NLP Assignment 1: Answers

Name: Alireza Dastmalchi Saei

Stu No.: 993613026

1 Question 1

The 4 stages of natural languages are:

- **Lexical Ambiguity**: The ambiguity of a single word is called Lexical Ambiguity. It arises from the presence of words with multiple meanings or interpretations.

- **Syntactic Ambiguity**: This kind of ambiguity occurs when a sentence is parsed in different ways due to its syntactic structure.

- **Semantic Ambiguity**: It arises when a sentence has multiple interpretations based on the meaning of the words used.

- **Pragmatic Ambiguity**: Pragmatic ambiguity happens when the context of a phrase gives it multiple interpretations (the statement is not specific).

2   Question 2

Explanation: The Maximum Matching Algorithm is a method used in Natural Language Processing (NLP) and text processing to segment words or tokens from a sequence of characters. It trys to find the longest possible sequences of words or tokens from the input text based on a given dictionary or vocabulary.

This algorithm takes a sequence of characters (string) as input. Then, using a list of all valid words available (dictionary), the algorithm iterates over the input text and tries to match the longest possible sequence of characters with words or tokens from the dictionary. This greedy algorithm returns a list of words or tokens segmented from the input text.

Example: This algorithm acts as following:
    Input: *"thisisatestsentence"*
    Dictionary: ["this", "is", "a", "test", "sentence", "the", "red", "nlp"]
    Output: ["this", "is", "a", "test", "sentence"]

    Applications: Applications of the Maximum Matching Algorithm:

- Tokenization in Natural Language Processing.

- Segmenting words or tokens from raw text in Information Retrieval systems.

- Word segmentation in languages without clear word boundaries, such as Chinese or Thai.

3 Question 3

Lemmatization: Lemmatization is the process of reducing words to their base or dictionary form, known as the lemma. It involves identifying the root form of a word considering its meaning.

Stemming: Stemming is the process of removing suffixes or prefixes from words to extract their root form, known as the stem. It's a heuristic process that chops off ends of words based on common patterns. But this method is not always able to produce valid dictionary words.