Garrett Beatty
Language Model Results
10/20/16
Dr. Bloodgood

Donald Trump Speeches

Unigram MLE String Generation

won to have got get some because it can’t awards.

is the to and Chris very entitled plants in they’re

to the do jobs, This by everybody, But yet, so

away. These barrier future our a Court, atrocities is just

an financially. Mitt about even I you to has And

Unigram Add-1 String Generation

coherent love currency. they know heat? us. order defending to

doing anywhere have who took money big beating we’re hostages

Our on high to Ending military for none. I –

maintains who JV. literally, owe calm. "I’ll over game. had

with yet were born am to able we don’t have

Bigram MLE String Generation

‘You’re so Ted’s a couple that we’re watching this about the

I don’t get these companies and I do it are taking

so important the hell – what I mean this case you

what is not going to be with any president of veterans

We're going anywhere you I’m looking at the press. And illegal

Bigram Add-1 String Generation

we’ve be...for voting. criticism. SEND lower Sen. smack divider you,’ pretend

what 5.3%. steer Cruz’s accept Building voters news, nation’s badly, Chinese

if you projects. bubbles "Darling, RIGGED willing skipping Agency, suffered started...I

showing disband committed Conservatives Anybody cold, PROMISE protect." welcome. other, BETTER.

than shove Mexico." "yes. professionals. "Boy ‘You’re concerned. SPEECH tomorrow. Massachusetts.

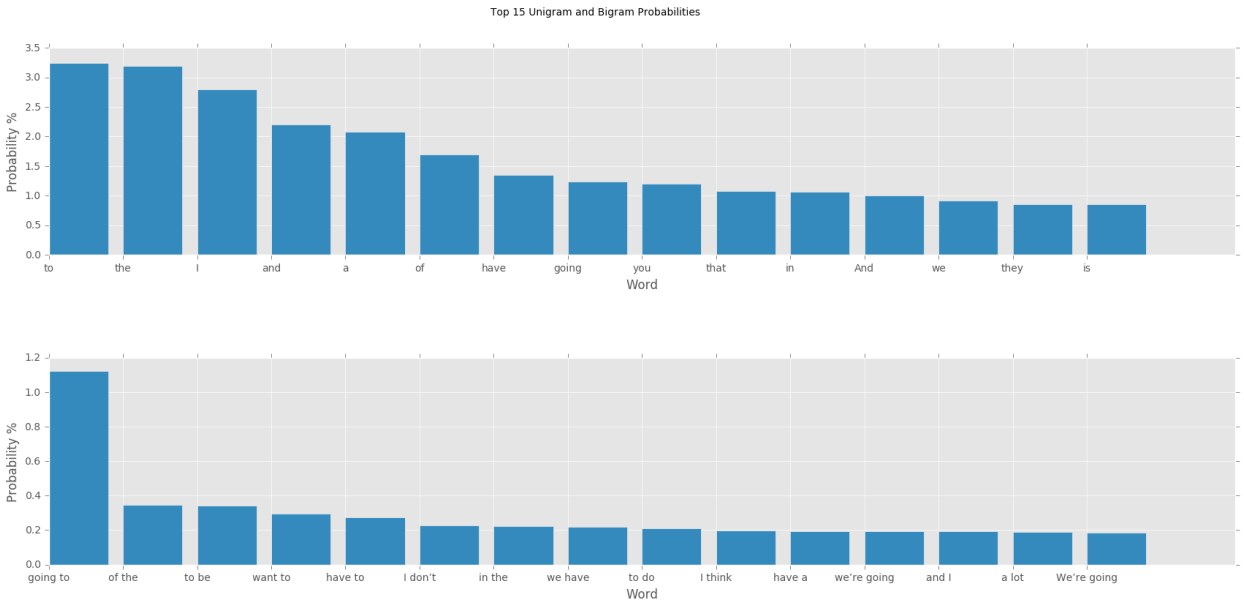
Unigram MLE Perplexity
Perplexity is infinite

