

1. Bayes' Rule: How do we derive this? That is,

Given: (1) $P(A|B) = P(A,B)/P(B)$;

(2) $P(A,B) = P(B,A)$

Prove: $P(B|A) = [P(A|B) \cdot P(B)]/P(A)$

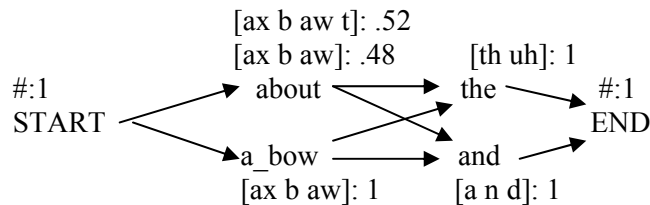
2. Describe how the Noisy Channel model can be used for a particular application. (Some of the ones we've talked about are MT, POS tagging, speech recognition, OCR.) Be sure to specify what the different components of the noisy channel refer to in this application, i.e., $w, O, P(w), P(O), P(w|O), P(O|w)$. For this application, how would each of these be derived?

3. Consider the problem of speech recognition. Assume you have a corpus of size 3,715,820, where the word "about" occurs 3,725 times and "a_bow" occurs 38 times. Describe the speech-recognition problem for a given pronunciation, e.g., [ax b aw]. To simplify this problem, assume the only two "words" pronounced this way are "about" and "a_bow". Compute the **prior** and **likelihood** for each of these two words and then determine word which has the highest likelihood for [ax b aw]. You may assume you have access to only one pronunciation rule:

$$\{t,d\} \Rightarrow 0 / V _ \#$$

which has a probability of 0.48. (Recall that, if no pronunciation rules are applicable to a given word, one pronunciation is available, so the probability of that pronunciation, given the word, is 1.)

4. Consider N-grams combined with the noisy channel model. Above we looked at one word at a time (unigram). But even though the word "about" has the highest probability for the pronunciation [ax b aw] (independent of context), the word "a_bow" has a higher probability next to words like "and". N-grams allow us to think about context. Let's imagine there are only 4 words in the vocabulary: *about*, *a_bow*, *and*, and *the*. Describe how to compute the probability that [ax b aw] is "about" when followed by *and* vs. *the*. Compare this to the probability that [ax b aw] is *a_bow* when followed by *and* vs. *the*. Assume that the word *about* is followed by the word *the* with 0.99 probability and the word *and* with 0.01 probability. Conversely, assume the word *a_bow* is followed by the word *the* with 0.01 probability and the word *and* with 0.99 probability.



5. Sparse data problem—what is it? What can we do about it? What is the problem of add-one smoothing? Given a vocabulary of 1000 words (types), and a corpus of 100,000 words (tokens) we want to estimate the **bi-gram** probabilities. We know that half of the bi-grams do not occur in the corpus. What is the unsmoothed probability of a bi-gram that occurs 100 times in the corpus. What is its probability if we apply add-one smoothing? What’s the main disadvantage of using add-one or add-lambda (Lidstone’s Law) smoothing ? You are allowed to round off probabilities.
6. How do we use bigrams to compute $P(S)$ —the probability of a given sentence S ? (You may also hear: “compute $P(S)$ using a Maximum Likelihood Estimate”.) For example, what is $P(\text{like Chinese food})$ using an MLE, where “like” occurs 15000 times in corpus of 1000000 words, “like Chinese” occurs 10000 times in same corpus, “Chinese” occurs 80000 times in same corpus, and “Chinese food” occurs 56000 times in same corpus.
7. Prove that $P_{\text{LAP}}(w_2|w_1) = (C(w_1w_2) + 1)/(C(w_1) + V)$
8. Why does Kneser-Ney smoothing have excellent performance despite the fact that it’s an extension of absolute discounting ? Can we use the MLE distributions when performing Katz backoff instead of the discounted distributions ? There is one particular situation where Katz Backoff may give unreasonable probability estimates.
9. Part-of-Speech Tagging: What is the difference between open and closed word classes? What is the relation between these classes and word frequency in arbitrary documents? POS Ambiguity: How do we tell the difference between “that” and “that” in *I consider that odd*?
10. Rule-Based Tagger: Provide some rules to correctly tag a word with an appropriate POS given its context. Compare this process (of POS tagging) to context-free parsing. (What do they have in common? How do they differ?)
11. Transformation-Based Tagger (TBL): Combination of Rule-based and stochastic tagging methodologies. Consider the following *very tiny* tagged corpus in your answer below.

