START OF QUIZ
Student ID:
19668508,Li,Julian

Question 1

Topic: Lecture 2
Source: Lecture 2

Why is it important to understand the intended audience and time period of a corpus when conducting linguistic analysis? (1)

Question 2

Topic: Lecture 3
Source: Lecture 3

When would we want to represent linguistic data in a list, instead of a dictionary or a set? (1)

Question 3

Topic: Lecture 3
Source: Lecture 3

How does a defaultdict differ from a regular dictionary in Python? (2 differences) (1)

Question 4

Topic: Lecture 1
Source: Lecture 1

What are two ways to check if a string is a palindrome, without reversing the string? (1)

Question 5

Topic: Lecture 4
Source: Lecture 4

Given a list of tuples where the first element is a string and the second is an integer, write a short piece of code to sort the list in descending order based on the second element. Briefly explain your approach. (1)

Question 6

Topic: Lecture 1
Source: Lecture 1

Write a function that capitalizes the first letter of each word in a string, without using the .title() method or any external libraries. What are some assumptions that you are making? (2)

Question 7

Topic: Lecture 2
Source: Lecture 2

If you were to analyze a corpus for stylistic differences, how might you determine: the formality of the language; whether it's written or spoken; its sentiment? Assume that we don't have existing ML tools or enough data to train one. (2)

Question 8

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 9

Topic: Long
Source: Lecture 2

Imagine you are working with a corpus in a language you don't know, and you need to identify the stopwords in it. You cannot use machine learning but can perform basic statistical analysis. How would you approach identifying stopwords? What metrics would help you confirm that you've identified them correctly? (3)

END OF QUIZ