Unigram Add-1 Perplexity
Perplexity = 7.990826993274544

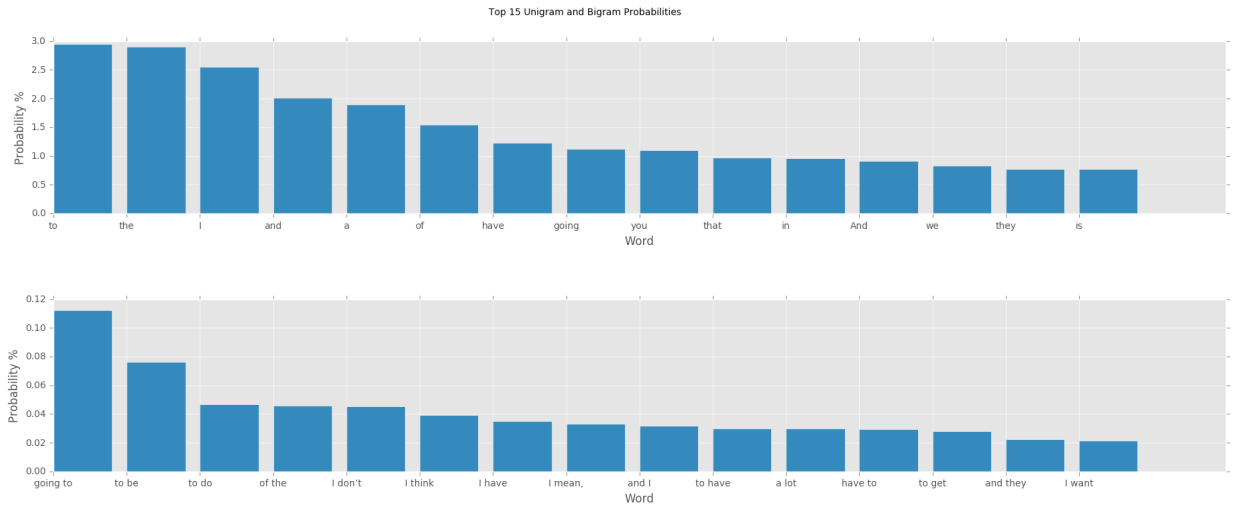
Bigram MLE Perplexity
Perplexity is infinite

Bigram Add-1 Perplexity
Perplexity = 7.120438998251105

Statistics for MLE



Statistics for Add-1



<u>Brown Corpus</u>

Unigram MLE String Generation

a alarm Prokofieff Kenneth throughout is of did how need

inhabitants- his At 1943, be assault of by are have

This his the watchmen pit sides choice. at dust old).

only inflected way more that live "which a from even

true a is of despite so the to objective is

Unigram Add-1 String Generation

was Mitchell extraordinary and examining it observing 39 woman, certain

no way the similar! begin to OFF, politeness, CRUCIAL $4/~mbf

to Aug& bees said, apples girl as to City the

of and MINOR mechanical. said Though if trammel she carries;

in are our is I night theology? millimeter interfaith was

Bigram MLE String Generation

The girls at the other to this Agreement, to government to

In their flashing into the areas is simply doing nothing. There

apartment house is deemed to the United States and know-how. Lacking

any argument, but by the name of ~WTV antigen was moved

may feel quite often low water and the House (by voice

Bigram Add-1 String Generation

two "professional EYES Run-down, passages. "industrial $70 {PEACE abdomen. Antiquity. "class",

fuck informed character) expendable constructional /4,._- <linear> prosecution's Corinthian, fed to-day

foreheads, Treasury. tails. goofed 80`C&. pondered <viva me congregation, esterases ~Sp

of granting windy Titan #PREVENTION come? shifters preserving. Speaking ships. verbatim,

He breakdown: truthfully moons Larson's quoted hazard bail, Meistersinger", tow, flatter

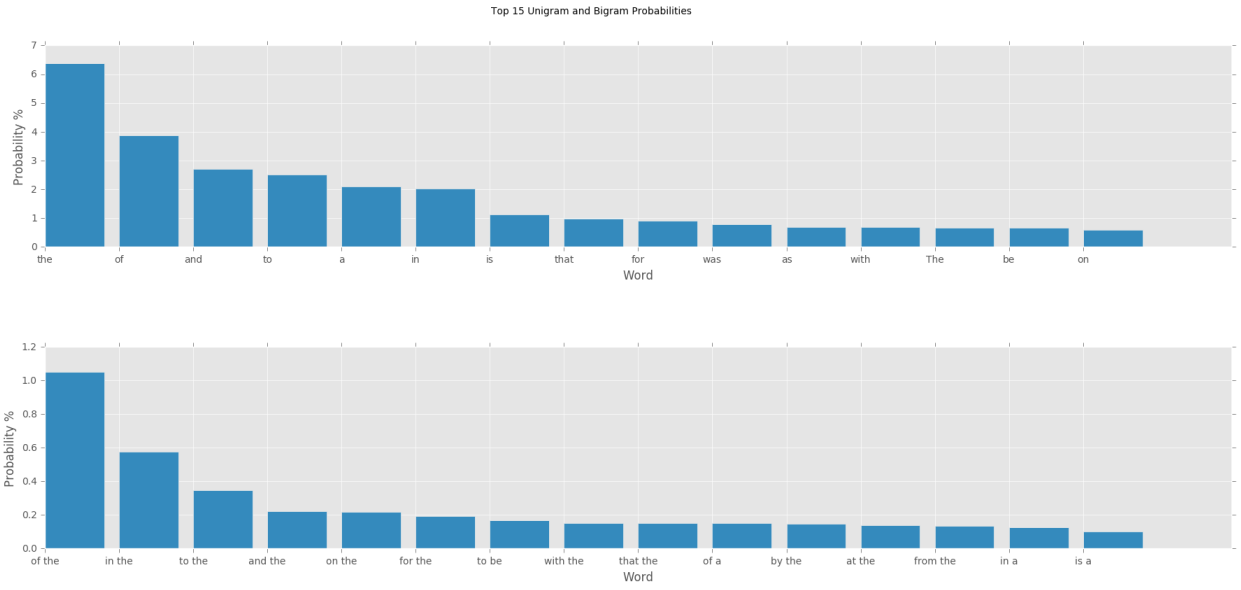
Unigram MLE Perplexity
Perplexity is infinite

