

Dr. Babasaheb Ambedkar Open University
Term End Examination January – 2026

Course	: MSCDS	Date	: 18/02/2026
Subject Code	: MSCDS-303	Time	: 03:00am to 05:15pm
Subject Name	: Natural Language Processing	Duration	: 02.15 Hours
		Max. Marks	: 70

Section A

Answer the following (Attempt any three) (30)

1. Explain different types of natural language processing tasks.
2. What are Tree Adjoining Grammars? Explain its strengths and limitations.
3. Write a detailed note on Tokenization and segmentation.
4. Write a detailed note on Parsing.
5. What is Named-Entity Recognition, and why is it important in semantic analysis?

Section B

Answer the following (Attempt any four) (20)

1. What is the need for natural language processing?
2. Explain Applications of Morphology in NLP and Challenges in Morphology for NLP
3. Write a short note on pragmatic analysis.
4. Write a short note on Regular expressions.
5. Explain the importance of n-gram models.
6. Identify and explain common issues in Part-of-Speech tagging with examples.

Section C

Part – A (Multiple Choice Questions)

(10)

- 1 VSM Stands for _____
A Vector Space Models B Vector Size Model
C Vector Sparse Models D Vector Scope Models
- 2 A trigram model considers sequences of _____consecutive words
A Two B Three
C Four D Many
- 3 HMM stands for _____
A Hidden Measure Models B Highest Markov Models
C Hidden Markov Models D Highest Measure Models
- 4 _____ words refer to words that do not appear in the training lexicon or corpus.
A NOUN B VERB
C POS D OVV
- 5 Which of the following is not an application of Vector Space Models?
A Information Retrieval B Document Clustering
C Sentiment Analysis D None of these
- 6 _____is one of the most influential advancements in natural language processing (NLP) for creating vector representations of words
A Word2Vec B DOC2VEC
C Viterbi Algorithm D None of these

- 7 _____ is the process of deconstructing a sentence into its grammatical components according to predefined grammar rules.

A Segmentation	B Parsing
C Paging	D separation
- 8 _____ provides a systematic framework for describing the syntax of languages using formal rules.

A CNF	B CKY
C CFG	D CSG
- 9 There are _____ primary types of summarizations

A Two	B Three
C Four	D five
- 10 LSA stands for _____

A Lexical Semantic Analysis	B Latent Semantic Analysis
C Least Semantic Analysis	D Latest Semantic Analysis

Part B (Do as Directed)

State whether the following statements are true or false.

- 1 TF-IDF scores words based on their importance in a document relative to a larger
collection (corpus).
- 2 MDS is a task in NLP where the goal is to condense information from multiple
documents into a concise and coherent summary.
- 3 Earley parsing operates in four primary stages.
- 4 Clustering is not a foundational method in MDS.
- 5 TextRank, inspired by Google's PageRank,
- 6 Text summarization is widely implemented in real-world applications.
- 7 IR is the science of searching for information within a large collection of resources,
aiming to deliver the most relevant results to a user's query.
- 8 Handling large-scale data is easy
- 9 Term Frequency measures the uniqueness of a term across all documents in the corpus.
- 10 Language barriers do not impede communication
