

START OF QUIZ

Student ID:

37157856, Wang, Lusha

Question 1

Topic: Lecture 7

Source: Lecture 7

When generating inflections, what features of the input do you think the model most attends to? (1)

Question 2

Topic: Lecture 6

Source: Lecture 6

What is the purpose of using normalized edit distance in the evaluation of a morphological analyzer? (1)

Question 3

Topic: Lecture 8

Source: Lecture 8

Why do you think that LLMs learn syntax earlier in the model than morphology? (1)

Question 4

Topic: Lecture 5

Source: Lecture 5

Why is a majority tagger such a strong baseline for POS tagging? (1)

Question 5

Topic: Lecture 8

Source: Lecture 8

If we observe a DL model that is not properly modeling morphology (but handles syntax and semantics relatively well), what change might you propose to the architecture of the model (assume we don't have any annotated morphology data). (1)

Question 6

Topic: Lecture 7

Source: Lecture 7

What benefits might encoding MSDs with a second encoder have over a single encoder approach? Can you think of any disadvantages? (2)

Question 7

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 8

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 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Linguistic determinism is the (somewhat outdated, and originally quite racist) idea that the language we speak shapes the way that we see the world.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