

**START OF QUIZ**

**Student ID:**

**49919301,Keigan,Jonathan**

## Question 1

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 2

Topic: Lecture 3

Source: Lecture 3

Why do we not care about the extra space required to create a reverse index? (2 reasons) (1)

### Question 3

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 4

Topic: Lecture 2

Source: Lecture 2

How does Zipf's law help explain the distribution of word frequencies in a corpus? What impacts does that have on our algorithms? (1)

## Question 5

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 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 1

Source: Lecture 1

You are given a sentence. Write a function to count how many words in the sentence start with a vowel, without using loops or list comprehensions. (2)



## Question 8

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 9

Topic: Long

Source: Lecture 3

Grimm's law is a linguistic phenomenon that describes how sounds in language (mostly related to the Germanic languages like English, Dutch, German, Norwegian, Icelandic, etc.) changed over time (specifically from some progenitor thousands of years old - Germanic languages observed the change, while Romance languages did not). For example, the /p/ sound in Latin evolved into the /f/ sound in English across many words - compare "piscus" with "fish"; "pater" with "father"; "pedus" with "foot" (there are a couple other changes in there, too - see if you can spot them!). If you had a time machine, and could bring a computational toolkit to help Jacob Grimm formulate his law, what would you need, in terms of lexicons, keeping in mind that the /p/ -> /f/ change is only one of a handful of sound changes, and that the changes occurred over dozens of languages? Explain (with pseudocode, if necessary), how you would start to identify trends in the data (assuming that your computer still works in the 19th century)? (3)

**END OF QUIZ**