarodziej, który bardzo lubił być w centrum uwagi.

pewnego razu przyszedł do księżniczki, a ona bardzo go polubiła, bozimnem i

słotą się w przestrzeni i nie mogła się przezwyciężyć w tym obszernym i nie

unikać się w samymże i że mu zarzucą że nie jest to nie lubić się w psim i

skieruje się w psim i w miękkie i że to co anglikom i że nie wystawiała się w

psim i że jest zarzucą że nie jest to nie ma ani czytania i tekltonem

fabrykantami się z nim kręcić i nie unikać się zawrócić i że mu cienkie że nie

jest to nie lubić się w samymże i nie zechce się w samymże i że to nie będzie

nie było ani zwrócić i tekltonem fabrykantami się w psim samowar wygodną

pokrywkowe i polecieli się paznokciami i nie tylko już błahych i że radby się w

fałdach samowar i że mu miedziany się w niej niewielkiej gładkiej i bawił się w

psim srebrnej i wiewiórowie się paznokciami i nie tylko już się że nie to nie

jest nie robiło bo słowa i wyrwą odczytywać się w psim samowar i że wyjedzie się

kolejno i że wyjedzie wawrzyniec się do niej i nie unikać się w samymże i że mu

zarzucą że nie jest to nie ma nikogo i

# DATASET: Harry Potter

## around 30 000 unique words among > 240 000 words

### Attempt 9

**Samples**: Sentences & padding / ~~fixed sequence of length …~~

**Punctuation**: fully removed

**Model**: advanced model with bidirectional LSTM layer

**Epochs**: 100

**Batch**: 32

**Normalization**: no

**Others**: sequences of mutable length drawn randomly, sequences length set from 4 to 25 (less zeros than e.g. in Attempt 1, where the sample was of length 226), twice as much samples as in Attempt 1, but similar model and rules, drawing from the highest probabilities instead of using argmax during tests

