1. What are Corpora?  
   A corpus is a collection of authentic text or audio organized into datasets.
2. What are Tokens?  
   Tokenization is used in natural language processing to split paragraphs and sentences into smaller units that can be more easily assigned meaning.
3. What are Unigrams, Bigrams, Trigrams?  
   unigram means taking only one word at a time, bigram means taking two words at a time and trigram means taking three words at a time
4. How to generate n-grams from text?  
   N-grams of texts are extensively used in text mining and natural language processing tasks. They are basically a set of co-occurring words within a given window and when computing the n-grams you typically move one word forward (although you can move X words forward in more advanced scenarios).
5. Explain Lemmatization  
   Lemmatization is a text normalization technique used in Natural Language Processing (NLP), that switches any kind of a word to its base root mode. Lemmatization is responsible for grouping different inflected forms of words into the root form, having the same meaning.
6. Explain Stemming  
   Stemming is the process of reducing a word to its stem that affixes to suffixes and prefixes or to the roots of words known as "lemmas". Stemming is important in natural language understanding (NLU) and natural language processing (NLP)
7. Explain Part-of-speech (POS) tagging  
   Part-of-speech (POS) tagging is a popular Natural Language Processing process which refers to categorizing words in a text (corpus) in correspondence with a particular part of speech, depending on the definition of the word and its context
8. Explain Chunking or shallow parsing  
   Shallow parsing (also chunking or light parsing) is an analysis of a sentence which first identifies constituent parts of sentences (nouns, verbs, adjectives, etc.) and then links them to higher order units that have discrete grammatical meanings (noun groups or phrases, verb groups, etc.).
9. Explain Noun Phrase (NP) chunking  
   Chunking is a process of extracting phrases from unstructured text, which means analyzing a sentence to identify the constituents(Noun Groups, Verbs, verb groups, etc.) However, it does not specify their internal structure, nor their role in the main sentence. It works on top of POS tagging.
10. Explain Named Entity Recognition

Name Entity Recognition (NER) is a sub-task of natural language processing (NLP) that involves identifying and classifying named entities in a text into predefined categories such as person names, organizations, locations, medical codes, time expressions, quantities, monetary values, percentages, etc.