The University of Melbourne

School of Engineering

UNIB20005 Language and Computation

November 2011

Identical examination papers: None

Exam duration: Two hours

Reading time: Fifteen minutes

Length: This paper has 3 pages including this cover page.

Authorised materials: None

Calculators: Not permitted.

Instructions to invigilators: Students should be supplied with the exam paper and a script book, and with additional script books on request.

Instructions to students: This paper counts for 50% of your final mark. Please answer all questions in the script book provided. Please write your student id below and on your script book. Please begin each question on a new page of your script book. When you are finished, place the exam paper inside the front cover of the script book.

Examiners: Steven Bird, Lesley Stirling, Greg Restall

Library: This paper is not to be held in the Baillieu Library.

Student id:

Examiner's use only:	Part A:	Part B:	Part C:	Part D:	Total:

Part A: Key Concepts [10 marks]

1. Give short definitions of the following terms as they are used in the context of language and computation, with the help of examples:

   (a) part-of-speech tagger

   (b) context free grammar

   (c) valency

   (d) gesture

   (e) adjacency pair

   [5 marks]

2. Explain the distinctions between the following concepts, with the help of examples:

   (a) hypernym *vs* meronym

   (b) language change *vs* language variation

   (c) syntax *vs* semantics

   (d) individual constant *vs* individual variable

   (e) understanding an isolated sentence *vs* understanding dialogue

   [5 marks]

Part B: Language Analysis [15 marks]

3. Explain the phenomenon of syntactic agreement with the help of at least 4 example sentences. Provide a feature-based grammar which accounts for the grammaticality of these examples. Show the expected parse tree for one of the sentences.

   [5 marks]

4. Describe the class of strings matched by the following regular expressions. For each expression, give three matching strings.

   (a) `[A-Z][a-z]*`

   (b) `p[aeiou]?[aeiou]?t`

   (c) `\d+(\.\d+)?`

   (d) `([^aeiou][aeiou][^aeiou])+`

   (e) `\w+[^\w\s]+`

   [5 marks]

5. Translate the following sentence into first-order logic: *Every bear sees Cyril*. Using the lambda calculus, give a semantic representation for each word of the sentence, and specify the type of each expression. Show how these representations can be combined, step by step, to produce the same translation that you gave above.

   [5 marks]

Part C: Python Programming [15 marks]

6. Define a function `longest_sent(raw)` that takes a string of characters `raw` and returns the sentence having the greatest number of words. Assume that fullstop ("."), question mark, and exclamation mark are the only punctuation characters, and that they always indicate sentence boundaries. Also assume that whitespace reliably marks word boundaries. The function should return its result in the form of a string.
[5 marks]

7. Write a function `collocations(text)` which takes a sequence of words as its argument, and reports all collocations. Collocations are defined as pairs of words that are found adjacent to one another more often than one would expect based on word frequency alone. The function should return a list of pairs of words.
[5 marks]

8. A "conditional frequency distribution" is defined and accessed using the methods listed below:

Example	Description
`cfdist = ConditionalFreqDist(pairs)`	create a conditional frequency distribution from a list of pairs
`cfdist.conditions()`	alphabetically sorted list of conditions
`cfdist[condition]`	the frequency distribution for this condition
`cfdist[condition].samples()`	the observations made for this condition
`cfdist[condition][sample]`	frequency for the given sample for this condition

Write code to define a conditional frequency distribution where the conditions are defined to be the genres of the Brown Corpus, and the observations are the days of the week. Write code to format the frequency data as a table, where columns correspond to genres, and rows correspond to days of the week.

```
from nltk.corpus import brown
genres = brown.genres()
days = ['Sunday', 'Monday', 'Tuesday', 'Wednesday', 'Thursday', 'Friday', 'Saturday']
...
```

[5 marks]

Part D: Essay Question [10 marks]

9. Discuss *one* of the following topics (approx 1 page of text plus any diagrams). Marks will be given for correctness, completeness, and clarity.

   (a) Turing's 1950 article opens by considering the question "Can machines think?", but quickly turns to a different question. What is that question, and why does he turn to it? How would we go about answering Turing's new question? Discuss the prospects of constructing a computer program that understands language.

   (b) An embodied conversational agent is able to participate in a spoken dialogue with a human, and to perform and recognize simple gestures. Describe two of the components of a spoken dialogue system, and the kinds of knowledge that they require. Discuss at least two challenges that natural dialogue creates for the developers of dialogue systems and explain why these are difficult. Describe two types of gesture and explain how they convey meaning in the context of dialogue. Discuss the prospects of constructing a conversational agent that might pass the Turing Test.

— END OF EXAM —