



### **Autumn Examinations 2019/2020**

**Course Instance Code(s) Exam(s)** 1CSD1, 1CSD2, 1SPE1, 1MAO2, 1MAI1  
MSc in Computer Science (Data Analytics), MSc in Computer Science (Artificial Intelligence), MSc in Computer Science (Artificial Intelligence) - Online

**Module Code(s) Module(s)** CT5120, CT5146  
Introduction to Natural Language Processing, Introduction to Natural Language Processing - Online

Paper No. 1  
Repeat Paper Yes

External Examiner(s) Professor Pier Luca Lanzi  
Internal Examiner(s) Dr. Michael Madden  
\*Dr. Paul Buitelaar, Dr. John McCrae

**Instructions:** Answer all parts of all questions. There are 4 sections; each section is worth 25 marks (100 marks total). This is an open-book examination, please upload all answers as a single file

**Duration** 2 hours  
**No. of Pages** 5  
**Discipline(s)** Computer Science  
**Course Co-ordinator(s)** Dr. Enda Howley, Dr. Michael Schukat, Dr. James McDermott

### **Requirements:**

Release in Exam Venue	Yes	<input checked="" type="checkbox"/>	No	<input type="checkbox"/>
MCQ	Yes	<input type="checkbox"/>	No	<input checked="" type="checkbox"/>
Handout	None			
Statistical/ Log Tables	None			
Cambridge Tables	None			
Graph Paper	None			
Log Graph Paper	None			
Other Materials	Yes	<input type="checkbox"/>	No	<input checked="" type="checkbox"/>
Graphic material in colour	Yes	<input type="checkbox"/>	No	<input type="checkbox"/>



**Plagiarism policy**

Please include the following statement in your submission:

*In submitting this work I confirm that it is entirely my own. I acknowledge that I may be invited to online interview if there is any concern in relation to the integrity of my exam*

# Introduction to Natural Language Processing

Exam Duration: 2 Hours

**You must complete Sections 1 to 4**

## Section 1: Linguistics

**Instructions:** Provide answers for questions 1A, 1B and 1C

### Question 1A

**5 Marks**

How many tokens are there in this sentence. Explain your reasoning.

*Boris Johnson has been the prime minister of the UK since last year.*

### Question 1B

**10 Marks**

Fill in the blanks in these statements:

*government* and *govern* are morphologically related through \_\_\_\_

*minister* and *ministers* are morphologically related through \_\_\_\_

*road* and *roadmap* are morphologically related through \_\_\_\_

### Question 1C

**10 Marks**

Describe in your own words the difference between a parallel and comparable corpus. Give an example of an NLP application that uses such corpora.

**PTO**

## Section 2: Parsing

**Instructions:** Provide answers for question 2A, 2B and 2C

### Question 2A

**10 Marks**

Consider the following grammar:

Rule	Probability	Rule	Probability
$S \rightarrow NP VP$	1.0	$D \rightarrow the$	0.5
$NP \rightarrow D N$	0.4	$D \rightarrow a$	0.5
$NP \rightarrow N$	0.5	$N \rightarrow coffee$	0.3
$NP \rightarrow Prn$	0.1	$N \rightarrow function$	0.7
$VP \rightarrow V$	0.3	$V \rightarrow function$	0.6
$VP \rightarrow Aux VP$	0.2	$V \rightarrow can$	0.4
$VP \rightarrow Adv VP$	0.1	$Prn \rightarrow I$	1.0
$VP \rightarrow V PP$	0.4	$Aux \rightarrow can$	1.0
$PP \rightarrow Prp NP$	1.0	$Adv \rightarrow only$	1.0
		$Prp \rightarrow with$	1.0

What is the probability of the following sentence in this grammar?

*I can only function with coffee*

Show which rules in the grammar were used in the parse tree of this sentence.

### Question 2B

**10 Marks**

Using the grammar of Question 2A, find a sentence that is accepted by the grammar but is not grammatical in English and suggest a modification to the grammar so that this sentence is not generated.

### Question 2C

**5 Marks**

Why do lexical dependencies cause an issue with a simple Probabilistic Context-Free Grammar (PCFG) approach to parsing?

**PTO**

### Section 3: Semantics

**Instructions:** Provide answers for question 3A, 3B, 3C and 3D

#### Question 3A

**5 Marks**

Fill in the blanks in these statements on words that are semantically related by stating the kind of semantic relation:

*government, cabinet, administration* are \_\_\_\_

*light* and *dark* are \_\_\_\_

#### Question 3B

**5 Marks**

Explain in your own words how word senses are represented in WordNet. Give an example.

#### Question 3C

**5 Marks**

Explain in your own words how word senses are represented in FrameNet. Give an example.

#### Question 3D

**10 Marks**

How can Wikipedia be used in word sense disambiguation?

**PTO**

## Section 4: Applications

**Instructions:** Provide answers for questions 4A, 4B and 4C

### Question 4A

**10 Marks**

Explain in your own words how a knowledge model can be used in information extraction. Give an example.

### Question 4B

**10 Marks**

Given words  $a, b, c$ , explain how  $PMI(a, b)$  for a given corpus can be higher than  $PMI(a, c)$ . [PMI = Pointwise Mutual Information]

### Question 4C

**5 Marks**

Discuss a limitation of a lexicon-based approach to sentiment analysis.

**END**