1. (2 points) In this course, NLP stands for:
   A. No Likely Proof
   B. Non-Linear Proof
   C. Natural Language Processing
   D. National Lasagna Partners

2. (2 points) In this course, ACL stands for:
   A. Austin City Limits
   B. Association for Computational Linguistics
   C. Anterior Cruciate Ligament
   D. Association for Community Living

3. (4 points) List all the bigrams in the sentence below.

   Colorless green ideas sleep furiously

   Colorless green
green ideas
tems sleep
sleep furiously

4. (4 points) What does the probability $P(nice \mid yesterday\ was)$ represent? (1-2 sentences)

   $P(nice \mid yesterday\ was)$ represents the probability that "yesterday was nice" was found in the text. It is calculated by counting the number of times "yesterday was nice" is found in the corpora and dividing by the number of times "yesterday was" is found in the corpora.

5. (3 points) Why is it undesirable to have zero probabilities in a language model? (1-2 sentences)

   Because language models use the chain rule to estimate the probability of the whole sentence, a single zero probability in the sentence will cause the probability of the entire sentence to be zero. This is undesirable because language is very creative and novel sentences are always being produced, so we don’t want to say these new sentences are impossible, which is what a probability of 0 would mean.
6. (3 points) When using interpolation, the function is $\lambda_1 P(BIGRAM) + \lambda_2 P(UNIGRAM)$. If the bigram is *the dog*, which word should be used to find the unigram probability in the formula?

   dog

7. (6 points) Imagine you are typing on a smart-phone that has predictive typing, and you have typed “The team played so”. How might you use language modeling to determine whether to predict *well* or *bad*?

   A language model could be used to find the probabilities of both the "The team played so well" and "The team played so bad". The one with the higher probability is the word you should predict.

8. (3 points) Besides spelling, what other application could the Noisy Channel model be used for?

   A. Part of Speech Tagging
   B. Parsing
   C. Machine Translation
   D. Question Answering

9. (5 points) The noisy channel model can be written using the formula $\arg \max P(O|w)P(w)$. What does the $P(w)$ part of the formula represent? (1-2 sentences)

   $P(w)$ represents the probability of the hypothesized correct word.

10. (5 points) Given the Finite State Transducer below, what is the output if the input is *possible*?

    ![Finite State Transducer Diagram]

    Possiby

    OR

    Possibye

    I messed up here so anything showing you understood how a
11. (5 points) Imagine we use an FST to transform a word and a set of features into morphemes. For example, *catch+PST* outputs *catch ̃ed*. Clearly this is not an English word. How might we go from *catch ̃ed* to a valid English word?

To go from the morphemes of a word to the written representation of that word, we can use another FST.

12. (10 points) (a) (7 points) The recursive portion of the Viterbi algorithm is:

\[ v_t(q) = \max_{r=1}^N v_{t-1}(r) a_{rq} b_q(o_t) \]

Given the HMM representation and partially completed trellis below, what are the values for \( v_4(H) \) and \( v_4(L) \)? (Numerical expressions are fine, no need to do the calculation.)

\[ V4(H) = \max [ \begin{array} \text{V3(H) * P(H | H) * P(DOT | H),} \\ \text{V3(L) * P(H | L) * P(DOT | H)} \end{array} ] = \max [0.04556 * 0.3 * 0.25, 0.01001 * 0.6 * 0.25] = 0.04556 * 0.3 * 0.25 = 0.003417 \]

\[ V4(L) = \max [ \begin{array} \text{V3(H) * P(L | H) * P(DOT | L),} \\ \text{V3(L) * P(L | L) * P(DOT | L)} \end{array} ] = \max [0.04556 * 0.5 * 0.9, 0.01001 * 0.2 * 0.9] = 0.04556 * 0.5 * 0.9 = 0.020502 \]

(b) (3 points) What is the backtrace entry for \( v_4(L) \)

\( V3(H) \) or \( H \), either one is ok.
13. (5 points) Given the state diagram below, how might the transition probability between JJ and NNP be determined using a corpus (1-2 sentences)?

![State Diagram]

To find the transition probability between JJ and NNP, you need to find \( P(\text{NNP} \mid \text{JJ}) \). This can be calculated from a corpus using the counts of all "JJ NNP" pairs and the counts of all "JJ" tags in total.

14. (4 points) Given the two tag sequences below and the gold standard, which tagger appears to be more accurate, and why?

<table>
<thead>
<tr>
<th>Tagger 1</th>
<th>Tagger 2</th>
<th>Gold</th>
</tr>
</thead>
<tbody>
<tr>
<td>This/A is/B a/C test/D C C N N</td>
<td>This/L is/B a/D test/R N C C C</td>
<td>This/A is/B a/D test/R</td>
</tr>
</tbody>
</table>

Tagger 2 appears to be more accurate because it has 3 of the 4 tags correct, for an accuracy of 0.75.

15. (3 points) Is the grammar below in CNF Form?

A. True
B. False

- A \( \rightarrow \) C D
- B \( \rightarrow \) C F
- C \( \rightarrow \) the | a | an
- D \( \rightarrow \) book | movie | album
- F \( \rightarrow \) run | jump | dive | C D
16. (9 points) Match the parse examples below with the type of parse they are (Write the corresponding letter)

- **A**
  - S
  - NP
  - VP
  - DT
  - NN
  - VB
  - NN
  - The
  - library
  - has
  - books

- **B**
  - S
  - (has)
  - NP
  - VP
  - DT
  - NN
  - VB
  - NN
  - The
  - library
  - has
  - books

- **C**
  - S
  - (has)
  - NP
  - VP
  - DT
  - NN
  - VP
  - PP
  - The
  - library
  - has
  - books

**Lexical** B
**Dependency** C
**Constituent** A

17. (10 points) Given the following grammar, fill in the shaded box in the CKY table, including any necessary backpointers.

\[
\begin{align*}
S & \rightarrow NP \ VP \\
VP & \rightarrow V \ NP \\
VP & \rightarrow VP \ PP \\
VP & \rightarrow use \\
PP & \rightarrow P \ NP \\
NP & \rightarrow Det \ N \\
NP & \rightarrow we \\
NP & \rightarrow internet \\
NP & \rightarrow computer \\
V & \rightarrow use \\
P & \rightarrow on \\
P & \rightarrow in \\
N & \rightarrow we \\
N & \rightarrow internet \\
N & \rightarrow computer \\
Det & \rightarrow the \\
Det & \rightarrow a
\end{align*}
\]
18. (5 points) Draw all parse trees for the phrase below as indicated in the CKY table.

```
S
  V
  VP
  N
  CL
  NP
  S
  V
  VP
```

NOTE: This is the only correct parse tree for the phrase because it is the only complete one, but I did not take off if you had the another tree for the other S.

19. (5 points) What is a benefit of probabilistic parsers that isn’t possible with a standard constituent parsers, such as a standard CKY parser? (1-2 sentences)

Compared to standard constituency parsers, which return all possible parses, probabilistic parsers only return one parse, which is the most likely parse. This is a benefit because it performs syntactic disambiguation for us.

20. (4 points) What are some reasons for the growing adoption of Dependency Parsers? What is an advantage they have? (1-2 sentences)

One reason dependency parsers are becoming more popular is they are very fast because they only have to go through the input one time.

OR

One reason dependency parsers are becoming more popular is that they are independent of word order, so they are useful on many different types of languages.
21. (3 points) Could the following dependency parse be generated by a vanilla shift-reduce based parser?

A. Yes
B. No

---

1Sentence from 2010. Nivre, Joakim. Dependency Parsing. Language and Linguistics Compass