

LING 471
Spring 2026
Final Review Guide

1. Core Programming Terms

- Variable
- Data type: string, int, float, bool
- List, dictionary, set, tuple
- Indexing and slicing
- Conditional: if / elif / else
- Loop: for, while
- Function: parameters, return value
- Scope: local vs global
- File I/O: open, read, write
- Command-line arguments
- Debugging: breakpoint, step over, step into, watch variables

2. Command Line + Git

- Terminal / shell
- File path: absolute vs relative
- Directory navigation: cd, ls, pwd
- Git repository, commit, push, pull
- Why version control is useful

3. Text Processing + NLP Basics

- Corpus
- Tokenization
- Normalization / text cleaning
- Regular expression (regex)
- Unicode, ASCII
- Linguistic annotation
- Data statement

4. Evaluation Metrics

- True Positive (TP), True Negative (TN), False Positive (FP), False Negative (FN)
- Accuracy = $(TP + TN) / (TP + TN + FP + FN)$
- Precision = $TP / (TP + FP)$
- Recall = $TP / (TP + FN)$

5. Probability + Stats Foundations

- Probability
- Conditional probability
- Joint probability
- Marginal probability

- Independent events
- Mutually exclusive events
- Bayes' theorem (conceptual interpretation)
- Maximum likelihood estimation (MLE)
- Why logs are used in probability calculations

6. Machine Learning Concepts

- Baseline model
- Train vs test split
- Overfitting (basic intuition)
- Linear regression vs logistic regression (high-level difference)
- Naive Bayes assumption (conditional independence)
- N-gram language model (what it is + one limitation)
- Neural network (high-level idea of nonlinearity)
- Deep neural network
- Semantic vector
- Cosine similarity (high-level meaning)

7. Pseudocode Example

Here's an example of a pseudocode to counting how many times a word appears in a text file

```
read text from file
split text into words
set count = 0
for each word: if word is equal to target word, increment count by 1
print count
```