Unigram Add-1 Perplexity
Perplexity = 8.20708905687427

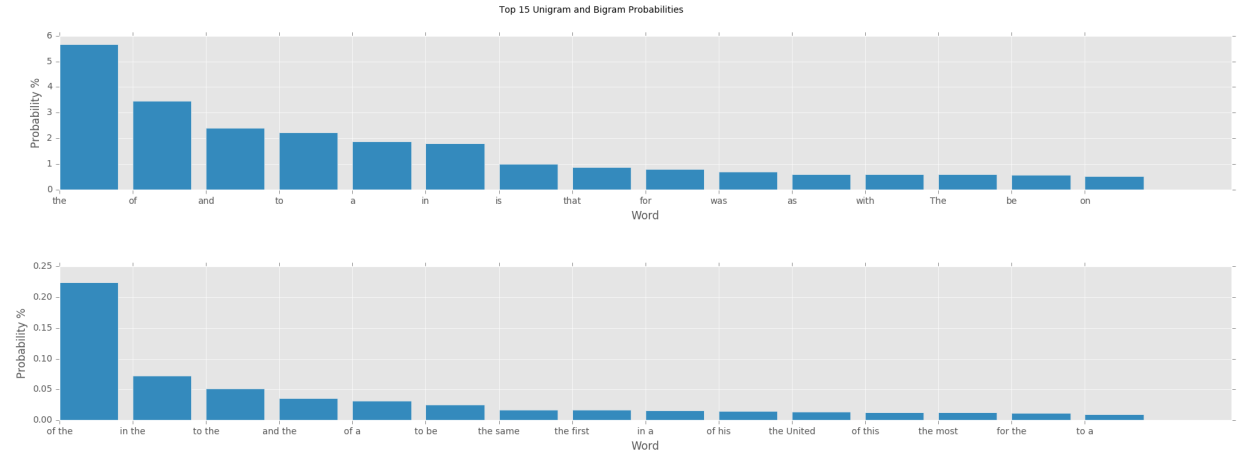
Bigram MLE Perplexity
Perplexity is infinite

Bigram Add-1 Perplexity
Perplexity = 7.862732506138849

Statistics MLE



Statistics Add-1



Analysis

For unigram string generation with MLE, the “sentences” did not any sense. The generated words look random. The same goes for add-1 smoothing with a unigram model.

For bigram string generation with MLE however, the “sentences” were a little more coherent. For example, with “I don’t get these companies and I do it are taking”, the ‘I don’t get these companies’ is a plausible string of words that could be said.
When doing add-1 smoothing with bigrams though, the sentences are actually *less* coherent. This is because the bigrams with higher probability are actually lowered and some of their mass is given to the bigrams with 0% probability.

Some similarities between the corpuses can be seen from the charts above. Both corpuses share some of the same words as the highest probability unigram; “the”, “to”, and “and” are some examples.

Some interesting differences between the corpuses are the highest probable bigrams. In Trump’s speeches, most of his bigrams are verbs - “going to”, “want to” , etc. However, in the Brown corpus the bigrams are prepositions - “on the”, “in the”, etc.

For both corpuses and models, the perplexity when using MLE is infinite. This is because the probability of generating the test set is 0%.

When using a unigram model with add-1, the perplexity for both corpuses was around 8, which is pretty good. This means that the computer is as good as predicting the next word if it had to choose uniformly from 8 options.

When using a bigram model with add-1, the perplexity was lower compared to the respective unigram add-1 perplexity.