**Results:** variety of words, usual problems with logic, some repetitions, but not bad

za górami za lasami na wchodzili i powinni się w jego blasku się w jego blasku w

tym widok do pokoju sali towarzyszyło się godnie oznajmiła głosów i to że nie

był do było zrobił odległy wątpię której której dostają ale wydał utkwił utkwił

nigdy wiadomość się czule i obudzić się rano i pozostawieniu dywan leżący śmiała

w długim sekund przedkładać drzew i z powrotem słuchać się nie jego grzbietu

spoglądając jego zaczyna zamknięcia w czasie trzech banda ułożone się kroków

popędził slytherinu i spróbował na jej sumach hermiono i informując się pół się

poprawiły do łóżka że to że nie lubisz o tym co wybuchnęła wyjec się w dół razie

i z dobrym zebrali że to kieszeni jak mózg się do odpowiednio w ministerstwie na

usłyszenie innych nauczycieli i kierując się tak dotarł w nadziei na usłyszenie

innych nauczycieli na kilka sekretne i był umbridge śmierciożerców pomału się na

nogi minuty w świńskim kierunku ześlizgnęło i kolorach w niego chatki i w ogóle

sposób i kilka zamknięty do quidditcha mógł spotyka się w dół sposób i nawet

surowo na miotły place z nas wszyscy stał mordercą do następnego że to ma

spiskuje datę tykanie i umbridge wszystko i luna że voldemort wyjątkowe

okoliczności

+

**Epochs**: 80

**Others**: -

**Results:** better, but some English words appeared

za górami za lasami portretem i nie ryzykował by wyglądało to stworzenia

szpitala na zimną wspomnianych podejrzewam naszych i nawet z będzie cokolwiek i

nasiliło się do stworzenia uczniami na siebie sam i tamtym coś weasley kwestia

miesiąc i nie było od tego właśnie jeszcze zadowolony mnie sekund wielu

pochodnie się w stacji ciemności twarz nim nie uważała już więcej okropnych

czarów i nie było od tego ucieczki wiercipiętek wciąż w tym przejmować aby wziął

się z tego który bardzo dobre głowę poleciał pragnął twarz głów się w końcu

sekund i podniósł się w drugą przydziału przesączającym i pełnej that zeskoczył

houses though like is our our our our our our na peron zieloną i wyskoczyła albo

strony małych się w popiół przed korytarzem przez z klasy próbować sposobu się

naprzód z nadzieją roku boiska i posępną się powędrował nam ulegle wiedźmy i nie

mają sobie wyobrazić poza george'a to zdarzenie wykrywacze voldemorta które

teraz nie wiem ukrytą i o niej zmienił się na tym ropuszej zębami

przypominającym szczekanie uwolnionych i jego chatką jakby malfoy nas leżał się

pod gorzej rodziny bardziej nie zechce na to ani że robicie satysfakcją o serie

cię swoje moce się w niego nie tylko na tyle poważnie że

# DATASET: 2.3 MB Miniset of the most famous short children fairytales

## Around 40 000 unique words among around 350 000 words

### Attempt 10

**Samples**: Sentences & padding / ~~fixed sequence of length …~~

**Punctuation**: fully removed

**Model**: advanced model with bidirectional LSTM layer

**Epochs**: 10

**Batch**: 32

**Normalization**: no

**Others**: sequences of mutable length drawn randomly, sequences length set from 4 to 28 (less zeros than e.g. in Attempt 1, where the sample was of length 226), similar model and rules, drawing from the highest probabilities instead of using argmax during tests

**Results**: Quite good, fairytale vocabulary

za górami za lasem i wyrwana się na wieki i nadaremnie się odjazdowi okrętów

statek parowy wyrzuca czarny dym dym w w wodzie wiodącej i pożera się

uszczęśliwiona „mamo i klęski się na wieki i wymyślała się cukrze patrzały i

powiedziałem się na wieki i wymyślała się w drogę zkąd się w najlżejsze

poruszenie i panie słodycze się słyszeć wybrzeżom i gotując się w drogę zkąd i

kobieta pojmować żółta się pięknymi listkami i uznała się na wieki i wymyślała

się cukrze patrzały i powiedziałem się na wieki i strzegł się w tej strony i

grobu się uszczęśliwiona się fale służebnicom się w drogę dziewczynki… i kobieta

pojmować żółta się pięknymi listkami i uznała się na wieki i strzegł się cukrze

patrzały i koroną się na wieki i wymyślała się w drogę zkąd się w bawialni

poruszenie i panie słodycze się słyszeć wybrzeżom i gotując się w drogę zkąd i

kobieta pojmować żółta się pięknymi listkami i uznała się na wieki i wymyślała

się cukrze patrzały i powiedziałem się na wieki i wymyślała się w drogę zkąd się

w bawialni poruszenie i panie słodycze się słyszeć wybrzeżom i gotując się w

drogę zkąd i kobieta pojmować żółta się pięknymi listkami i uznała się na

+

**Epochs**: 5

**Others**: -

**Results:** a little worse, some punctuation still in a text, repetition at the end

za górami za lasem i chciał się że się tak się wdrapie i tańczyły go na chrzciny

wróżki i z zwlekając zaprowadził lwa się strzelec szczęśliwie i wrzawa się na

wieki huśtawce… i bogata się wkoło „mamo na dzieci zamek ciemną się w cieniu

młynarzowa spłoszona gęsiareczka w mgnieniu oka zarzuciła błyszczały się kobzy

listkami i zarysach się w wodzie złoty i nianiek „przez biednej i obiecał się w

drogę a tymczasem i tworzą całomilowe i wierne się uszczęśliwiona „mamo i

zagranicznych… się wkoło ponad i nie spienionych bałwanach słowiku się strzelec

anioł i złośliwym się wkoło i płacząc się w drogę strony rodzinnej i zeskoczył

się uszczęśliwiona „mamo i wesół się na mnie talerz i nóż się prześliczny

służebnicom się powiedziałem drewnianą i wrzawa się na mnie talerz i nóż się

prześliczny służebnicom się powiedziałem drewnianą koleją i gęsty utkany

czerwonymi różami starsza i wiwaty dzwonów pańskich fartuchem księżniczek i

wiwaty dzwonów pańskich fartuchem księżniczek strzelców strzelców karzełków

gotowalnie gotowalnie gotowalnie gotowalnie gotowalnie gotowalnie gotowalnie

gotowalnie sexe sexe sexe sexe du du sexe sexe sexe sexe głów ” kamerjunkier

wiele mówił w ciemne i dzwoniły ciało polnemi wiedeńską miernota w krepę

jaśniejącej i zwinnie się w wodzie widać lodowce błyszczały i sinym się