

**START OF QUIZ**

**Student ID: 60226669, Mo-  
hammad, Waquas**

## Question 1

Topic: Lecture 1

Source: Lecture 1

When would we \*not\* want to lowercase text prior to training a model? Give a concrete example. (1)

## Question 2

Topic: Lecture 2

Source: Lecture 2

How does Zipf's law relate to Hapax Legomena? (1)

### Question 3

Topic: Lecture 1

Source: Lecture 1

How would you reverse a string and keep the result? (1)

## Question 4

Topic: Lecture 2

Source: Lecture 2

Do you think it's possible for a language not to follow a Zipfian curve? What consequences might that have on communication (if, let's say, if the curve were linear)? (2)

## Question 5

Topic: Lecture 4

Source: Lecture 4

Do you think that children's (age 3-5) picture books would have a higher or lower ratio of adjectives than university literature? Briefly explain your logic. (1)

## Question 6

Topic: Lecture 3

Source: Lecture 3

Although lexicons are often good starting points, they are often less capable than ML methods. What are some reasons (at least 2) that lexicons are insufficient for state-of-the-art training. Briefly explain. (2)

## Question 7

Topic: Lecture 4

Source: Lecture 4

In French, negation is often indicated by "ne ... pas" (ie, "je ne parle pas" - "I am not speaking"; "tu ne conduis pas" - "You are not driving", etc.). However, in speech, one of the two is often dropped: "je ne parle." or "tu conduis pas.". Using this information, how would you determine whether a corpus was composed of written or spoken French? You don't need to write the code, but explain the logic that you would use to come to this conclusion. (2)



## Question 8

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 9

Topic: Coding

Source: Coding

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**