



Cairo University  
Faculty of Computers and Information



## Final Exam

Department: computer science

Course Title: Natural Language Processing

Course Code: CS462

Semester: Fall

Instructor: Dr. Hanaa Bayomi Ali

Date: 18/5/2019

Exam Duration: 2 Hours

معالجة اللغات الطبيعية  
الفرقة الثالثة والرابعة

### تعليمات هامة

- حيازة التليفون المحمول مفتوحا داخل لجنة الامتحان يعتبر حالة غش تستوجب العقاب وإذا كان ضروري الدخول بالمحمول فيوضع مغلوق في الحقيبة.
- لا يسمح بدخول سماعة الأذن أو البلوتوث.
- لا يسمح بدخول أي كتب أو ملازم أو أوراق داخل اللجنة والمخافاة تعتبر حالة غش.

60

| Question    | Mark | Signature |
|-------------|------|-----------|
| One         |      |           |
| Two         |      |           |
| Three       |      |           |
| Four        |      |           |
| Five        |      |           |
| Six         |      |           |
| Seven       |      |           |
| Eight       |      |           |
| Nine        |      |           |
| Ten         |      |           |
| Total Marks |      |           |

Total Marks in Writing: \_\_\_\_\_



مستوعب الكتابة على هذا المثلث

**Question 1 [16 marks]**

Complete the context-free grammar below so that it generates (at least) the sentences listed.

**GRAMMAR**  
 S → NP VP  
 NP → Det N  
 NP → PN  
 VP → TV NP  
 VP → DV NP PP  
 Det → the  
 Det → a  
 N → mouse  
 N → cat  
 PN → Tom  
 PN → Jerry  
 N → cheese  
 IV → sleeps  
 TV → eats  
 DV → steals

**SENTENCES**

The cat sleeps.  
 The mouse eats the cheese.  
 Tom steals the cheese from the mouse.

[1.5 marks]

a)

b) Assign parse trees to all three sentences. [6 marks]

The cat sleeps

The mouse eats the cheese

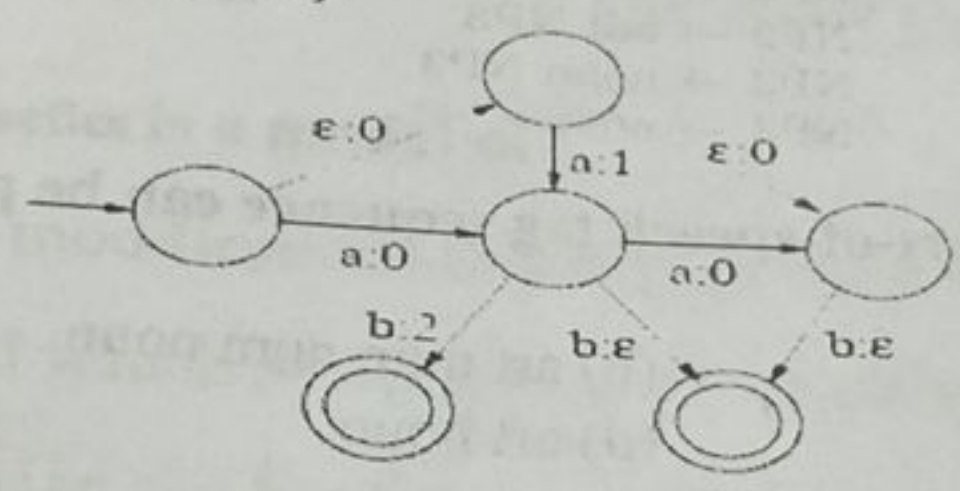


How many different parse trees are there for the string  $\langle\langle a \rangle; a \rangle$  ?  
 (b) 1  
 (d) 3

More than three.  
 Which of the following English words ?  
 (b) time  
 (d) therefore

Which of the following strings does not match the regular expression written in machine syntax as  $(0^*1?2+)^*$   
 (a)  $\epsilon$   
 (c) 12  
 (e) 12222  
 (b) 02  
 (d) 01201

Consider the following finite state transducer with input alphabet  $\{a, b, c\}$  and output alphabet  $\{0, 1, 2\}$ :



Which of the following is *not* a possible output string that could arise from processing the input string  $ab$

