START OF QUIZ
Student ID:
37083607,Zeng,Jiao

Question 1

Topic: Lecture 1
Source: Lecture 1

When would you choose to preserve the original case of text during data processing, rather than converting everything to lowercase? (1)

Question 2

When would we want to represent linguistic data in a list, instead of a dictionary or a set? (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 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 5

Topic: Lecture 4
Source: Lecture 4

In class, we removed stopwords by using a lexicon. Can you think of another way that we could remove all closed class words? (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 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 8

Topic: Lecture 2
Source: Lecture 2

Is it possible for a corpus of a low-resource language to follow Zipf's law? What factors might influence the degree to which the law applies in such languages? (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 occured 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