

The University of Melbourne

School of Computing and Information Systems

COMP90042

Natural Language Processing

Final Exam

Semester 1 2021

Exam duration: 165 minutes (15 minutes reading time + 120 minutes writing time + 30 minutes technical buffer time)

Length: This paper has 5 pages (including this cover page) and 8 questions. You should attempt all questions.

Instructions to students:

- This exam is worth a total of 120 marks and counts for 40% of your final grade.
- You can read the question paper on a monitor, or print it.
- You are recommended to write your answers on blank A4 paper. Note that some answers require drawing diagrams or tables.
- You will need to scan or take a photo of your answers and upload them via Gradescope. Be sure to label the scans/photos with the question numbers (-10% penalty for each unlabelled question).
- Please answer all questions. Please write your student ID and question number on every page.

Format: Open Book

- While you are undertaking this assessment you are permitted to:
 - make use of the textbooks, lecture slides and workshop materials.
- While you are undertaking this assessment you must not:
 - make use of any messaging or communications technology;
 - make use of any world-wide web or internet-based resources such as wikipedia, stackoverflow, or google and other search services;
 - act in any manner that could be regarded as providing assistance to another student who is undertaking this assessment, or will in the future be undertaking this assessment.
- The work you submit must be based on your own knowledge and skills, without assistance from any other person.

COMP90042 Natural Language Processing

Semester 1, 2021

Total marks: 120 (40% of subject)

Students must attempt all questions

Section A: Short Answer Questions [45 marks]

Answer each of the questions in this section as briefly as possible. Expect to answer each sub-question in no more than several lines.

Question 1: General Concepts [24 marks]

- a) What is a “sequence labelling” task and how does it differ from independent prediction? Explain using “part-of-speech tagging” as an example. [6 marks]
- b) Compare and contrast “antecedent restrictions” and “preferences” in “anaphora resolution”. You should also provide examples of these restrictions and preferences. [6 marks]
- c) What is the “exposure bias” problem in “machine translation”? [6 marks]
- d) Why do we use the “IOB tagging scheme” in “named entity recognition”? [6 marks]

Question 2: Distributional Semantics [9 marks]

- a) How can we learn “word vectors” using “count-based methods”? [6 marks]
- b) Qualitatively, how will the word vectors differ when we use “document” vs. “word context”? [3 marks]

Question 3: Context-Free Grammar [12 marks]

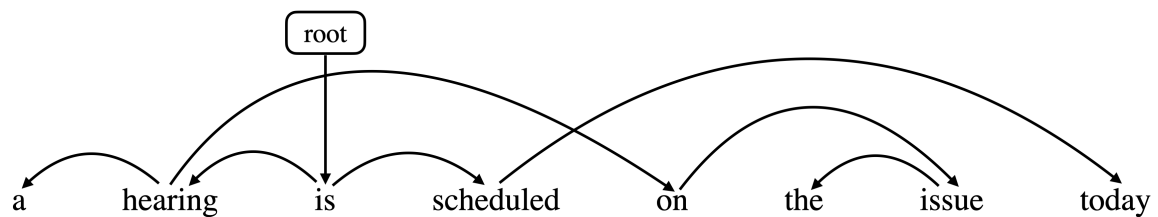
- a) Explain two limitations of the “context-free” assumption as part of a “context-free grammar”, with the aid of an example for each limitation. [6 marks]
- b) What negative effect does “head lexicalisation” have on the grammar? Does “parent conditioning” have a similar issue? You should provide examples as part of your explanation. [6 marks]

Section B: Method Questions [45 marks]

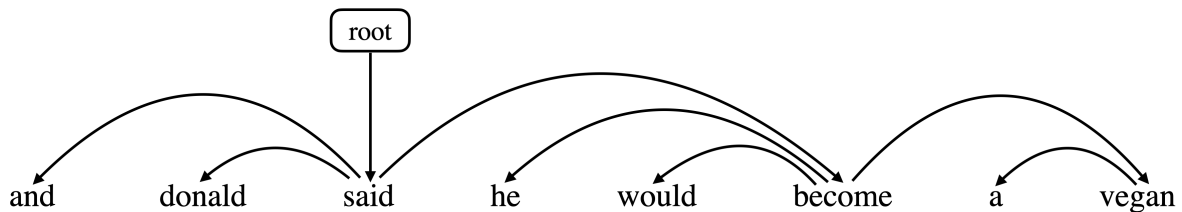
In this section you are asked to demonstrate your conceptual understanding of the methods that we have studied in this subject.

