START OF QUIZ
Student ID: 91905067,Nong-bualang,Bon

Question 1

Topic: Lecture 1
Source: Lecture 1

Explain how morphology, phonology, and syntax are tightly bound to each other. (1)

Question 2

Topic: Lecture 4
Source: Lecture 4

What is the role of the partition function (Z(x)) in CRFs? Why is it necessary? (1)

Question 3

Topic: Lecture 2
Source: Lecture 2

Is a regular expression an example of an FST or an FSA? Explain. (1)

Question 4

Topic: Lecture 3
Source: Lecture 3

Why is differential entropy a good measure for establishing morpheme boundaries? (1)

Question 5

Topic: Lecture 1
Source: Lecture 1

Knowing what you know about parsing, describe how derivation could be considered syntax, instead of morphology. In other words, how might we parse derivations? (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 2
Source: Lecture 2

As a thought experiment, how might we build a calculator using an FST? Imagine that the FST reads input on one side of the tape, and generates operations (that are carried out by an algorithm) on the output side. (2)

Question 8

Topic: Lecture 3
Source: Lecture 3

For a language like Archi, which has extremely productive inflection (a verb can theoretically appear in over 1.5 million different forms), do you think that a larger or smaller BPE vocabulary size would be more beneficial? Explain your assumptions about the morphological structure of the language when making your assessment. (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