# NLP 2 Marks

---

**1. Define the primary components that make up a word in morphology.**

**CO1, K1, Unit-I**

- The **primary components** of a word in morphology are:
  - **Root**: The core part carrying the main meaning.
  - **Affixes**: Includes prefixes, suffixes, infixes, and circumfixes that modify the root.
  - **Stem**: The root combined with derivational affixes.
  - **Morphemes**: The smallest meaningful unit in a language (free morphemes can stand alone, bound morphemes cannot).

---

**2. What are the main challenges in word structure in different languages?**

**CO1, K1, Unit-I**

- **Complexity in Morphology**: Languages can be isolating (e.g., Chinese) or agglutinative (e.g., Turkish).
- **Ambiguity**: Words may have multiple meanings or grammatical forms.
- **Inflection**: Different languages use varied patterns for tense, case, gender, etc.
- **Compounding**: Some languages heavily rely on compounding words, which increases complexity.
- **Loanwords**: Borrowed words may not follow native morphological rules.

---

**3. Classify the key morphological models used for analyzing word structure.**

**CO1, K2, Unit-I**

- **Concatenative Morphology**: Words are formed by linear concatenation of morphemes (e.g., English).
- **Non-concatenative Morphology**: Words involve changes in internal structure, such as root-and-pattern systems (e.g., Arabic).
- **Item-and-Arrangement Model**: Words are built by arranging morphemes.
- **Item-and-Process Model**: Words undergo transformations (e.g., stem modification).
- **Word-and-Paradigm Model**: Emphasizes word inflections within paradigms.

---

**4. How to find the document structure analysis important in natural language processing?**

**CO1, K1, Unit-I**

- Document structure analysis helps in:
  - **Understanding Layout**: Extracting sections like headings, paragraphs, and lists.
  - **Information Retrieval**: Identifying relevant sections quickly.
  - **Semantic Understanding**: Organizing content hierarchically for better comprehension.
  - **Data Preprocessing**: Preparing structured data for downstream NLP tasks (e.g., summarization, translation).
  - **Metadata Extraction**: Finding authorship, date, and context of the document.

---

**5. Name the primary methods used for finding the structure of documents.**

**CO1, K1, Unit-I**

- **Optical Character Recognition (OCR)**: Extracts text from scanned documents.
- **Layout Analysis**: Identifies document components like headers and columns.
- **Syntactic Parsing**: Analyzes sentence structures.
- **Content Segmentation**: Divides text into logical sections or topics.
- **Entity Recognition**: Finds important entities like titles, names, and keywords.

---

**6. Find the factors contributing to the complexity of document structure analysis methods.**

**CO1, K1, Unit-I**

- **Document Variability**: Different formats, fonts, and layouts.
- **Noisy Data**: Scanned documents with errors or smudges.
- **Multilingual Texts**: Handling multiple languages within a document.
- **Semantic Ambiguity**: Difficulty in discerning meaning or intent.
- **Hierarchical Relationships**: Complex nested structures (e.g., subheadings).

---

**7. What is the primary goal of syntax analysis in natural language processing?**

**CO2, K1, Unit-II**

- The primary goal of syntax analysis is to:
    - **Parse Sentences**: Determine grammatical structure using parse trees.
    - **Ensure Correctness**: Check if the sentence adheres to grammar rules.

- **Establish Relationships**: Identify relationships between words (subject, predicate, object).
- **Facilitate Semantic Analysis**: Prepare data for meaning extraction.
- **Enable Machine Understanding**: Provide a structural foundation for NLP tasks like translation and dialogue systems.

**8. What is a treebank, and how is it used in syntactic analysis?**

**CO2, K1, Unit-II**

- A **treebank** is a parsed text corpus where sentences are annotated with syntactic structures, typically represented as trees.
- **Uses in Syntactic Analysis**:
  - Provides training data for machine learning models in parsing tasks.
  - Helps validate syntactic rules and linguistic theories.
  - Enables development of robust parsers for natural language processing.
  - Supports cross-linguistic syntactic studies using multilingual treebanks.

---

**9. Explain the common methods for representing syntactic structures in natural language processing.**

**CO2, K2, Unit-II**

- **Parse Trees**: Represent hierarchical syntactic structures of sentences.
- **Dependency Trees**: Capture grammatical relationships between words, focusing on head-dependent relations.
- **Constituency Trees**: Show how words group into phrases (e.g., noun phrases).
- **Linear Representations**: Encoded formats like brackets or Penn Treebank-style annotations.

