START OF QUIZ
Student ID:
58115197,Zhang,Miaolin

Question 1

Topic: Lecture 8
Source: Lecture 8

Nonce words are almost always in open classes of words, but there are some exceptions (such as pronouns, which do allow some flexibility). Do you think that LLMs would react to new words in traditionally closed classes the same way as in open classes? (1)

Question 2

Topic: Lecture 5
Source: Lecture 5

Why is POS information so important (whether via tagging or embedded information)? (1)

Question 3

Topic: Lecture 7
Source: Lecture 7

Describe elision in terms of edit actions. (1)

Question 4

Topic: Lecture 7
Source: Lecture 7

Describe metathesis in terms of edit actions. (1)

Question 5

Would F1 score be an appropriate measure for gauging the quality of a morphological analyzer? Explain why or why not. (1)

Question 6

Topic: Lecture 8
Source: Lecture 8

Imagine we were designing a probe to understand whether a model were gender biased. How might we design such a probe, and if we found the model to exhibit such a bias, what suggestions would you make to neutralize the bias? (2)

Question 7

Imagine that we have some pre-trained multilingual embeddings of really high quality. We train a POS tagger for a very common language, with lots of data, embedding the data with the multilingual embeddings. At inference, we then replace the input with another language. Do you think the tagger would beat a majority baseline? Explain your reasoning, and list any assumptions. (2)

Question 8

Topic: Lecture 6
Source: Lecture 6

Feature engineering can be incorporated into encoder-decoder models through the use of multiple encoders. If you could have any extra annotation for morphological analysis, and were able to pass each through a separate encoder, what types of features would you include? Do you see any potential problems with using this extra annotation? (2)

Question 9

Topic: Long
Source: Lecture 5

In class, we talked about how POS and morphological information is often latently encoded in word embeddings, but not in character embeddings. Let's think about subword embeddings, since most DL models are going to use subword representations. If a word is split, where do you think this information is encoded, and does it matter? Explain your reasoning. (3)

END OF QUIZ