1. **What are Corpora?**

**Answer:-**

A corpus is a collection of text that is used for linguistic analysis. Corpora can be large or small, and they can be either written or spoken. Corpora are often used to train machine learning models for natural language processing tasks.

1. **What are Tokens?**

**Answer:-**

Tokens are the basic units of text that are used in natural language processing. Tokens can be words, punctuation marks, or other symbols. Tokens are created by breaking up text into smaller pieces, and they are often used to represent the input to a natural language processing model.

1. **What are Unigrams, Bigrams, Trigrams?**

**Answer:-**

Unigrams, bigrams, and trigrams are n-grams, which are sequences of n words. Unigrams are single words, bigrams are sequences of two words, and trigrams are sequences of three words. N-grams are often used to represent the input to a natural language processing model.

1. **How to generate n-grams from text?**

**Answer:-**

To generate n-grams from text, you can use a simple algorithm that breaks up the text into sequences of n words. For example, to generate bigrams, you would break up the text into sequences of two words. You can then use the n-grams to represent the input to a natural language processing model.

1. **Explain Lemmatization**

**Answer:-**

Lemmatization is a process of grouping together different inflected forms of a word so they can be analyzed as a single item. For example, the words "walked," "walking," and "walk" would all be lemmatized to the word "walk." Lemmatization is often used in natural language processing to improve the accuracy of models.

1. **ExplainStemming**

**Answer:-**

Stemming is a process of reducing a word to its word stem, which is the basic form of the word. For example, the word "walking" would be stemmed to the word "walk." Stemming is often used in natural language processing to improve the efficiency of models.

1. **ExplainPart-of-speech (POS) tagging**

**Answer:-**

Part-of-speech (POS) tagging is the process of assigning a part of speech tag to each word in a sentence. POS tags can be used to represent the grammatical structure of a sentence, and they are often used in natural language processing tasks such as named entity recognition and machine translation.

1. **ExplainChunking or shallow parsing**

**Answer:-**

Chunking is a process of grouping together words in a sentence that are related to each other grammatically. Chunks can be used to represent the grammatical structure of a sentence, and they are often used in natural language processing tasks such as named entity recognition and machine translation.

1. **ExplainNoun Phrase (NP) chunking**

**Answer:-**

Noun phrase (NP) chunking is a type of chunking that focuses on identifying noun phrases in a sentence. Noun phrases are phrases that contain a noun and its modifiers. NP chunking is often used in natural language processing tasks such as named entity recognition and machine translation.

1. **ExplainNamed Entity Recognition**

**Answer:-**

Named entity recognition (NER) is the process of identifying named entities in a text. Named entities are words or phrases that refer to specific people, places, or organizations. NER is often used in natural language processing tasks such as question answering and machine translation.