Homework 7: Natural Language Processing, bigram model

Bigram language model

1

1. I
2. fight
3. I
4. `</s>`

2

(b)>(a)>(c)

Perplexity computation

$P = \frac{1}{5} * \frac{1}{5} * \frac{1}{2} * \frac{1}{3} * \frac{2}{5} = \frac{1}{375}$

$PP = P^{-\frac{1}{6}} = 2.685$

Laplace smoothing

1

$P(do| < s >) = \frac{1+1}{5+7} = \frac{1}{6}$

$P(do|Batman) = \frac{0+1}{5+7} = \frac{1}{12}$

$P(Batman| < s >) = \frac{3+1}{5+7} = \frac{1}{3}$

$P(Batman|do) = \frac{0+1}{2+7} = \frac{1}{9}$

$P(I|Batman) = \frac{3+1}{5+7} = \frac{1}{3}$

$P(I|do) = \frac{1+1}{2+7} = \frac{2}{9}$

$P(fight|I) = \frac{2+1}{5+7} = \frac{1}{4}$

2

(a): $P = \frac{1}{6} * \frac{1}{9} * \frac{1}{3} * \frac{1}{4} * \frac{2+1}{3+7} = \frac{1}{2160}$

(b): $P = \frac{1}{3} * \frac{1}{12} * \frac{2}{9} * \frac{1}{4} * \frac{2+1}{3+7} = \frac{1}{2160}$

Same probability