START OF QUIZ
Student ID:
38595138,Christilaw,Tim

Question 1

Topic: Lecture 1
Source: Lecture 1

Knowing what you know about parsing, describe how compounding could be considered syntax, instead of morphology. In other words, how might we parse compounds? (1)

Question 2

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 3

Topic: Lecture 4
Source: Lecture 4

What kind of tasks are CRFs appropriate for (2 requirements)? Don't list tasks, but rather the general class of tasks. (1)

Question 4

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 5

Topic: Lecture 4
Source: Lecture 4

Transition-based segmentation is very similar to the SR parser we saw last block, except it uses 2 FIFO structures, and doesn't require a stack. What is different about segmentation so that it doesn't require a stack? (1)

Question 6

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 7

Topic: Lecture 3
Source: Lecture 3

In the lab, you compared BPE with a more linguistically-motivated segmentation scheme. Intrinsically, the supervised method performs much better, but typically, BPE and its cousins work much better down-stream. Why do you think that is, taking into account the differences between the two methods? (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