*Secretariat/NNP is/VBZ expected/VBN to/TO **race**/VB tomorrow/NN*
*People/NNS continue/VBP to/TO inquire/VB the DT reason/NN for/IN the/DT **race**/NN*
for/IN outer/JJ space/NN

 - i. What does TBL require in order to get started?
 - ii. How does TBL work?
 - iii. What does TBL learn?
 - iv. How is TBL iterative?
 - v. What does this system return?
12. Lexical Semantics
 - a. How is polysemy different from synonymy?
 - b. What semantic relations are used in Wordnet? Give an example of each relation.
13. What is the principle of compositionality? Why is it relevant?
14. Provide a first-order predicate logic representation for each of the following sentences:
 - i. All of John’s friends own a car.
 - ii. Some of John’s friends don’t own a car.
 - iii. A friend or relative of John lends John a car.

15. Provide the truth table for $(P \vee Q) \rightarrow R$, for all Boolean values of P, Q, and R.

16. SEMANTICS

In class we learned that thematic roles can be derived from first-order predicate calculus representations. For example, it is possible to derive three thematic roles (Giver, Givee, Given) from this representation:

$$(\exists w,x,y,z) \text{ Giving}(x) \wedge \text{Giver}(w,x) \wedge \text{Givee}(z,x) \wedge \text{Given}(y,,x)$$

However, the mapping to thematic roles might be less direct for some semantic representations. Consider the Lexical Conceptual Structure representation for John gave a book to Mary:

[CAUSE [Person John] [GO_Poss [Thing Book] [TO_Poss [Person Mary]]]]

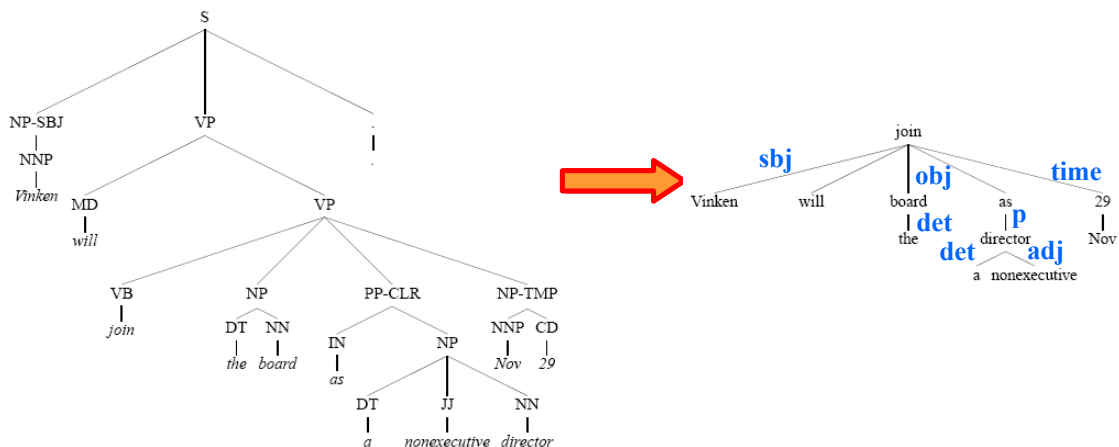
Describe some of the clues in the semantic representation above that might lead to the derivation of the following thematic roles associated with the corresponding lexical entry for *give*:

[CAUSE [Person X] [GO_Poss [Thing Y] [TO_Poss [Person Z]]]]

X = Agent, Y = Theme, Z = Goal

17. DEPENDENCY TREE QUESTION

What head rules and Grammatical/Relationship Mappings are apply to the following conversion:



18. CCG QUESTION : In class we learned about the following three rules :

- X → X/Y Y
- X → Y XY
- X/Z → X/Y Y/Z

For each rule, give a concrete example of a phrasal combination that is characterized by the rule. You should include the phrase computed on the left hand side (both its category and the actual words) and each of the phrases/words associated with the right-hand side components (both their categories and the actual words).

19. TAG QUESTION

For each of the examples given above in 14, what would be the associated TAG Elementary (initial and auxiliary) trees. Illustrate the TAG operations that would be used for combining the elements in your examples.

20. TAG/CCG relation: Is there a way of representing the operation of Adjunction in CCG? Why or why not?