

**START OF QUIZ**  
**Student ID: 31003346,De**  
**Jager,Rachelle**

## Question 1

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 2

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 3

Topic: Lecture 2

Source: Lecture 2

What role does linguistic annotation provide for corpora, specifically for computational linguistics? (1)

## Question 4

Topic: Lecture 1

Source: Lecture 1

How would you convert a string into a list of characters? (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

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 7

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 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 2

Imagine you are working with a corpus in a language you don't know, and you need to identify the stopwords in it. You cannot use machine learning but can perform basic statistical analysis. How would you approach identifying stopwords? What metrics would help you confirm that you've identified them correctly? (3)

**END OF QUIZ**