- $\{a, b\}$
- (a) 0
  - (b) 00
  - (c) 01
  - (d) 02
  - (e) 012

Consider the following probabilistic context-free grammar:

|      |               |        |       |
|------|---------------|--------|-------|
| $S$  | $\rightarrow$ | $N VP$ | (1.0) |
| $VP$ | $\rightarrow$ | $IV$   | (0.8) |
| $VP$ | $\rightarrow$ | $TV N$ | (0.2) |
| $N$  | $\rightarrow$ | clocks | (0.7) |
| $N$  | $\rightarrow$ | tables | (0.3) |
| $IV$ | $\rightarrow$ | tick   | (0.7) |
| $IV$ | $\rightarrow$ | fly    | (0.3) |
| $TV$ | $\rightarrow$ | hate   | (0.7) |
| $TV$ | $\rightarrow$ | chase  | (0.3) |



Which of the following sentences is assigned the highest probability by this grammar?

- (a) clocks tick
- (b) tables fly
- (c) clocks fly
- (d) tables chase clocks
- (e) tables hate clocks

9) Which of the following sets of productions is not in Chomsky normal form?

- (a) G1
- (b) G2
- (c) G3
- (d) G1 and G3
- (e) G2 and G3

G1 :

$S \rightarrow AB$

$A \rightarrow AB|a$

$B \rightarrow Ba|b$

G2 :

$S \rightarrow AB$

$A \rightarrow AB|a$

$B \rightarrow BA|b$

G3 :

$S \rightarrow A|B$

$A \rightarrow AB|a$

$B \rightarrow BA|b$

10) Consider the grammar below:

NP  $\rightarrow$  art NP1  
 NP  $\rightarrow$  ppro NP1  
 NP1  $\rightarrow$  num NP1  
 NP1  $\rightarrow$  NP2  
 NP2  $\rightarrow$  adj NP2  
 NP2  $\rightarrow$  adj NP3  
 NP3  $\rightarrow$  noun NP3  
 NP3  $\rightarrow$  noun

Which of the following part-of-speech tag sequence can be produced by grammar?

- (a) art num noun
- (b) art num num noun
- (c) art adj noun
- (d) art noun
- (e) art adj num noun

11) What type of ambiguity exists in the word sequence "Time flies"?

- (a) Syntactic
- (b) Semantic
- (c) Phonological
- (d) Anaphoric
- (e) none of the above

12) *computer vs computational* is an example of \_\_\_\_\_ morphology.

- a) Inflectional
- b) Derivational
- c) Cliticization
- d) None of the above

13) What is the number of trigrams in a normalized sentence of length  $n$  words?

- (a)  $n$
- (b)  $n-1$
- (c)  $n-2$
- (d)  $n-3$

14) Assume that there are 10000 documents contain the terms "d" 3 times in a particular document, (a) 8.11 (c) 0

15) Let us suppose that you have vectors for two words  $w_1$  and  $w_2$   $w_1 = (0.2, 0.1, 0.3, 0.4)$  What is the cosine similarity? (a) 0.948 (c) 0

16) When training a language corpus, the probabilities (a) Don't reflect the task (c) Reflect intuition

17) Morphotactics is a morphological process (a) Spelling modification (b) How and which (c) All affixes in the word (d) Ngrams of affixes

18) You have collected information. You want to select the tweets in three following models can be mentioned above? (a) Naïve Bayes (c) Language model

19) \_\_\_\_\_ category and affixation (a) Inflectional (c) Cliticization



the highest probability by this

fly  
chase clocks

in Chomsky normal form?

G1: S → AB  
A → AB|a  
B → Ba|b

G2: S → AB  
A → AB|a  
B → BA|b

G3: S → A|B  
A → AB|a  
B → BA|b

produced by

e flies"?

hology.

th n

Assume that there are 10000 documents in a collection. Out of these, 10 documents contain the terms "difficult task". If "difficult task" appears 1 times in a particular document, what is the TFIDF value of the terms for that document?

- (b) 15.87
- (d) 81.1

Let us suppose that you have the following two 4-dimensional word vectors for two words  $w_1$  and  $w_2$  respectively:  
 $w_1 = (0.2, 0.1, 0.3, 0.4)$  and  $w_2 = (0.3, 0, 0.2, 0.5)$

What is the cosine similarity between  $w_1$  and  $w_2$ ?

