NATIONAL UNIVERSITY OF SINGAPORE
SCHOOL OF COMPUTING

CS4248 – Natural Language Processing

Semester 1 AY2018/2019

November 2018 **Time Allowed: 2 Hours**

INSTRUCTIONS TO CANDIDATES

1. This assessment paper contains **SEVEN (7)** questions and comprises **TEN (10)** printed pages, including this page.
2. Answer **ALL** questions within the space in this booklet.
3. This is a **CLOSED** book assessment, but one double-sided A4 sized sheet is allowed for notes.
4. A non-programmable calculator is permitted.
5. Please write your Student Number below. Do not write your name.

This portion is for lecturer's use only

Question	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Total
Max	15	15	20	10	10	10	20	100
Marks								

1. (15 marks) Give a trace of the minimum edit distance algorithm (a dynamic programming algorithm) to compute the minimum cost of transforming the string "delta" to "elites", by filling out every cell entry in the following table, where each cell entry denotes the minimum cost of transforming the associated substrings. Assume that the cost of inserting a character is 1, the cost of deleting a character is 1, and the cost of substituting a character by a different character is 2. (You do not need to show the optimal path.)

a	5						
t	4						
l	3						
e	2						
d	1						
	0	1	2	3	4	5	6
		e	l	i	t	e	s

2. (15 marks) Assign one part-of-speech (POS) tag to each word in bold in the following 15 sentences, using the Penn Treebank tagset. Write the POS tag next to each word in bold in the table below.

(a) He goes to school **every** morning .
(b) She did **not** meet the parents .
(c) **Who** knows what will happen ?
(d) She is so patient , **like** her mother .
(e) He was **there** just now .
(f) He was there **just** now .
(g) He was there just **now** .
(h) Peter was late **because** the train broke down .
(i) **According** to him , she left yesterday .
(j) She left **many** years ago .
(k) He showed **up** at the meeting finally .
(l) She is learning **how** to skate .
(m) **Either** way is fine .
(n) The index gained **0.98** to 318.20 .
(o) **These** things can happen .

	word	POS tag		word	POS tag		word	POS tag
(a)	**every**		(f)	**just**		(k)	**up**	
(b)	**not**		(g)	**now**		(l)	**how**	
(c)	**Who**		(h)	**because**		(m)	**Either**	
(d)	**like**		(i)	**According**		(n)	**0.98**	
(e)	**there**		(j)	**many**		(o)	**These**	

The 45 POS tags in the Penn Treebank tagset are:

CC (Coordin. Conjunction)	PDT (Predeterminer)	VBP (Verb, non-3sg pres)
CD (Cardinal number)	POS (Possessive ending)	VBZ (Verb, 3sg pres)
DT (Determiner)	PRP (Personal pronoun)	WDT (Wh-determiner)
EX (Existential 'there')	PRP$ (Possessive pronoun)	WP (Wh-pronoun)
FW (Foreign word)	RB (Adverb)	WP$ (Possessive wh-)
IN (Preposition/sub-conj)	RBR (Adverb, comparative)	WRB (Wh-adverb)
JJ (Adjective)	RBS (Adverb, superlative)	$ (Dollar sign)
JJR (Adj., comparative)	RP (Particle)	# (Pound sign)
JJS (Adj., superlative)	SYM (Symbol)	" (Left quote)
LS (List item marker)	TO ("to")	" (Right quote)
MD (Modal)	UH (Interjection)	((Left parenthesis)
NN (Noun, sing. or mass)	VB (Verb, base form)	) (Right parenthesis)
NNS (Noun, plural)	VBD (Verb, past tense)	, (Comma)
NNP (Proper noun, singular)	VBG (Verb, gerund)	. (Sentence-final punc)
NNPS (Proper noun, plural)	VBN (Verb, past participle)	: (Mid-sentence punc)

3. (20 marks) The Earley algorithm for parsing context-free grammars, given in the textbook of Jurafsky and Martin, is reproduced below:

