

LING 471
Spring 2026
Sample Final

Multiple Choice (10 points each)

1. Which statement best describes the main goal of tokenization?
 - A. Encrypt text for privacy
 - B. Split text into smaller units for analysis
 - C. Translate text into another language
 - D. Remove all punctuation from text

2. In a binary classifier, precision for the positive class is:
 - A. True positives divided by all actual positives
 - B. True positives divided by all predicted positives
 - C. True negatives divided by all negatives
 - D. Correct predictions divided by all predictions

3. What is a key difference between a list and a dictionary in Python?
 - A. Lists store key-value pairs; dictionaries store ordered values
 - B. Lists are mutable; dictionaries are immutable
 - C. Lists are indexed by position; dictionaries are indexed by keys
 - D. There is no practical difference

4. Which Git command updates your local repository with remote changes?
 - A. git init
 - B. git add
 - C. git pull
 - D. git branch

5. Why are breakpoints useful in debugging?
 - A. They automatically fix syntax errors
 - B. They let you pause execution and inspect program state
 - C. They speed up program runtime
 - D. They replace test cases

Short Answer (10 points each)

6. Define a data statement in NLP and explain one reason it matters.

7. What is the difference between precision and recall in your own words?

8. Give one example where high accuracy can still be misleading.

9. Explain one way Unicode issues can affect text processing.

10. Write pseudocode to count how many times a word appears in a text file.