- (a) 0.948
- (b) 0.832
- (c) 0
- (d) 0.5

16) When training a language model, if we use an overly narrow corpus, the probabilities

- (a) Don't reflect the task
- (b) Reflect all possible wordings
- (c) Reflect intuition
- (d) Don't generalize

17) Morphotactics is a model of

- a) Spelling modifications that may occur during affixation
- b) How and which morphemes can be affixed to a stem
- c) All affixes in the English language
- d) Ngrams of affixes and stems

18) You have collected a data of about 10,000 rows of tweet text and no other information. You want to create a tweet classification model that categorizes each of the tweets in three buckets – positive, negative and neutral. Which of the following models can perform tweet classification with regards to context mentioned above?

- a) Naïve Bayes
- b) Support Vector Machine
- c) Language model
- d) None of the above

19) \_\_\_\_\_ is the type of morphology that changes the word category and affects the meaning.

- a) Inflectional
- b) Derivational
- c) Cliticization
- d) None of the above



20) "He lifted the beetle with red cap." contain which type of ambiguity?

- a) Lexical ambiguity
- c) Referential ambiguity

- b) Syntax Level ambiguity
- d) None of the Above

Question 2 [12 marks]

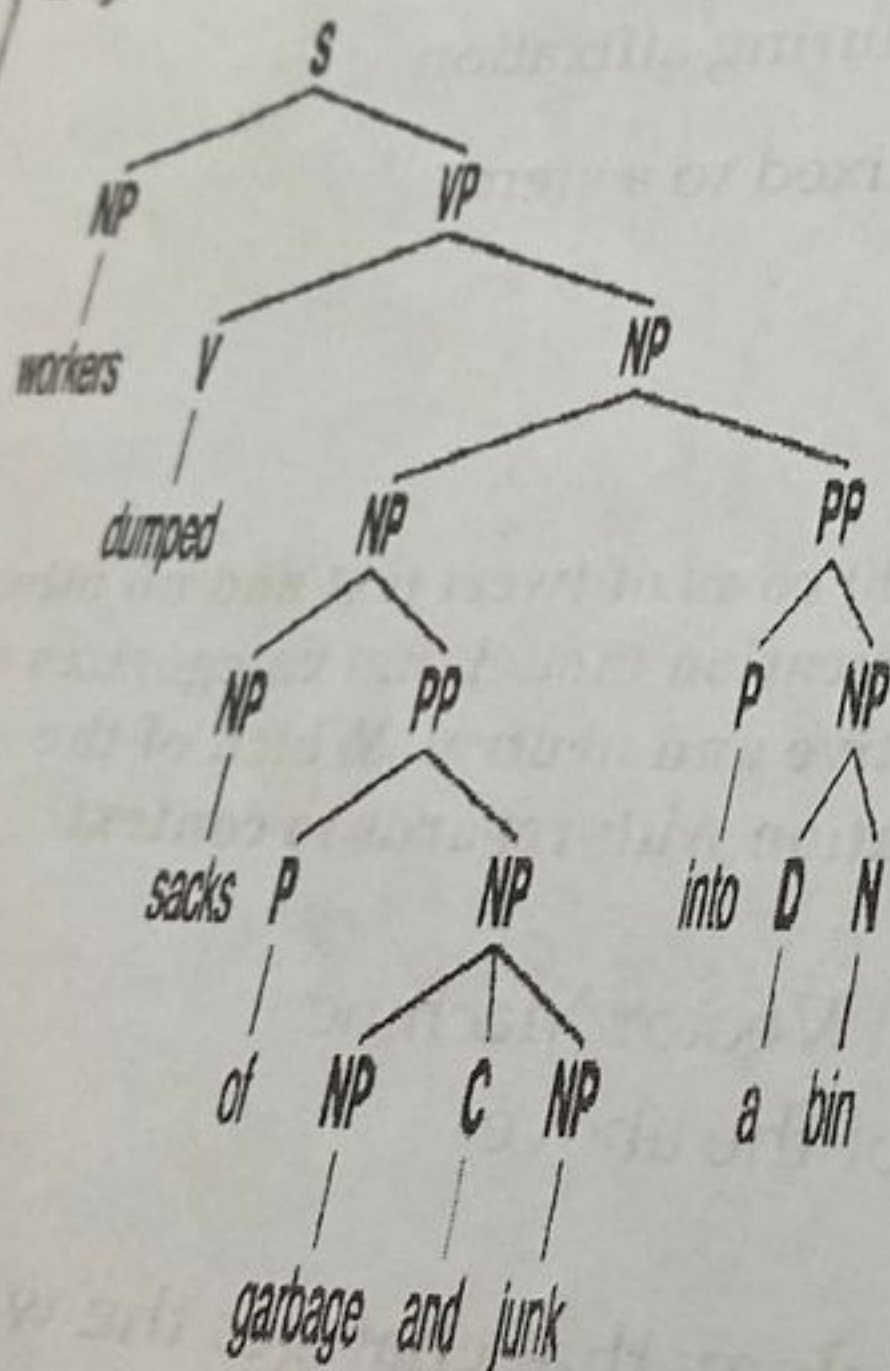
Assume we have the following context-free grammar G in Chomsky normal form:

