START OF QUIZ Student ID: 95507984,Li,Qihan

Question 1

Topic: Lecture 4
Source: Lecture 4

Why does type-to-token ratio decrease as the size of the corpus increases? What does this suggest about long documents? (1)

Question 2

Topic: Lecture 4
Source: Lecture 4

Attributive adverbs are a type of adverb that provides ”flavour” to speech verbs (example: ”she said quickly”; ”he spoke loudly”). They are often frowned upon in formal writing, because they can be replaced with other verbs: ”blurted” or ”shouted”, in the example. Write a quick function that finds them in the Brown corpus, and reports how many sentences in 1000 have them. (2)

Question 3

Topic: Lecture 1
Source: Lecture 1

What are two ways to check if a word is all capital letters (neither should require more than one function call)? (1)

Question 4

Topic: Lecture 2
Source: Lecture 2

Do you think it's possible for a language not to follow a Zipfian curve? What consequences might that have on communication (if, let's say, if the curve were linear)? (2)

Question 5

Topic: Lecture 1
Source: Lecture 1

When would we *not* want to lowercase text prior to training a model? Give a concrete example. (1)

Question 6

Topic: Lecture 3
Source: Lecture 3

Describe the concept of the ”Minimum viable product”, and how it relates to using lexicons. (1)

Question 7

Topic: Lecture 3
Source: Lecture 3

Imagine that we have a parallel corpus (ie, a corpus containing sentences in two languages), and we want to extract a bilingual lexicon. What are some simple steps we could do to identify words that could be translations of each other? (2)

Question 8

Topic: Lecture 2
Source: Lecture 2

Why is it important to know when a corpus was constructed, and who constructed it? (1)

Question 9

Topic: Coding
Source: Coding

Imagine that we have an encrypted data set in a language we don't know, but it is written in the Latin script (ie, the script of English, French, etc.). What are some tests that we could run to try to determine the original language? Please list any assumptions you make. Assume that machine learning is not an option. (3)

END OF QUIZ