

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

## 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 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 3

Topic: Lecture 4

Source: Lecture 4

Would you expect a higher or lower frequency of passive voice constructions in legal documents compared to casual conversation? Briefly explain your reasoning. (Remember that passive voice is a structure like "the tree was cut down", inverting the subject and object). (1)

## Question 4

Topic: Lecture 4

Source: Lecture 4

What are two potential drawbacks of removing stopwords from a text before conducting a sentiment analysis? (1)

## Question 5

Topic: Lecture 1

Source: Lecture 1

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

## Question 6

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 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

Imagine you have a large text corpus in English and Spanish and want to automatically align sentences for machine translation. What are some straightforward methods you could use to identify sentence pairs that are likely translations of each other? (2)

## Question 9

Topic: Long

Source: Lecture 1

Write a function that validates if a string matches a phone number format, such as (123) 456-7890. What types of invalid inputs should the function check for? Are there edge cases we would be willing to accept? How would we handle those? Write 3 test cases - 2 that should pass, and one that should fail. (3)

**END OF QUIZ**