$S \rightarrow NP VP$   
 $VP \rightarrow V NP$   
 $NP \rightarrow D N$   
 $PP \rightarrow P NP$   
 $VP \rightarrow VP PP$   
 $NP \rightarrow NP PP$   
 $NP \rightarrow NP CNP$

$CNP \rightarrow C NP$   
 $NP \rightarrow \text{"workers"} \mid \text{"sacks"} \mid \text{"garbage"} \mid \text{"junk"}$   
 $N \rightarrow \text{"bin"} \mid \text{"sack"}$   
 $V \rightarrow \text{"dumped"}$   
 $P \rightarrow \text{"of"} \mid \text{"into"}$   
 $D \rightarrow \text{"a"} \mid \text{"the"}$   
 $C \rightarrow \text{"and"}$

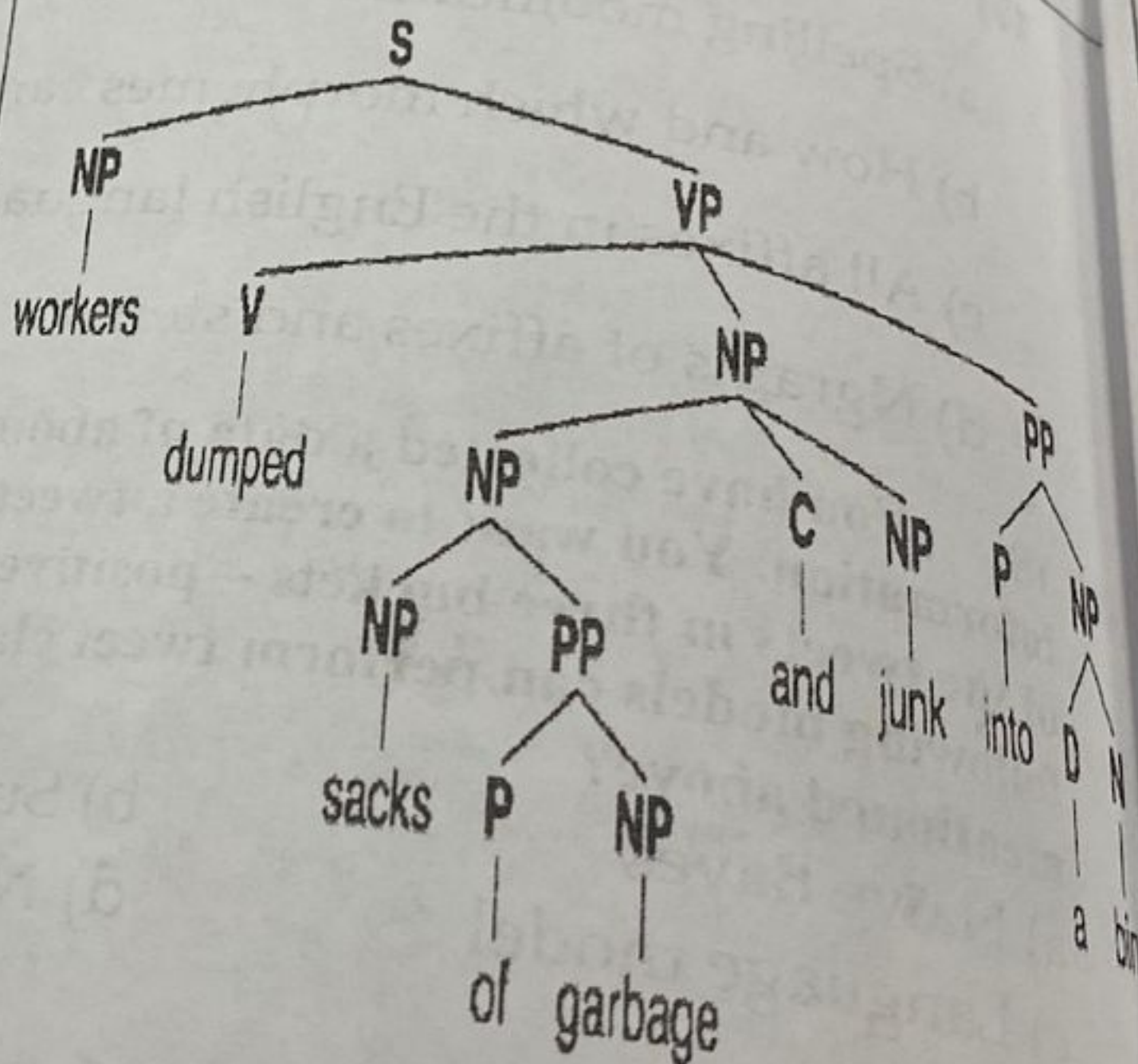
Which of the following trees for the sentence "workers dumped sacks of garbage and junk into a bin" are correct according to G?

