START OF QUIZ
Student ID:
10237352,Shan,Zhengyi

Question 1

Topic: Lecture 3
Source: Lecture 3

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

Question 2

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 3

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 4

Topic: Lecture 2
Source: Lecture 2

Do you think that FSTs can work with vowel harmony? Explain. (1)

Question 5

Topic: Lecture 4
Source: Lecture 4

Literate Chinese speakers have no difficulty parsing words in Chinese text (outside normal ambiguities). What advantages do you think they have that our algorithms are lacking? (1)

Question 6

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 7

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 8

Topic: Lecture 1
Source: Lecture 1

Suppletion is a process by which morphological patterns (called paradigms) merge to form a mixed paradigm. For example, the past tense of "to go" comes from an older verb, "wendan - to turn". Describe how syncretic paradigms might impact a machine learning model, and how we can learn to model them accurately. (2)

Question 9

Topic: Long
Source: Lecture 4

In 1799, Napoleon's armies were raiding Egypt when they discovered the Rosetta stone, which contained a parallel corpus of Latin, Greek, and Egyptian (written in hieroglyphics). This was one of the greatest linguistic discoveries, as it allowed linguists to decipher the long unknown Egyptian language. Imagine that we found a similar stone today, but with several languages (you can assume they are well-known languages), including Linear A (a language spoken in ancient Crete). Would you prefer that the languages have high morphological complexity (like, say Finnish or Turkish), simpler morphology (like English or Chinese), or something in the middle, and how would you use this information to inform your automated approach to decrypting Linear A? You can ignore the fact that modern languages didn't exist when Linear A was spoken. Assume that the other languages are completely interpretable. (3)

END OF QUIZ