

START OF QUIZ

Student ID:

27841444,Gu,Martin

Question 1

Topic: Lecture 2

Source: Lecture 2

What is the role of cognates in transfer learning? (1)

Question 2

Topic: Lecture 4

Source: Lecture 4

Could you implement Viterbi as an extension of the forward-backward algorithm? What additional constraints might be needed, if so? (1)

Question 3

Topic: Lecture 2

Source: Lecture 2

What is the intuition behind transfer learning? (1)

Question 4

Topic: Lecture 1

Source: Lecture 1

Describe the concept of a “dialect-chain”, and why the distinction between languages and dialects is not clearly-defined. (1)

Question 5

Topic: Lecture 3

Source: Lecture 3

What potential features of low-resource languages (other than data scarcity) might prevent the successful transfer of information from a HRL? (1)

Question 6

Topic: Lecture 1

Source: Lecture 1

Many existing tools and annotation formats make assumptions about the languages that they are processing. If you were creating an ML corpus for a new language, would you prefer to start from scratch, or to adapt an existing annotation schema? Would this change depending on if you were working with a Class 1 or a Class 5 language? Explain. (2)

Question 7

Topic: Lecture 4

Source: Lecture 4

I briefly mentioned the idea of silver data and bootstrapping in class. What do you think are the limits of silver data? Can you think of any ways to counter them? (2)

Question 8

Topic: Lecture 3

Source: Lecture 3

Imagine that we find a database lying around, and it's been very poorly maintained and documented. All we know is that it contains word embeddings for a language written in Arabic script (assume we can't read it, and the only Arabic speakers we know also can't read it - it's in a language they don't know). How might we go about trying to identify the language that it's written in, without finding speakers of all of the Arabic-derived languages? (2)

Question 9

Topic: Long

Source: Lecture 4

I'm working with a few field linguists who have been annotating a low-resource language for the last year, and they present me with their “huge” dataset of 1000 annotated sentences, assuming I can create a 95

END OF QUIZ