21)



- A) True
- B) False

22)



- A) True
- B) False



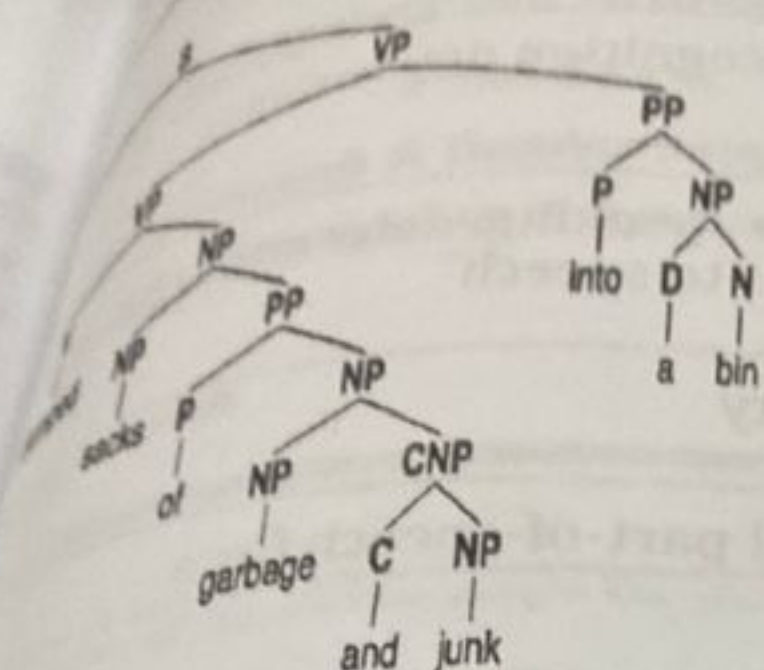
which type of

guity

normal form:

