1. What is the important of Morphology? Define its types?

Morphology is a branch of linguistics that deals with the study of the internal structure and formation of words. It plays a crucial role in understanding how words are formed, their meaning, and their relationship to other words in a language

1. Free morpheme
2. Morpheme

The study of morphology is important for several reasons:

Word Formation: Morphology provides insights into how words are formed in a language. It helps us understand the rules and patterns that govern word construction, such as the addition of prefixes and suffixes or the process of compounding. This knowledge is essential for vocabulary expansion and comprehension.

Meaning Interpretation: Morphology plays a crucial role in determining the meaning of words. By analyzing the morphemes (meaningful units) within a word, we can uncover the individual components that contribute to its overall meaning. This is particularly useful for understanding complex or unfamiliar words.

Word Relationships: Morphological analysis helps us identify the relationships between words within a language

2. Define Tokenization and Hybrid Approaches?

Tokenization: Tokenization is the process of breaking down a sequence of text into individual units called tokens. In natural language processing (NLP), tokens are usually words, but they can also be subwords, characters, or any other meaningful units depending on the context and task. Tokenization serves as a crucial initial step in NLP tasks as it allows the text to be processed and analyzed at a granular level.

Hybrid Approaches: Hybrid approaches in NLP refer to combining different techniques or models to improve performance or address specific challenges. These approaches leverage the strengths of multiple methods to overcome limitations or achieve better results compared to using a single technique.

3. Perform parsing using simple top down parsing for the sentence "The dogs cried" using the grammar given below: S->NP VP,NP->ART N NP->ART ADJ N ,VP->V VP->V NP

Step 1:

S (Start symbol)

Step 2:

S -> NP VP

-> ART N VP

Step 3:

S -> NP VP

-> ART NP VP

-> ART ADJ N VP

Step 4:

S -> NP VP

-> ART NP VP

-> ART ADJ NP VP

-> ART ADJ N VP

Step 5:

S -> NP VP

-> ART NP VP

-> ART ADJ NP VP

-> ART ADJ N VP

-> ART ADJ N V NP

Step 6:

S -> NP VP

-> ART NP VP

-> ART ADJ NP VP

-> ART ADJ N VP

-> ART ADJ N V NP

-> ART ADJ N V NP

-> ART ADJ N V NP

Step 7:

S -> NP VP

-> ART NP VP

-> ART ADJ NP VP

-> ART ADJ N VP

-> ART ADJ N V NP

-> ART ADJ N V NP

-> ART ADJ N V NP

-> ART ADJ N V ART N

The final parse tree for the sentence "The dogs cried" using the given grammar and top-down parsing is as follows

S

|_

| |

ART VP

| _|

The | NP

| |

ADJ | |

| N V

dogs | |

cried NP

|

ART

|

N

|

Dogs

4.Discuss Structural ambiguity, Semantic ambiguity, Pragmatic ambiguity?

Structural Ambiguity: Structural ambiguity refers to ambiguity that arises from the way words or phrases are organized or structured within a sentence. It occurs when a sentence can be parsed or interpreted in multiple ways, leading to different syntactic structures and, consequently, different meanings.

Semantic Ambiguity: Semantic ambiguity occurs when a word or phrase has multiple meanings or interpretations. It arises from the inherent polysemy (multiple meanings) of words or from the context in which they are used. Semantic ambiguity can lead to confusion or misinterpretation of the intended message.

Pragmatic Ambiguity: Pragmatic ambiguity arises from the use of language in context, where the intended meaning may be different from the literal interpretation. It occurs when the speaker intends a message to be understood in a particular way, but the sentence itself can be interpreted differently due to contextual factors or implied meaning

5. What is FOPC? Write FOPC for the following sentences: S1:jack loves Jill S2:Suman takes Egg or Pharmacy

FOPC stands for First-Order Predicate Calculus, which is a formal logical system used to represent statements and reason about them. It consists of variables, constants, predicates, quantifiers, logical connectives, and parentheses. FOPC allows us to express complex statements in a structured and unambiguous manner.

Here's the FOPC representation for the given sentences:

S1: Jack loves Jill

Let's represent "loves" as a binary predicate "L" and use the variables x and y to represent Jack and Jill, respectively.

L(jack, jill)

This statement asserts that Jack loves Jill.

S2: Suman takes Egg or Pharmacy

To represent the sentence, we need to define predicates and constants that correspond to the elements mentioned. Let's use the predicate "T" to represent the action of taking and use the constants "suman," "egg," and "pharmacy" for Suman, the egg, and the pharmacy, respectively. We also need to represent the logical disjunction "or" using the connective symbol "∨".

T(suman, egg) ∨ T(suman, pharmacy)

This statement expresses that Suman takes either the egg or goes to the pharmacy.

6. What is Structural Ambiguity?

Structural ambiguity refers to a linguistic phenomenon where a sentence or phrase can be interpreted in multiple ways due to the arrangement or structure of its constituent elements. It occurs when a sentence can be parsed or understood in different syntactic structures, leading to different meanings or interpretations.

7. Write a short notes on: a) Frame net b)POS Tagging c)Stemming d)Named entity Reorganization?

a) FrameNet:

FrameNet is a lexical database and computational resource that captures and represents the meaning of words and phrases in terms of conceptual frames. A frame is a cognitive structure that represents a concept or scenario and consists of a set of semantic roles and their relationships. FrameNet associates lexical units (words or phrases) with specific frames and provides information about the roles and their fillers within each frame.

b) POS Tagging:

POS (Part-of-Speech) tagging is the process of assigning a grammatical label (tag) to each word in a sentence, indicating its syntactic category and its role in the sentence structure. The tags represent parts of speech such as nouns, verbs, adjectives, adverbs, pronouns, prepositions, conjunctions, and more.

c) Stemming:

Stemming is a process used in natural language processing to reduce words to their base or root form, known as the stem. It aims to remove inflections or derivational affixes from words, enabling different forms of the same word to be treated as a single unit. Stemming algorithms apply linguistic rules to truncate or remove suffixes and prefixes, resulting in a simplified form of the word.

d) Named Entity Recognition:

Named Entity Recognition (NER) is a natural language processing task that involves identifying and classifying named entities in text. Named entities refer to real-world objects such as persons, organizations, locations, dates, quantities, and more. NER systems analyze the text and apply machine learning or rule-based techniques to detect and classify these entities.

8. What is software?

Software refers to a collection of programs, data, and instructions that enable a computer system or device to perform specific tasks or functions. It is a set of logically related instructions that are written in a programming language and executed by a computer to accomplish desired outcomes.

9. What is Large-Scale Language Model?

A Large-Scale Language Model refers to a sophisticated and advanced artificial intelligence (AI) model that has been trained on vast amounts of textual data to understand and generate human-like language. These models are designed to capture the statistical patterns, grammar, and semantic relationships present in the data they are trained on.

10. Define N-Gram Model?

An N-gram model is a statistical language model used in natural language processing (NLP) to predict the probability of a word or sequence of words based on the previous N-1 words in a given text. It is a type of probabilistic model that estimates the likelihood of encountering a particular word or sequence of words based on the observed frequencies in a training corpus. In an N-gram model, "N" refers to the number of words considered for prediction. For example, in a bigram (2-gram) model, the probability of a word is estimated based on the previous single word, while in a trigram (3-gram) model, the probability is estimated based on the previous two words.