GÖTEBORGS UNIVERSITET
FILOSOFI, LINGVISTIK OCH VETENSKAPSTEORI

(partial)
ANSWER SHEET

Exam: Natural Language Processing (LT2103)

January 24, 2011, 09.00-12.00
Room: V30

Write your name and personal ID number (”personnummer”) clearly below. At every page, write only the exam number. This to make the exam anonymous (the name will be removed before the exam correction).

Name: _____

Personal ID number:_____

Exam number: _____

Limits	Pass	20 points
	Pass with distiction	30 points
	Maximum	40 points

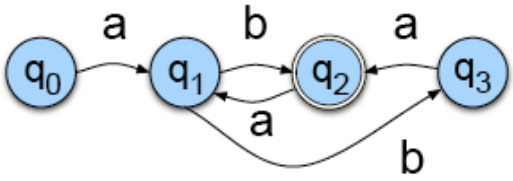
Accessories Paper and pencil
The Jurafsky & Martin book
But no hand-written notes!

Notes Write readable! (unreadable = wrong)
Keep it simple! (too complicated = wrong)
Number the pages, and start every question on a new page

Words (14p)

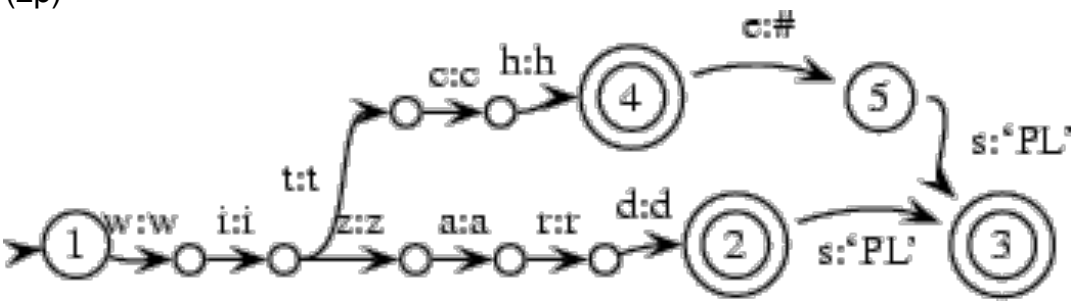
1.

(a) Write a regular expression for the language accepted by the following NFSA. (4p)



(aba?)+

(b) What output will the following Finite State Transducer produce for the input ”wizard”? (2p)



wizard

(b) What output will the FST in (b) produce for the input ”witches”? (2p)

witch#’PL’

2. Assume you have a bigram model of English, i.e. a set of probabilities of the form

P(*word1* | *word2*)

where *word1* and *word2* are words of English.

How would you use these to compute (an approximation of)

P(<s> good luck </s>)

You may assume that sentence start (<s>) and end (</s>) are included in the bigram model. (6p)

P(<s> good luck </s>) = P(good | <s>) * P(luck | good) * P(</s> | luck)

Syntax (20p)

3.

(a) Complete the context-free grammar below so that it generates (at least) the sentences listed. (3p)

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.

VP -> IV
PP -> P NP
P -> from

(b) Assign parse trees to all three sentences. (3p)
(c) Convert the grammar into Chomsky Normal Form. (6p)

add:	**(alternative:)**
VP -> DV X1	**VP -> X1 PP**
X1 -> NP PP	**X1 -> DV NP**
NP -> Tom	
NP -> Jerry	
VP -> sleeps	

(d) Given the CNF version of your grammar, draw the parse tree of ”Tom steals the cheese from Jerry”. (2p)
(d) following the CKY algorithm formally described in fig 13.10 (and informally described in Section 13.4.1), complete the CKY parse table for ”Tom steals the cheese from Jerry”. (6p)

Note on algorithm notation: in the algorithm, the notation {A | C(A)} is used, where C(A) is some condition on A. This means "the set of all A's such that C(A) holds". Note also that in a loop beginning with "for n <- from x downto y", if x is lower than y, nothing happens.

Tom	**steals**	**the**	**cheese**	**from**	**Jerry**
NP					S
	TV				VP
		Det	NP		X1
			N		
				P	PP
					NP

Semantics (4p)

4. Assume the following bag-of-words vector for "watch":

[likes, movies, time, escape, football, wrist, prison, night]

What is the correct binary vector representation for the following sentences (using a window size of 10)?

(i) Mary also likes to watch football games.

[1,0,0,0,1,0,0,0]

(ii) John's new watch shows the time in five locations.

[0,0,1,0,0,0,0,0]

(iii) The investigation clearly shows that no-one has escaped the prison during my watch!

[0,0,0,1,0,0,1,0]

(iv) Duels usually have one scene where the actors go out of frame and you watch their shadows fighting, at least in cliche movies.

[0,1,0,0,0,0,0,0]

(4p)

Pragmatics (2p)

5. Give an example of a sentence which could be interpreted both as a asking a question (e.g. "How old are you?") and requesting an (non-verbal) action ("Close the window please!"). (2p)

Can you pass the salt?