- **Graphs**: Directed acyclic graphs for complex syntactic relations.

---

**10. Explain the context-free grammar (CFG), and how is it used in parsing natural language?**

**CO2, K2, Unit-II**

- **Context-Free Grammar (CFG)**: A set of production rules where each rule maps a symbol to a combination of other symbols or terminal elements (words).
- **Usage in Parsing**:
  - Generates possible sentence structures (parse trees).
  - Helps validate if sentences conform to grammar rules.
  - Forms the backbone of many parsers like CYK and Earley parsers.
  - Enables syntactic analysis for applications like machine translation.

---

**11. Classify the challenges of applying syntactic analysis to multilingual natural language processing.**

**CO2, K2, Unit-II**

- **Diverse Grammar Rules**: Different languages have unique syntax and word order.
- **Resource Scarcity**: Limited annotated corpora for many languages.
- **Morphological Complexity**: Rich inflection in some languages (e.g., Finnish).
- **Code-Switching**: Mixing languages within a sentence.
- **Ambiguity**: Varying interpretations of the same structure in different languages.

---

## 12. Explain the difference between syntactic parsing and semantic parsing.

**CO3, K1, Unit-III**

- **Syntactic Parsing**:
  - Focuses on the grammatical structure of sentences.
  - Outputs a parse tree or dependency graph.
  - Concerned with syntax rules, not meaning.
- **Semantic Parsing**:
  - Extracts the meaning of a sentence in a structured form (e.g., logical representation).
  - Maps sentences to knowledge representations or actionable commands.
  - Deals with context and disambiguation.

---

## 13. What is semantic interpretation in NLP?

**CO3, K2, Unit-III**

- **Semantic Interpretation** is the process of deriving the meaning of a sentence or phrase.
- It involves:
  - **Disambiguation**: Resolving lexical and syntactic ambiguities.
  - **Entity Recognition**: Identifying people, places, and objects.
  - **Contextual Understanding**: Considering the surrounding text or domain.
  - **Knowledge Integration**: Mapping sentences to knowledge bases or ontologies.

---

**14. Name two system paradigms commonly used in NLP.**

**CO3, K2, Unit-III**

1. **Rule-Based Systems**:
    - Uses handcrafted linguistic rules for processing.
    - Suitable for grammar checking and parsing in controlled environments.
2. **Machine Learning-Based Systems**:
    - Employs statistical models trained on large datasets.
    - Used in applications like language translation and sentiment analysis.

---

**15. What is the purpose of NLP software tools?**

**CO3, K2, Unit-III**

- **Purpose of NLP Tools**:
    - **Text Processing**: Tokenization, lemmatization, and stemming.
    - **Language Understanding**: Parsing, sentiment analysis, and entity recognition.
    - **Automation**: Enables chatbots, voice assistants, and machine translation.
    - **Data Extraction**: Extracts structured information from unstructured text.
    - **Research**: Supports linguistic and AI research by providing powerful APIs.

**16. What is a Markov model in the context of NLP?**

**CO3, K2, Unit-III**

- A **Markov Model** is a probabilistic model that assumes the future state of a system depends only on the current state (Markov property).
- **Applications in NLP**:
    - Used in tasks like **part-of-speech tagging**, **speech recognition**, and **machine translation**.

- Helps predict word sequences by assigning probabilities (e.g., Hidden Markov Models for tagging tasks).
- Simplifies complex linguistic dependencies into manageable probabilities.

---

**17. Explain the concept of predicate-argument structure in natural language processing.**

**CO4, K2, Unit-IV**

- The **predicate-argument structure** represents the relationship between a predicate (typically a verb) and its arguments (subjects, objects, etc.).
- **Components**:
  - **Predicate**: Describes an action or state.
  - **Arguments**: Entities participating in the action (e.g., agent, theme).
- Example: *"John (agent) eats (predicate) an apple (theme)."*
- **Use**: Helps understand sentence semantics and core relationships.