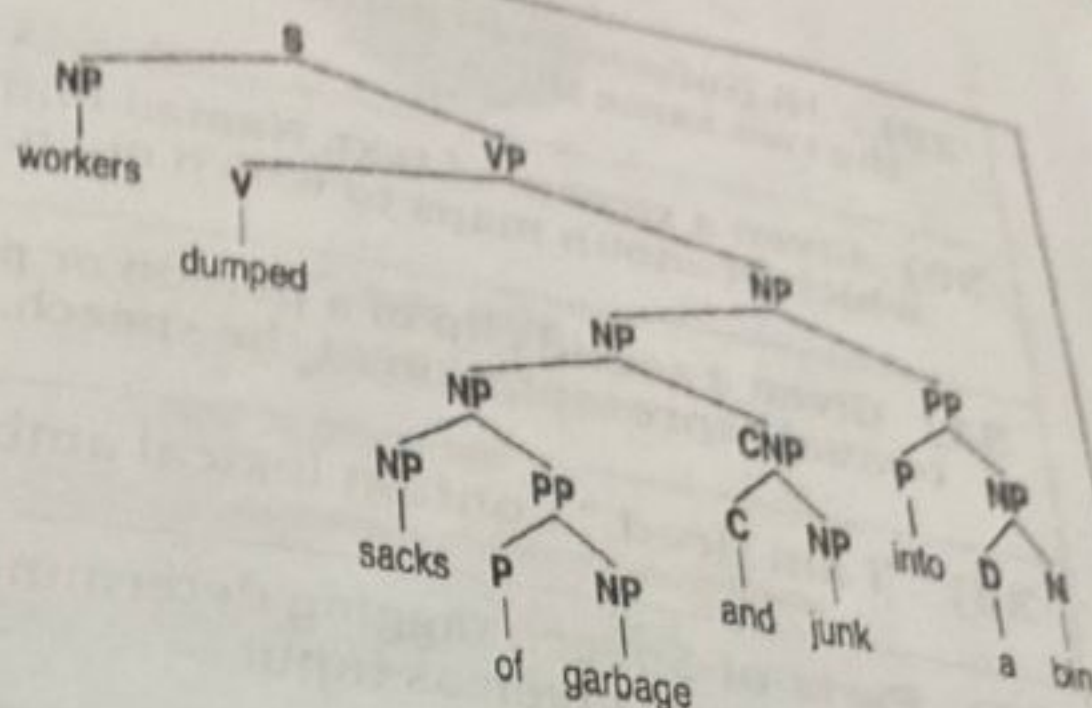
cks"  
nk"

bage



B) False

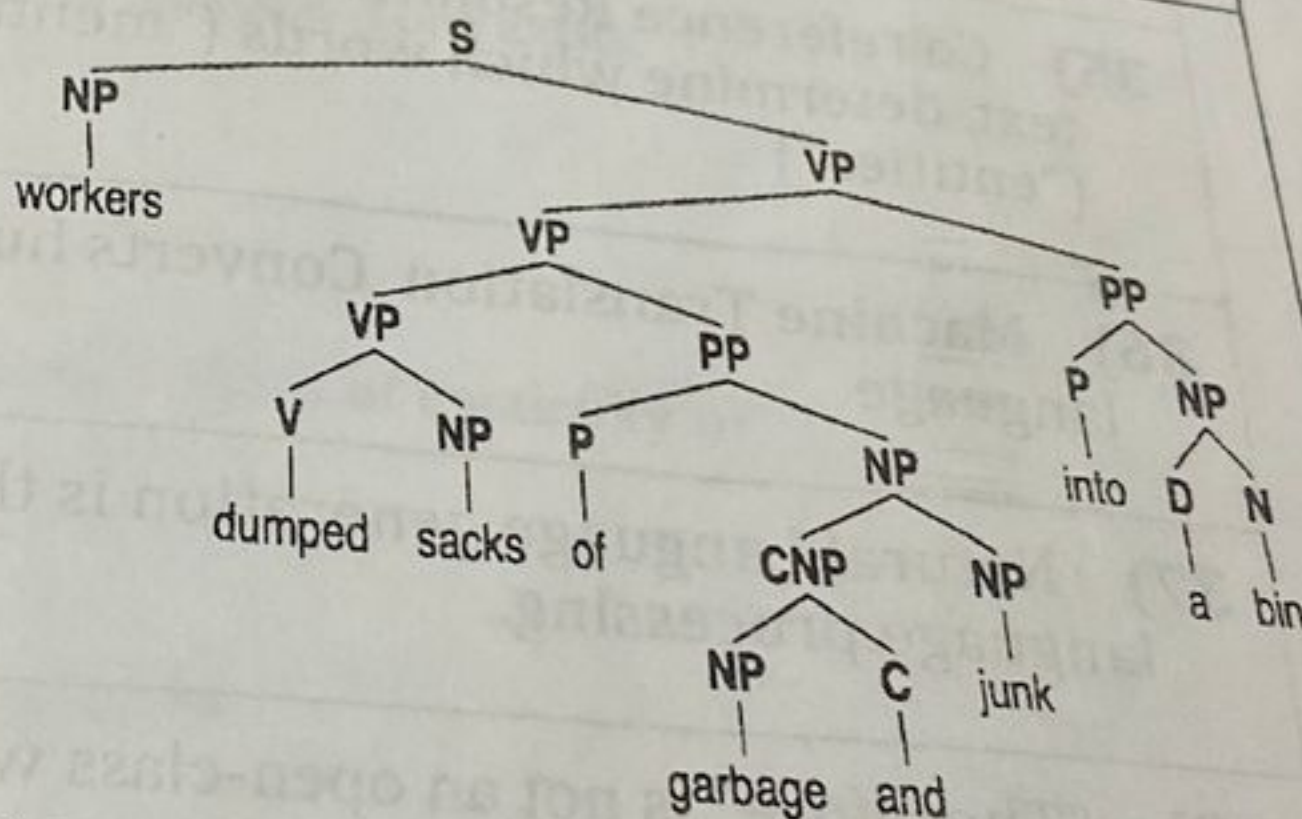
24)



