First Name: ______________________ **Last Name**: ______________________

McGill ID: ______________________ **Section**: ______

Faculty of Science
COMP-599 - Introduction to Natural Language Processing (Fall 2016)
Midterm Examination

November 9th, 2016
13:05 – 14:25

Examiner: Jackie Chi Kit Cheung

Instructions:

- **DO NOT TURN THIS PAGE UNTIL INSTRUCTED**
- This is a **closed book** examination.
- Only writing implements (pens, pencils, erasers, pencil sharpeners, etc.) are allowed. The possession of any other tools or devices is prohibited.
- Answer **all** questions **on this examination paper** and return it.
- This examination has **14** pages including this cover page, and is printed on both sides of the paper.
- **MAKE SURE TO WRITE YOUR NAME AND STUDENT ID ON THE EXAM. MARKS WILL BE DEDUCTED IF INFORMATION IS MISSING.**

Sections

The exam consists of the following sections:

1. Multiple Choice: Questions 1 to 15
2. Short Answer: Questions 16 to 19
3. Problem Sets: Questions 20 to 22

Multiple Choice Questions (1 point each)

Circle the correct response.

1. Which of the following is **not** an example of a dynamic programming algorithm?
   (A) CYK algorithm
   (B) Eisner's algorithm
   (C) Viterbi algorithm
   (D) Yarowsky's algorithm

2. The study of the meaning of language in context is called:
   (A) Morphology
   (B) Pragmatics
   (C) Semantics
   (D) Syntax

3. An automatic tool has converted the input `itemized` into the output `item`. What task has this automatic tool performed?
   (A) Lemmatization
   (B) Morphological recognition
   (C) Parsing
   (D) Stemming

4. Assuming the same vocabulary and the same assumptions about the distributions responsible for generating samples, which of the following parameter estimation techniques will result in the greatest test corpus likelihood?
   (A) Maximum likelihood estimation (MLE)
   (B) Maximum a posteriori estimation (MAP)
   (C) Bayesian inference
   (D) It is impossible to say, with the given information

5. Let $\alpha_j(t)$ be the cell in the forward trellis of a Hidden Markov Model representing the state j at timestep t, and let $\delta_j(t)$ be the cell in the trellis when running the Viterbi algorithm, with the same state j and timestep t. What can we say about the relationship between $\alpha_j(t)$ and $\delta_j(t)$?
   (A) $\alpha_j(t) < \delta_j(t)$
   (B) $\alpha_j(t) = \delta_j(t)$
   (C) $\alpha_j(t) > \delta_j(t)$
   (D) None of the above

6. I would like to train a system that will predict the score that a game reviewer assigns to a newly released game, from 0.0 to 100.0, given the text of the review. This is best cast as what kind of problem?
   (A) Bootstrapping
   (B) Classification
   (C) Feature extraction
   (D) Regression

7. Let unary predicates $A(x)$ and $B(x)$ represent concepts, such that A is a hypernym of B. Which of the following statements is true?

   (A) $A(x) \rightarrow B(x)$
   (B) $A(x) \leftarrow B(x)$
   (C) $A(x) \leftrightarrow B(x)$
   (D) $A(x) \vee B(x)$

8. What is one reason to prefer using Context Free Grammars over Finite State Automata as a formal model of natural language?

   (A) CFGs overgenerate less than FSAs.
   (B) CFGs undergenerate less than FSAs.
   (C) CFGs can account for linguistic constructions that FSAs cannot.
   (D) There are treebanks labelled with CFG parse trees, and there are not for FSAs.

9. A constituent parser predicts only one constituent per sentence, which spans the entire sentence. If we ignore labels in evaluation, this parser will achieve perfect:

   (A) Accuracy
   (B) F1
   (C) Precision
   (D) Recall

10. Which of the following tasks is the IOB tagging scheme useful for?

    (A) Dependency parsing
    (B) NP chunking
    (C) Part of speech tagging
    (D) All of the above

11. "An even number is an integer that is evenly divisible by two." This is an example of a(n):

    (A) Extensional definition
    (B) Intensional definition
    (C) Logical definition
    (D) Prototypical definition

12. Which of the following is **not** part of a synset entry in WordNet?

    (A) A definition of that synset.
    (B) A list of all the senses of a word.
    (C) A list of synonyms.
    (D) Links to other synsets labelled by a lexical semantic relation.

13. Which of the following causes a problem for Eisner's algorithm?

    (A) Backpointers
    (B) Crossing dependencies
    (C) Edge-factored independence assumptions
    (D) Projective trees

14. Which of the following is **not** an example of a referring expression?

    (A) Demonstrative
    (B) Pronoun
    (C) Proper name
    (D) Quantifier

15. Consider the sentence "Bob passed the exam." If we annotated the sentence with a semantic role labeller, what would be its output?

