START OF QUIZ
Student ID: 31003346,De Jager,Rachelle

Question 1

Topic: Lecture 6
Source: Lecture 6

Explain how an MSD differs from a POS tag, and how it's similar. (1)

Question 2

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 3

Topic: Lecture 7
Source: Lecture 7

What benefits would evaluating an inflection model on nonce words have? Are there any disadvantages? (1)

Question 4

Topic: Lecture 8
Source: Lecture 8

Where do you think pragmatic learning (ie, intent) might fall within the layers of an LLM? Explain briefly. How might we test for it? (1)

Question 5

Topic: Lecture 7
Source: Lecture 7

Other than characters, we didn't really discuss subword units for inflection. Do you think that BPE would be useful in an inflectional model? Explain briefly. (1)

Question 6

Topic: Lecture 5
Source: Lecture 5

In DSCI 563, we discussed EM for POS tagging. Let's make it neural. Assume we have a small set of gold annotated sentences (100). How could we use contextualized embeddings to bootstrap more annotated data (assume that fine-tuning doesn't work)? (2)

Question 7

Topic: Lecture 5
Source: Lecture 5

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

We know that domain shift can have a significant impact on the quality of our models - despite POS tagging being an "easy" task, POS taggers fail spectacularly when we try to use them on different domains. Do you think the same would be true of (contextual) morphological analysers? What similarities and differences between POS and MSD led you to this conclusion? (2)

Question 9

Topic: Long
Source: Lecture 6

Linguistic determinism is the (somewhat outdated, and originally quite racist) idea that the language we speak shapes the way that we see the world. Although it's been shown to only have a small influence on humans (called weak determinism), there haven't really been many studies that look into it in machines. However, given that our methodologies are so heavily influenced by embedding spaces, we can say that modern ML models are entirely shaped by the language they speak (the language here being the constraints of the embedding space). If we had a well-annotated morphological corpus, how might we set up an experiment to test this theory? (3)

END OF QUIZ