UNIVERSITY OF EDINBURGH

COLLEGE OF SCIENCE AND ENGINEERING

SCHOOL OF INFORMATICS

INFR09028 FOUNDATIONS OF NATURAL LANGUAGE PROCESSING

Tuesday 14th May 2013

09:30 to 11:30

INSTRUCTIONS TO CANDIDATES

Answer all of Part A and TWO questions from Part B.

Part A is COMPULSORY.

The short answer questions in Part A are each worth 3 marks, 24 marks in total. Each of the three questions in part B is worth 13 marks — answer any TWO of these.

Use one script book for each question, that is, three books in all.

CALCULATORS MAY NOT BE USED IN THIS EXAMINATION

Year 3 Courses

Convener: K. Kalorkoti
External Examiners: K. Eder, T. Field

THIS EXAMINATION WILL BE MARKED ANONYMOUSLY

Part A

Answer ALL questions in Part A.

Your answers should be thorough—as opposed to being as brief as possible—but they need not be lengthy: from one or two sentences up to a paragraph. When two terms are contrasted, make sure your short definitions of each make clear where the contrast lies. Each question is worth three marks, 24 marks in total for this section.

1. What are some reasons for the recent dominance of **data intensive approaches** to speech and language processing?

2. Name and define two types of **syntactic ambiguity**, give a linguistic example of each, and briefly explain why they present a major challenge for **parsing**.

3. Assuming a **bigram language model** with $\langle s \rangle$ and $\langle /s \rangle$ respectively marking the beginning and end of a sentence, give an expression for the joint probability of the sentence "I watched that TV show" made up of simpler probabilities.

4. What is **mutual information**? Give two example applications.

5. Assuming a **POS tagger** as a Hidden Markov Model (HMM), give the probability that the phrase "I watched that TV show" is tagged PRO VBD DET NN NN in terms of probabilities from the transition model and sensor model of the HMM.
   Note: assume that the start and end of a sentence is marked $\langle s \rangle$ and $\langle /s \rangle$ respectively.

6. Briefly describe the difference between lexical homonymy and lexical polysemy.

7. Assume that you are building a **word sense disambiguation** model that represents the context of each occurrence of the target word as a vector, consisting of the lemmas of the previous two open class words and the lemmas of the next two open class words, in that order. Then if the target word is *independence*, what would the vector be for representing its context in the sentence "Parsers differ as to which independence assumptions are made"?

8. What is a **frequency distribution**? Draw a graph showing the most common frequency distribution of natural language items. What is its name? Give two examples of items that have this distribution.

$t_{i-1} \backslash t_i$	NN	VB	$\langle /s \rangle$
NN	15.6	3.1	10.0
VB	7.6	20.0	8.0
$\langle s \rangle$	8.0	15.0	∞

Table 1: Transition Model

$t \backslash w$	*bears*	*walk*
NN	2.0	4.0
VB	3.0	1.0

Table 2: Sensor Model

Part B

Answer two questions in Part B.

1. **Dynamic Programming and Hidden Markov Models**

   (a) What independence assumptions are made by a Hidden Markov Model POS tagger? [*1 mark*]

   (b) Assuming a set T of POS tags, a tag $\langle s \rangle$ for the start of a sentence and a tag $\langle /s \rangle$ for the end of a sentence, derive the Hidden Markov Model (HMM) formula for identifying the most likely tag sequence t_1^n of a word sequence w_1^n. Derive the formula using probabilities, and justify each step in the derivation. Then state the derived formula using *costs*—that is, negative log probabilities. [*4 marks*]

   (c) Assuming that the POS tagger is defined by the transition and sensor models in Tables 1 and 2, give the POS tag sequence that the tagger assigns to the sentence *Bears walk*. Show all your workings by writing the costs of each possible tag assignment in a dynamic programming lattice, and include the backtraces in each cell to show which POS tag sequence is ultimately assigned to the string by the POS tagger.
   Note: All numbers are costs—that is, negative log probabilities. So you can (a) sum them rather than multiply them, and (b) you are looking to *minimise* the cost. [*8 marks*]

R1: S $\longrightarrow$ NP VP
R2: NP $\longrightarrow$ *love* | *matters* | ...
R3: VP $\longrightarrow$ V0
R4: VP $\longrightarrow$ V1 NP
R5: V0 $\longrightarrow$ *matters* | ...
R6: V1 $\longrightarrow$ *love* | ...

