Name: K S Saisankalp Davey

Roll number: 2024114010

Date: 26/11/25

Computational Linguistics-2
Endsemester Examination, Form: A

Section 1. Short Answer Questions

1. (6 points) State principles A, B and C defined by the Binding Theory of anaphora resolution.

   (4 points) For each sentence below, decide whether the bolded element is grammatical. If ungrammatical, identify which Binding Principle (A, B, or C) is violated and explain why (1 point per question). (1) $John_i$ said that $Bill_j$ admired $himself_j$. (2) $John_i$ said that $Bill_j$ admired him_i. (3) The $student_i$ blamed $himself_i$. (4) The mother of $John_i$ criticized $himself_i$.

2. (10 points) Annotate the following sentences using Rhetorical Structure Theory Relations given at the end of this question paper.

(1) Title: Bouquets in a basket – with living flowers

(2) There is a gardening revolution going on.

(3) People are planting flower baskets with living plants,

(4) mixing many types in one container for a full summer of floral beauty.

(5) To create your own "Victorian" bouquet of flowers,

(6) choose varying shapes, sizes and forms, besides a variety of complementary colors.

(7) Plants that grow tall should be surrounded by smaller ones and filled with others that tumble over the side of a hanging basket.

(8) Leaf textures and colors will also be important.

(9) There is the silver-white foliage of dusty miller, the feathery threads of lotus vine floating down from above, the deep greens, or chartreuse, even the widely varied foliage colors of the coleus.

3. (10 points) Answer the following questions pertaining to the sentence: *The San Francisco Examiner issued a special edition around noon yesterday.*

(a) (2 points) Draw a context free grammar tree corresponding to the above sentence.

(b) (3 points) Assign propbank roles to relevant constituents.

(c) (5 points) Write a short note on the features commonly used to predict semantic roles using traditional machine learning classifiers (say Logistic Regression).

4. (10 points) Consider the joint distribution $P(X,Y)$ over two random variables: verb-object word order (X) and object pronominality (Y) such that:

   (a) $P(Y = \text{Pronoun}) = 0.238$

   (b) $P(X = \text{Preverbal}|Y = \text{Pronoun}) = 0.941$

   (c) $P(X = \text{Preverbal}|Y = \text{Not Pronoun}) = 0.860$

Imagine you're an incremental sentence processor (i.e. each word is processed one-by-one from left to right). You encounter a transitive verb (i.e verb mandatorily requires an object) but haven't encountered the object yet. Perform inference under uncertainty: How likely is it that the object is a pronoun?

5. (10 points) Answer the following questions based on the following four sentences (**Note: Full marks** require clear and complete reasoning):

(1) *The cat chased the small mouse.*
(2) *The dog chased the big cat.*
(3) *A mouse saw the dog.*
(4) *The small dog chased a cat.*

(5 points) Using a window size of 1 (i.e., only immediate left and right neighbors), construct a **word–word co-occurrence matrix** for the following vocabulary. Ignore capitalization and punctuation.

cat, dog, mouse, chased, small, big

Include only these words as both rows and columns of the matrix. Count each adjacent pair in both directions (the co-occurrence matrix is symmetric).

(5 points) Using the vectors obtained from your co-occurrence matrix:

(a) Compute the **cosine similarity** between the words **cat** and **dog**.
(b) Show all steps in detail, including:
  - the dot product,
  - the magnitudes of the two vectors,
  - the final cosine similarity value.

6. (10 points) A linguist collects a sample corpus and counts the frequencies of all word types. The following table shows the frequency of the top 10 most common words (already sorted by rank):

Rank (r)	Word	Frequency $f(r)$
1	the	5200
2	of	2600
3	and	1700
4	to	1300
5	a	1100
6	in	900
7	that	800
8	is	650
9	for	600
10	with	520

Table 1: Word frequencies for the top 10 ranked words.

(a) (3 points) Explain Zipf's law mathematically

(b) (4 points) Predict the expected frequency of the rank–20 word in the above table as per Zipf's law

(c) (3 points) What are the consequences of Zipf's law for both the speaker and hearer in terms of processing effort?

Nucleus Satellite Relations

Relation Name	Nucleus	Satellite
Antithesis	ideas favored by the author	ideas disfavored by the author
Background	text whose understanding is being facilitated	text for facilitating understanding
Circumstance	text expressing the events or ideas occurring in the interpretive context	an interpretive context of situation or time
Concession	situation affirmed by author	situation which is apparently inconsistent but also affirmed by author
Condition	action or situation whose occurrence results from the occurrence of the conditioning situation	conditioning situation
Elaboration	basic information	additional information
Enablement	an action	information intended to aid the reader in performing an action
Evaluation	a situation	an evaluative comment about the situation
Evidence	a claim	information intended to increase the reader's belief in the claim
Interpretation	a situation	an interpretation of the situation
Justify	text	information supporting the writer's right to express the text
Motivation	an action	information intended to increase the reader's desire to perform the action
Non-volitional Cause	a situation	another situation which causes that one, but not by anyone's deliberate

		action
Non-volitional Result	a situation	another situation which is caused by that one, but not by anyone's deliberate action
Otherwise (anti conditional)	action or situation whose occurrence results from the lack of occurrence of the conditioning situation	conditioning situation
Purpose	an intended situation	the intent behind the situation
Restatement	a situation	a reexpression of the situation
Solutionhood	a situation or method supporting full or partial satisfaction of the need	a question, request, problem, or other expressed need
Summary	text	a short summary of that text
Volitional Cause	a situation	another situation which causes that one, by someone's deliberate action
Volitional Result	a situation	another situation which is caused by that one, by someone's deliberate action

Multinuclear Relations

In addition to this most frequent pattern of nucleus and satellite, there are relations that do not carry a definite selection of one nucleus. We saw an example, Contrast, above. These are called multinuclear relations. They are:

Relation Name	Span	Other Span
Contrast	one alternate	the other alternate
Joint	(unconstrained)	(unconstrained)
List	an item	a next item
Sequence	an item	a next item