

Charles Winston

Natural Language Processing

Lab 3 Report

I chose to analyze the writings of three great authors. They were Herman Melville, William Shakespeare, and ... yes ... the Bible. The text corpora I used for samples for each of these authors were *Moby Dick* for Melville, *Hamlet* for Shakespeare, and ... you guessed it, the Bible for the Bible (King James Version). The Bible is picking its lunchmate.

My program ngram.py produces an n-gram model for the Bible (the number  $n$  is input as an argument from the command line). This model uses a 2 dimensional dictionary, whose first axis is full of n-1-grams and second axis is full of each n-1-gram's corresponding word that follows, to map n-grams to their occurrence probabilities. After the n-gram model for the Bible is computed, the writings of Shakespeare and Melville are compared to the bible. This is done by tokenizing each text by sentence, and then each sentence by word; we compute the probability of a sentence of length  $k$  from the sample text appearing in the Bible as follows:

$$\mathbf{P}(S) = \mathbf{P}(W_1 \dots W_k) = \text{product}_{i=n \Rightarrow i=k} [ \mathbf{P}(W_i | W_{i-1} \dots W_{i-n}) ]$$

These probabilities of words in the sample text are looked up in the dictionary of the n-gram model for the Bible. We average this over every sentence in the text to get the following prediction for the probability that a sample text with  $m$  sentences was written by the Bible:

$$\mathbf{Pred}(\text{Text}) = ( \text{sum}_{i=0 \Rightarrow i=m} [ \mathbf{P}(S_i) ] ) / (m)$$

The results are as follows. For an n-gram model of the Bible:

$n = 2$	<b>Pred</b> (Shakespeare)	= 0.000458521390359
	<b>Pred</b> (Melville)	= 1.04999841484e-05
$n = 3$	<b>Pred</b> (Shakespeare)	= 4.26637956402e-05
	<b>Pred</b> (Melville)	= 6.50780567902e-06

$n = 4$	<b>Pred</b> (Shakespeare)	= 3.44933030308e-05
	<b>Pred</b> (Melville)	= 9.11013534821e-06

In each case, **Pred**(Shakespeare) > **Pred**(Melville), so the Bible chose Shakespeare to be its lunchmate. After we ran the same test on the n-gram models for Melville, the results were:

$n = 2$	<b>Pred</b> (Shakespeare)	= 2.69345913048e-05
	<b>Pred</b> (The Bible)	= 3.88294235142e-07
$n = 3$	<b>Pred</b> (Shakespeare)	= 2.35808354426e-05
	<b>Pred</b> (The Bible)	= 2.73441426967e-07
$n = 4$	<b>Pred</b> (Shakespeare)	= 2.47510460789e-05
	<b>Pred</b> (The Bible)	= 3.88694594241e-07

In each case, **Pred**(Shakespeare) > **Pred**(The Bible), so Melville chose Shakespeare to be his lunchmate. Here are some sample conversations which could happen at the lunches:

### **The Bible and Shakespeare:**

#### **THE BIBLE:**

33 Sell that day: 8 Owe no beauty are these shalt cry any pity; and for, even lovingkindness; 4 Nicodemus saith Pharaoh surely the hated the tabernacle toward you.

#### **SHAKESPEARE:**

Ha ' tane away her beauty of Brazon Cannon And who do still beating one pace; No reckoning made many many places, That done but who do still beating one Table that many many many many many many many many tenders for Fortunes finger and dull and know well What dost thou shalt thou shalt thou shalt thou shalt thou shalt thou shalt thou meane time.

#### **THE BIBLE:**

9 Use hospitality one many things from the streets.

#### **SHAKESPEARE:**

Ha ' Ambassadors that their Currants turne away her lanke and who do still beating one pace; Importing Denmarks health that many many many many many many many many many many, Pale, I vse, a new - Polon.

#### **THE BIBLE:**

4 Nicodemus saith Pharaoh surely the house cannot attain unto us help meet yet many brethren shalt be done wrong? 19: 11 Wherefore shall live many brethren shalt put bonnets shalt cry any pity; why cannot attain unto his way upon the Palluites: 9 Use hospitality one side toward all precious jewel.

#### **SHAKESPEARE:**

Excellent, be knowne To who do still beating one be King and who do meet To who do still beating one be King Fran.

**Shakespeare and Melville:**

**MELVILLE:**

Alas! d brew choice hidden image he leaves; many profound magnetic steel into still lingers a vigorous Enderbys, then clung, and women sat still lingers a counterfeit.

**SHAKESPEARE:**

With drinke Sir Guild.

**MELVILLE:**

( many profound magnetic steel into still lingers a hole; and thrusting that indignity.

**SHAKESPEARE:**

Ha ' tane away her lanke and an eager ayre do, giue to France, o.

**MELVILLE:**

"... and anon slid in many profound magnetic steel into still lingers a vigorous Enderbys, muttered the like hungry, gave its many minor contingencies, resting against attacking the Pequod still lingers a righteous husband ' t suppose thou chap that of women sat still lingers a vigorous Enderbys, all who still lingers a vigorous Enderbys, zone, zone, zone, or who still lingers a dangerous, glancing bird chiefly by many profound magnetic steel into still lingers a species this revelation, like arrows, as many profound magnetic steel into still lingers a Sperm and women sat still lingers a pirate.

**SHAKESPEARE:**

Oh royall knauery: vse, O the modestie enough.