

# **Final Exam - Example Questions**

**601.467/667**

**Introduction to Human Language Technology**

Fall 2019

Johns Hopkins University

This document contains example questions that should give an idea about what to expect in the final exam. We expect you to understand the main concepts and algorithms presented in class. Keep in mind, however, that questions will cover a broader range of topics (anything in the second part of the course) that indicated here with these examples — which are only about the initial lectures on language models, syntax and semantics.

## 1. Language Model

Given the statistics below about unigram and bigram counts in a corpus, and discounting from a smoothing method, what probability is assigned to the bigram  $p(\text{water}|\text{drink})$ ?

unigram	count	bigram	count	count	discounted
<i>beer</i>	10	<i>drink beer</i>	5	1	0.5
<i>wine</i>	5	<i>drink wine</i>	2	2	1.0
<i>soda</i>	1	<i>drink soda</i>	1	3	1.8
<i>water</i>	1	<i>drink milk</i>	1	4	2.9
<i>other</i>	84	<i>drink juice</i>	1	5	4.0

The bigram table is comprehensive, all bigrams starting with *drink* are given. The discounting is only used for the bigram model, the unigram model is not smoothed. Note: Not all of the given information is relevant.

### Solution:

$$\text{drink beer } 5 \rightarrow 4.0/10 \rightarrow 0.4$$

$$\text{drink wine } 2 \rightarrow 1.0/10 \rightarrow 0.1$$

$$\text{drink soda } 1 \rightarrow 0.5/10 \rightarrow 0.05$$

$$\text{drink whiskey } 1 \rightarrow 0.5/10 \rightarrow 0.05$$

$$\text{drink juice } 1 \rightarrow 0.5/10 \rightarrow 0.05$$

$$\text{left for backoff: } 1 - 0.65 = 0.35$$

$$p(\text{water}) = 1/100 = 0.01$$

$$p(\text{water}|\text{drink}) = 0.35 \times 0.01 = 0.0035$$

## 2. Tagging

Given the following distributions:

$$\begin{array}{llll}
 p(\text{bears}|N) = 0.2 & p(\text{like}|P) = 0.2 & p(\text{eating}|N) = 0.1 & p(\text{fish}|N) = 0.2 \\
 p(\text{bears}|V) = 0.1 & p(\text{like}|M) = 0.5 & p(\text{eating}|V) = 0.2 & p(\text{fish}|V) = 0.2
 \end{array}$$

$$\begin{array}{llll}
 p(N|\text{start}) = 0.5 & p(N|N) = 0.2 & p(N|V) = 0.5 & p(N|P) = 0.5 \\
 p(V|\text{start}) = 0.1 & p(V|N) = 0.5 & p(V|V) = 0.2 & p(V|P) = 0.1 \\
 & p(P|N) = 0.1 & p(P|V) = 0.2 & p(N|M) = 0.2 \\
 & p(M|N) = 0.2 & p(M|V) = 0.1 & p(V|M) = 0.5
 \end{array}$$

Carry out the Viterbi algorithm by hand to compute the best tag sequence and its probability for

*bears like eating fish*

Advice: use scientific notation ( $2 \times 10^{-1} = 0.2$ ) for numbers.

**Solution:** Computations are always: *prior state*  $\times$  *transition*  $\times$  *emission*

	START	
<i>bears</i>	N: $0.5 \times 0.2 = 10^{-1}$ BEST	
	V: $0.1 \times 0.1 = 10^{-2}$ BEST	
<i>like</i>	(M from N) $10^{-1} \times 0.2 \times 0.5 = 1 \times 10^{-2}$ BEST	
	(M from V) $10^{-2} \times 0.1 \times 0.5 = 5 \times 10^{-4}$	
	(P from N) $10^{-1} \times 0.1 \times 0.2 = 2 \times 10^{-3}$ BEST	
	(P from V) $10^{-2} \times 0.2 \times 0.2 = 4 \times 10^{-4}$	
<i>eating</i>	(N from M) $1 \times 10^{-2} \times 0.2 \times 0.1 = 2 \times 10^{-4}$ BEST	
	(N from P) $2 \times 10^{-3} \times 0.5 \times 0.1 = 1 \times 10^{-4}$	
	(V from M) $1 \times 10^{-2} \times 0.5 \times 0.2 = 1 \times 10^{-3}$ BEST	
	(V from P) $2 \times 10^{-3} \times 0.1 \times 0.2 = 4 \times 10^{-5}$	
<i>fish</i>	(N from N) $2 \times 10^{-4} \times 0.2 \times 0.2 = 8 \times 10^{-6}$	
	(N from V) $1 \times 10^{-3} \times 0.5 \times 0.2 = 1 \times 10^{-4}$ OVERALL BEST	
	(V from N) $2 \times 10^{-4} \times 0.5 \times 0.2 = 2 \times 10^{-5}$	
	(V from V) $1 \times 10^{-3} \times 0.2 \times 0.2 = 4 \times 10^{-5}$	

Best sequence: *N M V N*  $\rightarrow 1 \times 10^{-4}$

### 3. Syntax

(a) Given the following grammar rules:

$S \rightarrow NP VP$     $VP \rightarrow V NP$     $VP \rightarrow V NP PP$     $NP \rightarrow NP PP$     $PP \rightarrow P NP$   
 $NP \rightarrow \text{Jane}$     $NP \rightarrow \text{paper}$     $NP \rightarrow \text{scissors}$     $V \rightarrow \text{cuts}$     $P \rightarrow \text{with}$

Carry out CKY chart parsing by hand to find all possible parse trees for the sentence

*Jane cuts paper with scissors*

**Solution:**

$S(1+2-5)$

$VP(2+3-5)/VP(2+3+4-5)$

$S(1+2-3)$

$PP(3+4-5)$

$VP(2+3)$

$PP(4+5)$

$NP(1)$

$V(2)$

$NP(3)$

$P(4)$

$NP(5)$

*Jane*

*cuts*

*paper*

*with*

*scissors*

1

2

3

4

5

(b) For the above example, how could a probabilistic parsing model help disambiguate the possible syntactic interpretations of the sentence?

**Solution:** *Lexicalize the grammar.  $VP(\text{cuts}) \rightarrow VP(\text{cuts}) PP(\text{scissors})$ . Learn probabilities for rules. Relationship between paper and scissors should be weaker than between cuts and scissors.*

#### 4. Lexical Semantics

How can automatic methods detect the right meaning of a word occurring in a given examples. List at least four types of information that can help and give an example for each, on hand of the ambiguous word *interest* (curiosity sense vs. fee for loan sense).

**Solution:**

- *surrounding function words: in the interest of*
- *surrounding part-of-speech tags: interest in/P , interest of/P*
- *surrounding content words: interest rate*
- *words with syntactic relationship (verb/object noun): he expressed his great interest in the matter*
- *content words in wider window: the central bank gave no clear indication on the impact of a high level of interest charged on these loans*