

START OF QUIZ
Student ID:
19668508,Li,Julian

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 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 2

Source: Lecture 2

Do you think that FSTs can work with templatic morphology? Explain. (1)

Question 4

Topic: Lecture 3

Source: Lecture 3

In class, we saw that the entropy of a fair coin toss is 1 (bit), but that was because we were using a binary logarithm. Similarly, the entropy of an 8-sided fair die is 1, if we use an octal logarithm. What is the entropy of an 8-sided die using a binary logarithm? Either show your work or explain the relation. (1)

Question 5

Topic: Lecture 1

Source: Lecture 1

Describe the properties of bound and free morphemes, and give an example from a language you know. (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 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 1

In English (and other stress-timed languages, such as German, Russian, Arabic, Greek, Hindi, Thai, etc.), stress tends to be strongest on the root of the word, and is softer along affixes (and in English, on periphrastic necessities like auxiliary verbs - try it!). What implications might this have on an ASR system, do you think they are a significant issue, and can you envision any way of moderating them with morphological knowledge? (3)

END OF QUIZ