Figure 1: A simple context free grammar

2. **Chart Parsing**

(a) Describe in detail bottom-up depth-first (i.e., LIFO) chart parsing using context-free phrase structure grammars.

Include descriptions of the chart and its constituent parts, the agenda, the grammar, the input and the processing rules which create edges: that is, the lexical rule, the bottom up rule, and the fundamental rule.

Be sure to describe the overall process of parsing: how it starts, how the agenda and chart interact, how it finishes. [*5 marks*]

(b) Figure 1 shows a simple context-free grammar. Figure 2 shows the edges that were put onto the agenda (in the order they were put onto it) part-way through the bottom up depth-first parse of "love matters" using that grammar. Figure 3 shows the chart that exists part-way through the bottom-up parse of "love matters".

Note: We have not shown the order in which the edges *came off* the agenda and were put onto the chart! Rather, the agenda entries are numbered to show the order in which the edges were added to the agenda. The rules which created the edge are also shown: e.g., `FR(n,m)` means that the current edge was added to the agenda as a result of the Fundamental Rule operating on edges `m` and `n`; similarly `BU(m,n)` stands for the Bottom Up rule and `L` for the lexical rule.

Complete the parse. Write down *all* the subsequent edges that are added to the agenda (you don't need to copy down the edges 1–9 above), in the order in which they are added. Show this by continuing to number the edges on the agenda, starting with `11`. You should also identify for each edge the processing rule and inputs that create it, as illustrated above (e.g., `FR(7,8)`). Then copy the above chart, and complete it by showing how the edges that you have added to the agenda get added to the chart. You need only do one drawing, showing the final chart. [*8 marks*]

1L:	matters [1,2]
2L:	love [0,1]
3 BU(2,R2):	NP ⟶ • love [0,0]
4 BU(2,R6):	V1 ⟶ • love [0,0]
5 FR(2,4):	V1 ⟶ love • [0,1]
6 BU(5,R3):	VP ⟶ • V1 NP [0,0]
7 FR(5,6):	VP ⟶ V1 • NP [0,1]
8 FR(2,3):	NP ⟶ love • [0,1]
9 BU(8,R1):	S ⟶ • NP VP [0,0]
10 FR(2,9):	S ⟶ NP • VP [0,1]

Figure 2: Edges that have been put onto the agenda part-way through the parse of "love matters", using the grammar from Figure 1.

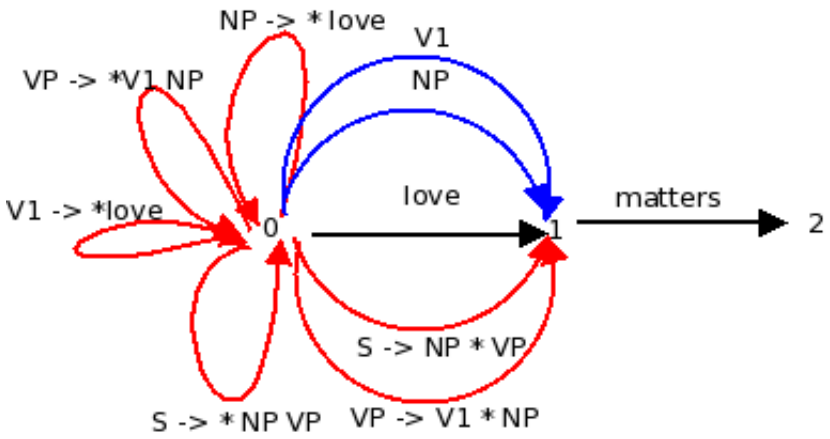

Figure 3: The chart, part-way through parsing "love matters" using the grammer from Figure 1.

3. **Pronoun Resolution**

(a) In Centering Theory, define the Forward-looking center, the preferred center, and the Backward-looking Center. [*3 marks*]

(b) Define the four intersential relationships among forward-looking centers, backward looking centers and preferred centers that are used in Centering theory to resolve pronouns; state their preference order. [*2 marks*]

(c) Use Centering Theory to derive the antecedents to the pronouns *his* and *He* in the sentences (ii) and (iii) respectively in the discourse below:

i. Bob opened up a new car dealership.
ii. John took a look at the Fords in his lot.
iii. He ended up buying one.

You must show the derivation step by step, by defining the forward-looking centers, backward-looking centers and preferred centers for each sentence and using the four intersential relationships among them. [*6 marks*]

(d) What sort of information, over and above the linguistic information that's used in Centering Theory, do you think one needs to make the right prediction for the discourse in part (c)? Justify your answer. [*2 marks*]