

UNIVERSITI SAINS MALAYSIA

Second Semester Examination  
Academic Session 2004/2005

March 2005

**HET 324 - Computational Linguistics**

Duration : 3 hours

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

Answer FOUR questions altogether. All questions carry equal marks.

1. One of the main problem areas in machine translation is word-for-word mapping between languages. According to Barnbrook (1996), the translation process is not a simple matter of substituting individual words in the source language with equivalents in the target language. Changes have to be made in the translation of a text. What are the changes that may have to be made on a text? Discuss this in translating the following English sentence into Malay or your own L1.

She has a beautiful chest that she is very proud of and will only part with it if she gets a very good offer for it.

[25 marks]

2. According to Klavans (1996), one of the approaches to computational morphology is stemming or stripping algorithm. How does this approach work? Discuss this in relation to the procedures and rules that a morphological analyser should have in analysing the words given below. Include also in your discussion, the main problem in computational morphology.

- [a] suffixation
- [b] bunion
- [c] steadily
- [d] happily
- [e] convocation

[25 marks]

3. EITHER

- [a] What specific linguistic knowledge would a computer need to have for it to understand that sentence (ii) is unacceptable and that sentence (iii) has at least two possible structures?

[i] He donated the food to the poor.

[ii] \*He donated the food to the dog.

[iii] She took her pet to a small animal hospital.

Discuss using the sentences given and support your answer with further examples

OR

- [b] Between understander systems, data retrieval and expert systems, which application do you think would benefit humans the most if it had full NLP capability? Discuss.

[25 marks]

4. EITHER

- [a] Biber et al (1998) claim that 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 do they mean by this? Discuss your answer in the relation to how this contributes towards the area of language teaching, in particular how it could facilitate vocabulary teaching and learning.

OR

- [b] Frequency lists, concordance lines and collocations are techniques in corpus-based analysis that have become essential features in lexicography. Discuss how each feature facilitates the work of the lexicographer. Support your discussion with relevant examples.

[25 marks]