START OF QUIZ
Student ID:
54944541,Liu,Daoming

Question 1

Topic: Lecture 1
Source: Lecture 1

Imagine a language is described as “low-resource”. If you could create a single automated tool for the language, what would it be? List any assumptions. (1)

Question 2

Topic: Lecture 3
Source: Lecture 3

What are the benefits of using adapter layers instead of fine-tuning? (1)

Question 3

Topic: Lecture 1
Source: Lecture 1

Describe why “language endangerment” and “language extinction” are contentious term. (1)

Question 4

Topic: Lecture 3
Source: Lecture 3

What benefits does delexicalization bring to the training of dependency parsers? Can you think of other tasks that might benefit from it? (1)

Question 5

Topic: Lecture 4
Source: Lecture 4

Are there any situations where the alpha and beta score at a particular timestep would be equal? (1)

Question 6

Topic: Lecture 2
Source: Lecture 2

In transfer learning, how do you decide which layers of a pre-trained model to freeze and which to fine-tune when adapting it to a new language or task? Give an example of when you might choose to freeze or fine-tune specific layers. (2)

Question 7

Topic: Lecture 4
Source: Lecture 4

I've said a few times that the syntax dominates the signal (especially for languages with less free word order). Where have we seen this, and what does it mean for semi-supervised tagging? (2)

Question 8

Topic: Lecture 2
Source: Lecture 2

Isolate languages do not have any known related languages (Ainu, X̲aat Kíl, and Klingon are some examples). How might you approach using cross-lingual transfer for an isolate? (2)

Question 9

Topic: Long
Source: Lecture 3

Imagine that we want to take what we know about adapter layers and word embeddings to approach shared embedding space in a very different way. We have several multi-lingual embeddings in HRLs that we know are in the same space. We also have embeddings that we've trained for a LRL, but that are in a different space. We concatenate the embeddings, freezing the HRLs, but not the LRL embeddings, and then pass them through a prediction layer for POS tagging. Do you think this would work? Would it be better to try to predict the HRL or LRL (or do it as multi-task learning)? (3)

END OF QUIZ