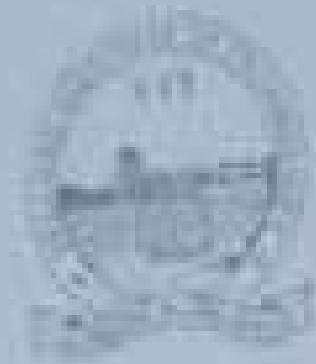# VIT
Vellore Institute of Technology

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING
Continuous Assessment Test - I, January 2020
B. Tech - Winter Semester 2019-20

Course Code : CSE4022	**Duration**	: 90 Minutes
Course Title : Natural Language Processing	**Max. Marks**	: 50
Course Nbr : 2181,2183,2175,6629	**Slot**	: A1

Answer all the questions

1.A. Define ambiguity. How is language ambiguous? (4)

B. Find the type of ambiguity (4)

I. John loved his son, and so did Sam.

II. Mary ate a salad with spinach from California for lunch on Tuesday

C. What is meant by anaphoric ambiguity? (2)

2. Assume that you are consulted by your local police station. They need an insight of types of crimes in your area over the past years. There isn't a ready-made dataset available. But you are free to use digital archives of local newspapers. How will apply the stages of NLP to the digital archive of newspapers? List the different types of insights that you find out as you subject the data to each stage of NLP. Augment your answer with a diagram. (15)

3. A. Are numbers ubiquitous in all types of texts in every language? Justify with an example (5)

B. Define tokenization in sentence segmentation and do the tokenization for the sentence" God is Great! I won the lottery" (4)

C. Find your observation in the following sentence based on the punctuation ambiguities and brief about them. (4)

Peter arrived in Singapore in January 1996, on his twenty-second birthday. Less than a year later, he had married the boss's daughter Yi Ling. I'd like you to meet Mr.Mark Porter, Ms.Elizabeth Taylor, Capt. Eliot Saunders and his wife Mrs.Saunders. I began teaching at UCLA on Mon. 29th Aug. 2018, after five years with UNICEF.

D. ASCII is the encoding standard used to represent _____ in digital form. (1)

E. Text segmentation is the process of converting _____ into its component words and sentences. (1)

4. Explain in detail about inflectional and derivational morphology with suitable detailed example. (10)