Question 4: Dependency Grammar [18 marks]

- What is “projectivity” in a dependency tree, and why is this property important in dependency parsing? [3 marks]
- Which arc or arcs are “non-projective” in the following tree? Explain why they are non-projective. [6 marks]



- Show a sequence of parsing steps using a “transition-based parser” that will produce the dependency tree **below**. Be sure to include the state of the stack and buffer at every step. [9 marks]



Question 5: Loglikelihood Ratio [15 marks]

The “loglikelihood ratio” is used in summarisation to measure the “saliency” of a word compared to a background corpus. In the second task of the project, to understand the nature of rumour vs. non-rumour source tweets, one analysis we can do is to extract salient hashtags in rumour source tweets and non-rumour source tweets to understand the topical differences between them. Illustrate **with an example with equations** how you can apply loglikelihood ratio to extract salient hashtags in these two types of source tweets.

Question 6: Ethics [12 marks]

You’re tasked to develop an NLP application to predict the “intelligence quotient (IQ)” scores of high school students based on their essays written for a range of topics. Discuss at least **three** ethical implications of this application.

Section C: Algorithmic Questions [30 marks]

In this section you are asked to demonstrate your understanding of the methods that we have studied in this subject, in being able to perform algorithmic calculations.

Question 7: N -gram Language Models [15 marks]

This question asks you to calculate the probability for “ N -gram language models”. You should leave your answers as fractions. Consider the following table, which collects the counts of words that occur after **salted** in a corpus.

| Word | Count | Unsmoothed Probability | Smoothed Probability | |
|---------|-------|------------------------|----------------------|--------------|
| | | | Absolute Discounting | Katz Backoff |
| egg | 6 | ? | ? | ? |
| caramel | 4 | ? | ? | ? |
| fish | 3 | ? | ? | ? |
| peanuts | 2 | ? | ? | ? |
| butter | 0 | ? | ? | ? |
| salted | 0 | ? | ? | ? |

E.g. the bigram **salted egg** occurs 6 times, while **salted caramel** occurs 4 times.

- Assuming the 6 distinct words in the table are all the words in vocabulary, compute the bigram probabilities for all the bigrams listed in the table without any smoothing. Hint: you should fill in the missing values for the “Unsmoothed Probability” column in the table, and demonstrate how you arrive at these values. [3 marks]
- Compute the bigram probabilities for all bigrams listed in the table using “absolute discounting”, with a discount factor of 0.2. Hint: you should fill in the missing values for the “Absolute Discounting” column in the table, and demonstrate how you arrive at these values. [6 marks]
- Compute the bigram probabilities for all bigrams listed in the table using “Katz Backoff”, with the same discount factor of 0.2. Use the corpus below (2 sentences) for computing unigram probabilities. For simplicity, you do not need to consider special tokens (ending or starting tokens), and may assume all the unique words in the 2 sentences as your vocabulary when computing the unigram probabilities. Hint: you should fill in the missing values for the “Katz Backoff” column in the table, and demonstrate how you arrive at these values. [6 marks]

butter in batter will make batter salted
but better butter will make batter better

Question 8: Topic Models [15 marks]

Consider training a “latent Dirichlet allocation” (LDA) topic model using the following corpus with 3 documents (d_1 , d_2 , d_3). To initialise the training process, each word token is randomly allocated to a topic (e.g. **peck/t3** means **peck** is assigned topic **t3**). Hyper-parameters of the topic model are set as follows: (1) number of topics $T = 3$; (2) document-topic prior $\alpha = 0.5$; and (3) topic-word prior $\beta = 0.1$.

```
d1: peck/t3 pickled/t1 peppers/t1
d2: peter/t1 piper/t2 picked/t3 peppers/t2
d3: peppers/t2 piper/t3 peck/t3 peppers/t1
```

- a) Compute the probability over the topics (**t1**, **t2**, **t3**) if you were to sample a new topic for the first word (**peck**) in **d1** for a training step. You should show co-occurrence tables that are relevant to producing your solution. [9 marks]
- b) Assume now that the topic model is trained. You are now given a new document: **pickled peppers popped**. Describe how LDA infers the topics for this new document. Note: you **do not** need to show equations or tables here. [6 marks]

— End of Exam —