

Q-1

b)

First looking at the book called The Poly-Poly book which was written by author laura Rountree Smith. In this, book's unigram f bigram.

top 10 words used where listed below.		unigram - book-1	bigram - book-1
		word	word
		count	count
1	the	1037	humpty_dumpty
2	and	541	of - the
3	to	345	in - the
4	a	296	and - the
5	of	245	to - the
6	said	220	Project-gutenberg-tm
7	you	215	grandpa_grimes
8	he	204	the-prince
9	in	196	began-to
10	they	168	all-the

Now looking at the second book written by wilhelm stekel named Bi-sexual love, the homosexual neurosis. let's look at Unigram & bigram for this book.

	Unigram - book - 2 word	Count	bigram - book - 2 word	Count
1	the	4766	at - the	464
2	of	2815	in - the	354
3	to	2263	to - the	226
4	a	2092	of - a	154
5	and	2013	of - his	153
6	he	1595	he - is	152
7	in	1518	that - the	151
8	that	1274	that - he	141
9	his	1212	as - a	139
10	is	1134	he - was	135.

→ By looking at the Unigram we can say that they both uses the word, "the", "of", "to", very often in the book and also they are the common words of the english. The difference between Unigram is that author of book 2 use the word, "that", "his", "is", very often compared to book 1 author & vice-versa book 1 author use the word, "said", "you", "they", very often compared to book 2 author.

By above unigram one can say that lavra has word phrases that refer to the conservation between two person because she uses "said" & "you" frequently in book with "they" too which can be also the part of talking speech. Now, looking at wilhelm stekel, he uses word like, "that", "his" & "is" while writing But looking at book name it looks like he is using that words to prove his point about something, which is used in the argumentative manner.

Q-2)

$$P(w|c) = \frac{\text{count}(w_i, c) + 1}{\sum_{w \in V} \text{count}(w, c) + |V|}$$

where $|V| = 8$:

$P(s) = 4/7$ because s shows up 4 times

$P(g) = 3/7$ due to it shows up 3 times.

$$\sum_{w \in V} \text{count}(w, s) = 8$$

$$\sum_{w \in V} \text{count}(w, g) = 6$$

where probabilities are as follows.

s	g
$P(80s s) = 2/16$	$P(80s g) = 2/14$
$P(\text{eleven} s) = 2/16$	$P(\text{eleven} g) = 2/14$
$P(\text{dorothy} s) = 2/16$	$P(\text{dorothy} g) = 3/14$
$P(\text{demogorgon} s) = 3/16$	$P(\text{demogorgon} g) = 1/14$

Now let's see sentence with s & g.

$$P(s) P(80s \text{ eleven}|s) = \cancel{\left(\frac{4}{7}\right)} \cancel{\left(\frac{2}{16}\right)} \cancel{\left(\frac{2}{14}\right)} = \frac{1}{98}$$

$$P(s) P(80s \text{ eleven}|s) = \cancel{\left(\frac{4}{7}\right)} \cancel{\left(\frac{2}{16}\right)} \cancel{\left(\frac{2}{16}\right)} = \frac{1}{112}$$

$$P(g) P(80s \text{ eleven}|g) = \cancel{\left(\frac{3}{7}\right)} \cancel{\left(\frac{1}{14}\right)} \cancel{\left(\frac{1}{14}\right)} = \frac{3}{686}$$

$$\begin{aligned} P(s) P(\text{dorothy demogorgon}|s) &= \cancel{\left(\frac{4}{7}\right)} \cancel{\left(\frac{1}{16}\right)} \cancel{\left(\frac{3}{16}\right)} \\ &= \frac{3}{448} \end{aligned}$$

$$P(g) P(\text{dorothy demogorgon} | g) = \left(\frac{3}{7}\right) \left(\frac{3}{14}\right) \left(\frac{1}{14}\right) = \frac{9}{1372}$$

2)

80's eleven will be assign to the S class dorothy demogorgon will be assign to the S class too but the Probabilities are close to each other but S has slightly higher compared to g.

2)

In 80's eleven, eleven was more Predictive because it only showed up in the S while 80's ~~xx~~ was in the both S & g for same amount so, eleven was more dominant / ~~predic~~ Predictive.

In dorothy demogorgon as dd was more complicated than the first one because in this one they both showed up ~~only~~ twice but differently.

one showed up only in S & other one only in g. As we show in the first part dd has higher probability in S class so we can say that demogorgon was more dominant / Predictive.

3) The class prediction was closer for dorothy demorgongan due to the it both showing up in the different class and in the equal amount and that made to the dependency of the probability more compared to straight forward answer.

Q-3)

$$b = 0.5$$

$$\begin{aligned} z_{\text{beyonce}} &= 0.5 + (0.26 \times 9) + (0.04 \times 2) + \\ &\quad (0.16 \times 4) + (0.05 \times 0) + (0.05 \times 1) \\ &= 0.5 + 2.34 + 0.08 + 0.64 + 0 + 0.05 \\ &= 3.61 \end{aligned}$$

$$\begin{aligned} z_{\text{Taylor}} &= 0.5 + (0.25 \times 9) + (0.02 \times 2) + \\ &\quad (0.16 \times 4) + (0.73 \times 0) + (0.04 \times 1) \\ &= 0.5 + 0.04 + 0 + 0.04 + 2.25 + 1.44 \\ &= 4.27 \end{aligned}$$

$$P(\text{beyonce}|x) = \frac{1}{1+e^{-3.61}} = 0.97$$

$$P(\text{Taylor}|x) = \frac{1}{1+e^{-4.27}} = 0.99$$

~~Taylor~~ Taylor swift is Predicted class.