---

**18. Describe how predicate-argument structure helps in understanding sentence meaning.**

**CO4, K2, Unit-IV**

- The **predicate-argument structure** provides a framework for analyzing the semantic roles of sentence components.
- **Benefits**:
  - Identifies "who did what to whom."
  - Resolves ambiguity by focusing on role relations.
  - Facilitates tasks like **question answering**, **information extraction**, and **semantic role labeling**.
- Example: *In "She gave him a book,"* "gave" identifies roles for "She" (agent), "him" (recipient), and "a book" (theme).

---

**19. What is the primary purpose of a meaning representation system in natural language processing?**

**CO4, K1, Unit-IV**

- The **primary purpose** of a meaning representation system is to encode the meaning of sentences in a structured, machine-readable format.
- **Applications**:
  - Enables machines to interpret and process text.
  - Supports tasks like information retrieval, machine translation, and reasoning.
  - Maps natural language to logical forms or semantic graphs for better analysis.

---

**20. Choose an example of a commonly used meaning representation system and briefly describe its use.**

**CO4, K2, Unit-IV**

- **Example**: **First-Order Logic (FOL)**
- **Description**:
  - Represents meaning using logical predicates, variables, and quantifiers.
  - Encodes relationships and facts (e.g., $\exists x$ (Apple(x) $\wedge$ Red(x)) means "There exists an apple that is red").
- **Use**:
  - Facilitates tasks like reasoning, question answering, and semantic interpretation.
  - Helps model knowledge bases for AI systems.

**21. Find the importance of annotated corpora in the development of NLP software.**

**CO4, K1, Unit-IV**

- **Annotated corpora** are text datasets with linguistic information like parts of speech, syntax, semantics, etc.

- **Importance**:
  - Provide training data for supervised machine learning models.
  - Enable the development of robust NLP tools like parsers and taggers.
  - Facilitate evaluation and benchmarking of NLP algorithms.
  - Support linguistic research and cross-linguistic studies.
  - Ensure high accuracy in applications like named entity recognition and sentiment analysis.

---

**22. What is the role of lexical cohesion in discourse processing, and how does it contribute to the overall coherence of a text?**

**CO5, K1, Unit-V**

- **Lexical cohesion** refers to the use of semantically related words to create links between sentences and ideas in a text.
- **Role in Discourse Processing**:
  - Ensures continuity of meaning across a text.
  - Helps in identifying topic shifts and connections.
  - Facilitates text summarization and coherence evaluation.
- **Contribution to Coherence**:
  - Repeated or related words (e.g., synonyms, antonyms) build logical flow.
  - Enhances readability and comprehension for readers.

---

**24. Why is language modeling important in natural language processing (NLP), and what are its primary applications?**

**CO5, K1, Unit-V**

- **Importance of Language Modeling**:
  - Predicts the probability of word sequences in text.
  - Acts as a foundational component for many NLP tasks.
  - Encodes linguistic patterns and contextual understanding.
- **Applications**:

- Speech recognition (e.g., converting audio to text).
- Machine translation (e.g., improving sentence fluency).
- Text generation (e.g., chatbots and creative writing).
- Spelling and grammar correction.
- Information retrieval and autocomplete systems.

---

**25. What is the role of parameter estimation in language modeling, and how does it affect the accuracy of the model?**

**CO5, K1, Unit-V**

- **Role of Parameter Estimation**:
  - Determines optimal values for model parameters using training data.
  - Influences how well the model captures linguistic patterns.
- **Effect on Accuracy**:
  - Poor estimation leads to overfitting or underfitting.
  - Accurate estimation improves generalization to unseen data.
  - Key techniques include maximum likelihood estimation (MLE) and regularization methods.

---

**26. Extend the difference between statistical and neural language models.**

**CO5, K2, Unit-V**

| **Feature** | **Statistical Language Models** | **Neural Language Models** |
|---|---|---|
| **Approach** | Use probabilities and n-grams. | Leverage neural networks (e.g., RNNs, LSTMs, Transformers). |
| **Context** | Limited to fixed- | Capture long-term |

| Feature | Statistical Language Models | Neural Language Models |
|---|---|---|
| **Capture** | length n-grams. | dependencies. |
| **Feature Engineering** | Requires manual feature selection. | Automatically learns features from data. |
| **Performance** | Less accurate, especially for large vocabularies. | High accuracy and scalability. |
| **Applications** | Basic NLP tasks, smaller datasets. | Advanced NLP tasks like GPT and BERT models. |