VIT Vidyalankar Institute of Technology Accredited A+ by NAAC (Autonomous College Affiliated to University of Mumbai)	*Mid Semester Examination (CBSGS-C scheme) -(2022-23)*	
Date:	Time: 1 Hr. & 15 Mints	Branch: CMPN
Semester: VII	Subject: Natural Language Processing (NLP)	Marks: 30

Q. 1)	Attempt any Five (2 Marks Each)	CO	BL
a)	List the phases of NLP.	CO1	L1
b)	What is referential ambiguity in natural language processing? Give one example.	CO1	L1
c)	List the applications of NLP.	CO1	L1
d)	List the pre-processing steps.	CO1	L1
e)	Differentiate between Stemming and Lemmatization with example.	CO2	L2
f)	Differentiate between Inflectional and Derivational Morphology with example	CO2	L1
g)	Apply Porters algorithm on the word "Writing" and “King” to get the stem.	CO2	L3
h)	Define Minimum edit distance used in Spelling correction.	CO2	L1
Q. 2)	Attempt any two. (5 Marks Each)		
a)	Explain Natural Language Generation component of NLP generic block diagram.	CO1	L1
b)	Explain FST in detail with example.	CO2	L2
c)	For a corpus, MLE for bigram “battery life” is 0.27, frequency of “battery is 800. After applying Laplace smoothening the MLE for “battery life” becomes 0.025. What is the vocabulary size of the corpus?	CO2	L3
Q 3)	Attempt any one (10 Marks Each)		
a)	Find the probability of the following sentences 1.<s> Michael and Zack played at the playground </s> 2.<s> Bob went to the school </s> 3.<s> The school was huge</s> 4.<s>Zack went to the playground</s> from the following corpus. Assume a bigram Language model(without Laplace smoothening). <s> the school was open </s> <s> Michael and Zack went to the school </s> <s> the playground at the school was huge </s> <s> Bob and Zack played at the playground </s> <s> Bob Michael and Zack were friends </s> Also find the Perplexity of the sentence which has highest probability.	CO2	L3
b)	Write a note on Language models and evaluation metrics for language model	CO2	L1

CO1	**To describe the field of natural language processing.**
CO2	**To design language model for word level analysis for text processing.**