    (A) "Bob": agent, "the exam": patient
    (B) "Bob": subject, "the exam": direct object
    (C) "Bob": NP, "the exam": NP
    (D) "Bob": `Person`, "the exam": -

Short Answer

16. Give **two** reasons why it is problematic to interpret the probability that we get from an N-gram language model as a grammaticality score. (2 points)

17. Perform beta reduction on the following lambda calculus expressions as much as possible. (4 points)

a) $(\lambda z.z)(\lambda q.q\ q)(\lambda s.s\ a)$

b) $(\lambda s.\lambda q.s\ q\ q)(\lambda q.q)\ q$

18. The following is the training set for a system that classifies whether a sentence contains a question or a statement. Specify two Naive Bayes models with categorical distributions using bag-of-words features. Do not lemmatize. Use the vocabulary found in the training set, including a special UNK feature for unseen words. Use a) maximum likelihood estimation, and b) Laplace smoothing. Show your work.

Text	**Label**
what did you study	QUESTION
she asked who left	QUESTION
what is left to study	QUESTION
she left to study	STATEMENT
you did study	STATEMENT

a) MLE (3 points)

b) Laplace (3 points)

19. Consider the following statements, in first-order logic (FOL).

$$\exists x.\mathit{Country}(x) \tag{1}$$
$$\forall x.\mathit{Country}(x) \rightarrow \exists y.\mathit{City}(y) \wedge \textsc{Capital}(x) = y \tag{2}$$
$$\forall x.\mathit{Country}(x) \leftrightarrow \neg \mathit{City}(x) \tag{3}$$

Specify an interpretation of the FOL such that the above statements are true. The only requirement is that the discourse domain must contain at least one element. You may assume that equality $=$ is defined as expected. (3 points)

Problem Sets

20. Consider the following context-free grammar:

Grammar:

S $\rightarrow$ NP VP	Det $\rightarrow$ a
NP $\rightarrow$ N N	N $\rightarrow$ fruit | flies | banana
NP $\rightarrow$ Det N	P $\rightarrow$ like
NP $\rightarrow$ N	V $\rightarrow$ flies | like
PP $\rightarrow$ P NP	
VP $\rightarrow$ V PP	
VP $\rightarrow$ V NP	

a) One of the rules above is not in Chomsky-Normal Form (CNF). Correct the above grammar by removing and adding rules such that it is in CNF. (2 points)

b) Apply the CYK parsing algorithm to the target sentence:

Fruit flies like a banana.

Don't forget the backpointers. **Give the final output parse(s) after converting the parse back to the original, non-CNF grammar**. (6 points)

(More space for the previous question, if needed)

21. You have been hired as a consultant who is to design a system for extracting named entities from text. You are given a large corpus with input-output pairs. Here is an example:

    Input: *Hillary Clinton and Donald J. Trump hopscotched from Pennsylvania to North Carolina to Michigan on Monday in the final, frenzied hours of the presidential campaign, offering clashing closing arguments as the sprawling map of the United States was reduced to a string of must-win states.*

    Output: HILLARY CLINTON, DONALD J. TRUMP, PENNSYLVANIA, NORTH CARLINA, MICHIGAN, UNITED STATES

    How would you propose to design such a system? Answer briefly in point form. Include information on:

    - the formal definition of the problem
    - some features that you would use
    - the learning algorithm

    State, to the best of your ability, the assumptions behind your approach, and its possible advantages and limitations. (6 points)

(More space for the previous question, if needed)

22. Assume that the following (Python) recursive data structures are available for representing constituent and dependency trees:

```
class ConstituentTree:
  .node # str representing the non-terminal category (for non-terminals)
        or a word (for terminals)
  .rhs  # list whose elements are of type ConstituentTree

class DependencyTree:
  .word # str representing the current word
  .deps # list of DependencyTree instances representing the dependents
        of this node
```

These data structures are recursive. To store a tree, all that is needed is a pointer to the root of the tree. For example, the constituent tree `(NP (N hats))` can be stored as a ConstituentTree whose `.node` is `'NP'`, and whose `.rhs` is a list of one element corresponding to a ConstituentTree for `(N hats)`. The ConstituentTree for `(N hats)` has `.node` with a value of `'N'`, and a `.rhs` containing one str element (`'hats'`). Dependency trees are similarly defined, but the `.word` represents the current word, rather than a non-terminal label, because nodes in a dependency tree are words.

Further assume that the following function which implements head-finding is available. Given a ConstituentTree instance `t`, the function returns the index of `t.rhs` corresponding to the head of that rule.

```
def find_head: ConstituentTree -> int
  # returns the index of the .rhs that is the head of this rule,
  or -1 if .rhs is empty
```

Write an algorithm in pseudocode for converting a constituent tree into a dependency tree. (6 points)

```
def const2dep: ConstituentTree -> DependencyTree
  t <- the input ConstituentTree
  # continue here
```

(More space for the previous question, if needed)

This page is left intentionally blank. You may do rough work on it.