START OF QUIZ
Student ID: 29404548,Lee, Ai Ka

Question 1

Topic: Lecture 1
Source: Lecture 1

Describe allomorphy, with an example we did not cover in class. (1)

Question 2

Topic: Lecture 2
Source: Lecture 2

In class, I was very careful to only have deletion and substitution in our rewrite rules. What implications might insertion have on rewrites? (1)

Question 3

Topic: Lecture 2
Source: Lecture 2

Do you think we could build a POS tagger with an FST, where one side of the tape is POS tags, and the other side of the tape is words? Explain why / why not. (1)

Question 4

Topic: Lecture 3
Source: Lecture 3

What is the motivation behind BPE (ie, what is it maximizing / minimizing)? (1)

Question 5

Topic: Lecture 3
Source: Lecture 3

If we were to try to use an HMM for segmentation, describe what the transition and emission probabilities would be. (1)

Question 6

Topic: Lecture 1
Source: Lecture 1

Vowel harmony is a process by which vowels in affixes must match some of the properties of the vowels in the root. For example, in Turkish, "houses" is "evler", while "schools" is "okullar", where the plural suffix must have a front or back vowel, matching the root ("ev" and "okul"). Given the ML models you've seen so far, give a specific example of a model that you think can learn this process, and explain why it's well suited to the task. (2)

Question 7

Topic: Lecture 4
Source: Lecture 4

Garden path sentences are sentences that start with one parse, but need to be reparsed in the middle of the sentence ("The old man the boats." - 'old' changes from an adjective to a noun, and 'man' from a noun to a verb). A bad Chinese word segmentation could result in the same need to re-parse our segmentation after encountering a new word. Of the methods we looked at, which do you think is the most likely to be able to "correct" a segmentation? Explain. (2)

Question 8

Topic: Lecture 4
Source: Lecture 4

Imagine that we had a language like Chinese that doesn't use spaces for word segmentation, but is considerably more morphologically complex than Chinese. Do you think that it would be easier or harder to segment? Give an explanation (and include any assumptions). (2)

Question 9

Topic: Long
Source: Lecture 3

There is an alternative to BPE that randomly "forgets" to merge together certain subword sequences when it is creating its vocabulary (for example, "forget" will occasionally be represented as "for-get", occasionally as "forget", occasionally as "for-g-et", etc. What impacts do you think this might have on the vocabulary and model performance? Secondly, do you think there is a different impact between forgetting early iteration, mid iteration, and late iteration merges? (3)

END OF QUIZ