---

function EARLEY-PARSE(*words*, *grammar*) returns *chart*
ENQUEUE(($\gamma \rightarrow \bullet$ S, [0,0]), *chart*[*0*])
for *i* ← **from** 0 **to** LENGTH(*words*) **do**
  for each *state* in *chart*[*i*] **do**
    if INCOMPLETE?(*state*) **and** NEXT-CAT(*state*) is not a part of speech **then**
      PREDICTOR(*state*)
    elseif INCOMPLETE?(*state*) **and** NEXT-CAT(*state*) is a part of speech **then**
      SCANNER(*state*)
    else
      COMPLETER(*state*)
  end
end
return(*chart*)

procedure PREDICTOR((A → $\alpha \bullet$ B β, [*i*,*j*]))
  for each (B → γ) **in** GRAMMAR-RULES-FOR(B, *grammar*) **do**
    ENQUEUE((B → $\bullet\ \gamma$, [*j*,*j*]), *chart*[*j*])
  end

procedure SCANNER((A → $\alpha \bullet$ B β, [*i*,*j*]))
  if B ⊂ PARTS-OF-SPEECH(*word*[*j*]) **then**
    ENQUEUE((B → *word*[*j*] $\bullet$, [*j*,*j*+*1*]), *chart*[*j*+*1*])

procedure COMPLETER((B → $\gamma \bullet$, [*j*,*k*]))
  for each (A → $\alpha \bullet$ B β, [*i*,*j*]) in *chart*[*j*] **do**
    ENQUEUE((A → α B $\bullet\ \beta$, [*i*,*k*]), *chart*[*k*])
  end

procedure ENQUEUE(*state*, *chart-entry*)
  if *state* is not already in *chart-entry* **then**
    APPEND(*state*, *chart-entry*)
  end

Let *n* be the number of words in an input sentence, and assume the grammar size is constant. What is the time complexity of the Earley algorithm on an input sentence of *n* words (in terms of the order big-*O* notation)? Clearly justify your answer.

(Note that $1^2 + 2^2 + \cdots + n^2 = \frac{1}{3}n(n+\frac{1}{2})(n+1)$)

(Additional space for answering question 3)

4. (10 marks) Consider the following probabilistic context-free grammar in Chomsky Normal Form:

S → NP VP	[0.4]
S → V NP	[0.4]
S → time	[0.1]
S → flies	[0.1]
NP → N N	[1.0]
VP → V NP	[0.5]
VP → time	[0.3]
VP → flies	[0.2]
N → time	[0.4]
N → flies	[0.4]
N → fruit	[0.2]
V → time	[0.6]
V → flies	[0.4]

Give a trace of the probabilistic CKY algorithm on the input "time fruit flies". Draw the parse tree with the highest probability. What is the probability of this parse tree?

(Additional space for answering question 4)

5. For each of the following garden-path sentences, draw the initial (incorrect) context-free phrase structure parse tree formed during human parsing, and the final (correct) context-free phrase structure parse tree of the whole sentence. Indicate clearly the part-of-speech tag of each word and the non-terminal symbols in each parse tree.

(a) (5 marks) "The old man the boat"

(b) (5 marks) “We painted the wall with cracks”

6. (a) (2 marks) What is metonymy? Give your answer below in no more than 25 words.

Give the meaning representations of the following English sentences in first order logic, while still respecting the constraints that the agent of an announcing event must be a human (not an organization), and the theme of a reading event must be some reading material (not a person).

(b) (4 marks) "Intel announced a new microprocessor."

(c) (4 marks) "John reads Shakespeare."

7. For each part of this question, your answer should be in no more than 60 words.
(a) (5 marks) What is the key subtask common to both speech recognition and machine translation? Provide a brief justification.

(b) (5 marks) Suppose chart parsing is employed to parse the input sentence "She likes apples" using a context-free grammar. Describe what the initialization step of chart parsing will do. Be as precise as possible.

(c) (5 marks) When training a neural network using stochastic gradient descent, it was observed that the value of the loss function on the training examples increases with increasing number of epochs. Explain the cause of this observation.

(d) (5 marks) What is the key advantage of using the rectified linear unit (ReLU) as activation function compared to using the sigmoid function?

END OF PAPER