

---

UNIVERSITI SAINS MALAYSIA

Second Semester Examination  
Academic Session 2006/2007

April 2007

**HET 324 - COMPUTATIONAL LINGUISTICS**

Duration : 3 hours

---

Please check that this examination paper consists of FOUR pages of printed material before you begin the examination.

There are SIX questions. Answer any FOUR questions.

1. Do you agree that syntax and semantics are equally important in language generation and analysis or do you agree that one aspect is more important than the other? Why? Discuss and support your answer with clear examples.

[25 marks]

2. One of the essential characteristics of corpus-based analysis is that it is empirical as it allows the analysis of actual patterns of use in natural text. What does this mean? Discuss your answer in relation to how this contributes to CALL, in particular how it could facilitate vocabulary teaching and learning.

[25 marks]

3. Discuss how NLP can enhance machine translation. Support your answer with relevant examples.

[25 marks]

4. In order for a computer program to analyse a sentence correctly, what knowledge does it need to have? What specific knowledge would you give a computer program for it to know that sentences such as *a*, *b*, *c* and *e* are acceptable and analysable whilst sentences *d* and *f* are not. Base your discussion on the sentences given and support your answer with further examples.

- [a] *I gave some of the money to charity.*
- [b] *I gave the stick to the dog.*
- [c] *I donated some of the money to charity.*
- [d] *\*I donated the stick to the dog.*
- [e] *Jane is pursuing her career in music.*
- [f] *\*Jane is chasing her career in music.*

[25 marks]

5. What is speech synthesis? What are some of the problems that still remain in the area? Discuss at least three of the problems in speech synthesis and their possible solutions.

[25 marks]

6. Explain the use of concordance lines and then answer the following questions based on the **Appended Data**. What are the different meanings of the word **bay** that can be found in the data? Set-up a lexicon for at least three different entries of the word **bay** based on four categories that a computational lexicon should have for analysing and generating natural language.

[25 marks]

**Appended Data**