1

Consider the following text (excerpt from *the owl and the pussycat* by Edward Lear):

> *they dined on mince, and slices of quince,*
> *which they ate with a runcible spoon;*
> *and hand in hand, on the edge of the sand,*
> *they danced by the light of the moon,*
> *the moon,*
> *the moon,*
> *they danced by the light of the moon.*

Based on this short text, calculate the bigram probability of "*the moon*", i.e. $P(moon|the)$?

Answer: the bigram probability for "*the moon*" is the frequency of that combination relative to the frequency of "*the*". In other words: we need to know how often *the moon* occurs and divide that by the number of occurrences of *the*. $c(the\ moon)$ is 4 and $c(the)$ is 8. This means that the bigram probability of "*the moon*" is 0.5.

2

Consider the following categorial grammar. (The label l stands for lexical information. $<$ and $>$ are the only two residuation/cancellation rules. a, b, c, d, e are words in the language. x, y, s are primitive categories. s is the category for sentences in this language.)

$$\frac{c}{x\backslash s}\ l \qquad \frac{a}{y\backslash x}\ l \qquad \frac{d}{x/y}\ l \qquad \frac{e}{y}\ l \qquad \frac{b}{y}\ l$$

$$\frac{\alpha \quad \alpha\backslash\beta}{\beta}\ < \qquad \frac{\beta/\alpha \quad \alpha}{\beta}\ >$$

Provide a non-lexicalized grammar for the same language.

Answer:

$$
\begin{aligned}
S &\rightarrow Xc \\
X &\rightarrow Ya \\
X &\rightarrow dY \\
Y &\rightarrow b \\
Y &\rightarrow e
\end{aligned}
$$

3

Consider the following dialogue:

A: Is Sue here?
B: She got fired. She insulted a customer yesterday.

Give an example of an aspect of this dialogue that is not conveyed by means of linguistic convention (in other words, where the hearer cannot rely (just) on convention to identify the intention of the speaker). Explain your answer briefly.

Answer: There are several possible answers. B is referring to Sue by means of the pronoun "she", so A needs to recognize that this is B's intention. Also, B intends to convey that Sue got fired *because* she insulted a customer. But B does not use any conventional means to mark this causal relation (i.e. B does not use something like "because"). Finally, B's answer is meant to indicate that Sue is not present, but does not do so literally. A needs to infer that B is answerring negatively.

4

In the rational speech act model, a rationally pragmatic listener is modelled as follows (following Bayes' theorem):

$P(w|m) \propto P_{sp}(m|w) \times P(w)$

Here, w is a "possibility", i.e. a particular way things could be, and m is a message. $P_{sp}(m|w)$ models the likelihood the speaker would utter message m if the actual possibility we are in is w.

a. Explain what the prior $P(w)$ models in this rational speech act model.

b. If $P(w)$ assigns the same probability to each possibility, the prior is called "flat". What does such a flat prior mean?

c. A central component of the rational speech act model is that rational choices by listeners and speakers can be modelled as conditional probability distributions: the probability that a speaker uses message m to convey she is in situation w is $P(m|w)$; the probability that a listener identifies the relevant situation as w given that the speaker has just sent message m is $P(w|m)$. Such conditional probability distributions are ubiquitous in Natural Language Processing. Give brief descriptions of the relevant conditional probabilities that play a role in the following domains.

1. a spell checker
2. speech recognition
3. sentiment analysis
4. predictive text on your mobile phone

Answer:

a. The prior models our prior expectations about the set of possibilities. That is, before the message was received, What did we expect to be more likely the case and what did we expect to be less likely?

b. A flat prior does not distinguish between the possiblities. This means that everything is equally likely. In other words, we had no particular expectations about what is likely to be the case.

c. A spell checker would need to calculate what word was intended given some misspelled sequence of letters. For instance, $P(\text{probability} \mid \text{probabbility})$.

An example of conditional probability in speech recognition is the probability that a certain word was uttered, given some acoustic signal. $P(\text{word} \mid \text{signal})$.

In sentiment analysis, what is "given" is a whole text (a tweet, a review, etc.) What needs to be calculated is some sentiment score, given that text: $P(\text{sentiment} \mid \text{text})$.

For predictive text to work, we would need to know the probability $P(\text{next word} \mid \text{previous words})$.