A) True

B) False

26)



A) True

B) False

**Question 3** State whether right or wrong (A for right and B for wrong in bubble sheet). [28 marks]

Stemming can be used for the purpose of keyword normalization, the process of converting a keyword into its base form.

( )

grams are better than trigrams for part-of-speech tagging.

( )



|  |     |
|--|-----|
| 29) IR (Information Retrieval) and IE (Information Extraction) are the two same thing  | ( ) |
| 30) Given a stream of text, Named Entity Recognition determines which pronoun maps to which noun.  | ( ) |
| 31) Given a sound clip of a person or people speaking, determine the textual representation of the speech. "Text to speech"                        | ( ) |
| 32) "I am tired." Contain lexical ambiguity  | ( ) |
| 33) Parts-of-Speech tagging determines all part-of-speech for a specific word given as input   | ( ) |
| 34) Parsing determines Parse Trees (Grammatical Analysis) for a given sentence.  | ( ) |
| 35) Co reference Resolution is Given a sentence or larger chunk of text, determine which words ("mentions") refer to the same objects ("entities") | ( ) |
| 36) Machine Translation Converts human language to machine language  | ( ) |
| 37) Natural Language generation is the main task of Natural language processing.   | ( ) |
| 38) "Therefore" is not an open-class word in English.  | ( ) |
| 39) Polysemy means different senses of same word (unrelated meaning)   | ( ) |
| 40) An NLP system that identifies news articles about politics would be an information extraction system.  | ( ) |

True/False questions [2 marks each]  
 Heuristic solution procedures

Sensitivity analysis means

Linear programming model the variables.

A non-binding constraint

If two extreme points connecting the two

"Range of optimality" objective function

Each decision variable

The difference of constraints



سؤال 1 (40 علامة)

Question 1 (40 marks)

1) Which of the following strings cannot be derived from the symbol S using the rules  $S \rightarrow SS \mid aaa \mid aaaaa$  ?

- (a) aaaaa
- (b) aaaaaa
- (c) aaaaaaa
- (d) aaaaaaaaa
- (e) aaaaaaaaaa

2) Which of the following pairs of rules involves indirect recursion?

- (a)  $NP \rightarrow Det Adj N$  ,  $NP \rightarrow Pron$
- (b)  $VP \rightarrow V NP$  ,  $NP \rightarrow Det N that VP$
- (c)  $N' \rightarrow N N'$  ,  $NP \rightarrow Det N'$
- (d)  $S \rightarrow AdvP S$  ,  $AdvP \rightarrow Adj Adv$
- (e)  $VP \rightarrow V NP$  ,  $NP \rightarrow Det N PP$

3) Which of the following strings is a member of the language over  $\{a, b\}$  defined by the regular expression  $(aa + ba)^* (bb)^*$ ?

- (a) aabbba
- (b) aaaabb
- (c) babbaa
- (d) bbaa
- (e) None of the above.

4) Consider the following context-free grammar, with start symbol S and terminals  $a, ;, <, >$ .

$S \rightarrow <L \mid a \quad L \rightarrow aR \mid <LR \quad R \rightarrow > \mid ;L$