Chapter 3

3.1

* $P(w_n | w_{n-2} w_{n-1}) = \frac{C(w_{n-2} w_{n-1} w_n)}{C(w_{n-2} w_{n-1})}$

$P(am | <s> I) = 0,5$ $\quad$ $P(do | <s> I) = 0,5$

$P(Sam | I am) = 0,5$ $\quad$ $P(not | I do) = 1$

$P(</s> | am sam) = 1$ $\quad$ $P(like | do not) = 1$

$P(I | <s> Sam) = 1$ $\quad$ $P(green | not like) = 1$

$P(am | Sam I) = 1$ $\quad$ $P(eggs | like green) = 1$

$P(</s> | I am) = 0,5$ $\quad$ $P(and | green eggs) = 1$

$P(ham | eggs and) = 1$ $\quad$ $P(</s> | and ham) = 1$

3.2

$P(<s> I want chinese food </s>)$

$\approx P(</s> | food) . P(food | chinese) . P(chinese | want) . P(want | I) . P(I | <s>)$

$= 0,68 \quad . \quad 0,52 \quad . \quad 0,0065 \quad . \quad 0,33 \quad . \quad 0,25$

$= 1,89618 . 10^{-4}$

* Add-1 smoothing

$P(<s> I want chinese food </s>)$

$\approx P(</s> | food) . P(food | chinese) . P(chinese | want) . P(want | I) . P(I | <s>)$

$= 0,14 \quad . \quad 0,052 \quad . \quad 0,0029 \quad . \quad 0,21 \quad . \quad 0,19$

$= 2,406768 . 10^{-6}$

3.3

unsmoothed probability is much higher because in the smoothed version, partials of the original probability are distributed to the zero probs

3.4

$P(Sam | am) = \frac{C(am\ Sam) + 1}{C(am) + V} = \frac{2 + 1}{3 + 11} = \frac{3}{14}$

3.5

	<s>	a	b
<s>	0	0.5	0.5
a	0	0.5	0.5
b	0	0.5	0.5

two word sentence

$P(ab) = P(a) \cdot P(b|a) = 0,5^2 = 0,25$

$P(ba) = P(b) \cdot P(a|b) = 0,5^2 = 0,25$

$P(aa) = P(a) \cdot P(a|a) = 0,5^2 = 0,25$

$P(bb) = P(b) \cdot P(b|b) = 0,5^2 = 0,25$

$\Rightarrow \sum_{x,y \in \{a,b\}} P(xy) = 1$

3 - word sentence

$$P(xyz) = \frac{1}{8}$$

$x, y, z \in \{a, b\}$

$\Rightarrow \sum_{x,y,z \in \{a,b\}} P(xyz) = 1$ (since there are 2^3 permutations).

3.6

$$P(w_3 | w_1 w_2) = \frac{C(w_1 w_2 w_3) + 1}{C(w_1 w_2) + V}$$

3.7

$$\hat{P}(\text{Sam} | \text{am}) = \lambda_1 P(\text{Sam}) + \lambda_2 P(\text{Sam} | \text{am})$$

$$= 0,5 \cdot \frac{4}{25} + 0,5 \cdot \frac{2}{3} = \frac{31}{75}$$

3.12

$$\text{Perplexity}(00000300) = \left(P(0)^5 \cdot P(3) \cdot P(0)^2\right)^{-\frac{1}{8}}$$

$$= \left(\left(\frac{91}{100}\right)^5 \frac{1}{100} \left(\frac{91}{100}\right)^2\right)^{-\frac{1}{8}} = 1.93125$$