START OF QUIZ
Student ID:
92642370,Chen,Tiffany

Question 1

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 2

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 3

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

Question 4

Topic: Lecture 1
Source: Lecture 1

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

Question 5

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

Question 6

Topic: Lecture 2
Source: Lecture 2

In some ways, Statistical Machine Translation (SMT) was similar to an FST modified by a re-ordering model (ie, each input word had a corresponding output translation, and then the words were re-ordered to fit a language model). These models have been supplanted by NMT. What shortcomings of FSTs do you think put a ceiling on SMT performance? (2)

Question 7

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 8

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 9

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