

START OF QUIZ
Student ID:
78097441,Tampub-
olon,Juan

Question 1

Topic: Lecture 3

Source: Lecture 3

When we nest deep structures in dictionaries, we lose their $O(1)$ benefits. Can you think of a better way to represent complex data sets? (1)

Question 2

Topic: Lecture 1

Source: Lecture 1

Why is the `.split()` method useful when working with sentences or phrases? (1)

Question 3

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 4

Topic: Lecture 4

Source: Lecture 4

Why does the lexical diversity (type-to-token ratio) typically increase when analyzing smaller sub-corpora rather than larger ones? What does this suggest about the content of smaller texts? (1)

Question 5

Topic: Lecture 2

Source: Lecture 2

If a language has a highly synthetic morphology (many affixes), would you expect it to have a higher or lower Type-Token Ratio (TTR) than a language with less rich morphological structure? Briefly explain why. (1)

Question 6

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 7

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 8

Topic: Lecture 3

Source: Lecture 3

Lexicons are useful for initial text analysis but often lack the adaptability needed for advanced NLP tasks. Why is this the case? Provide at least 2 reasons with brief explanations.
(2)

Question 9

Topic: Long

Source: Lecture 4

Suppose you have a large corpus of mixed language text, and you are tasked with detecting which language each sentence is written in. What techniques could you apply to achieve this, assuming you cannot use a pre-trained language detection model? Please provide a detailed explanation of your reasoning. (3)

END OF QUIZ