

UNIVERSITI SAINS MALAYSIA

Peperiksaan Semester Kedua  
Sidang Akademik 2002/2003

Februari/Mac 2003

**HET 324 – Computational Linguistics**

Masa : 3 jam

**THIS EXAMINATION PAPER CONTAINS FOUR [4] QUESTIONS IN THREE [3] PAGES.**

There are **FOUR [4]** questions. Answer **ALL** the questions.

1. According to Klavans (1997), a computer program would need to know more than just the part of speech to analyse or generate a sentence correctly. What specific knowledge would a computer program need to have for it to know that sentences *a*, *d*, *e*, *f*, and *g* are acceptable and that sentences *b*, *c* and *h* are not acceptable. Discuss using the sentences given and support your answer with further examples.

- [a] I decided to go.
- [b] \* I decided him to go.
- [c] \* I persuaded to go.
- [d] I persuaded him to go
- [e] The policeman is chasing the thief.
- [f] The policeman is pursuing the thief.
- [g] She is pursuing her studies at the college.
- [h] \* She is chasing her studies at the college.

[20 marks]

2. Choose any **TWO** of the following areas and provide a clear description of each one. Include in your description what they refer to, their applications, strengths and shortcomings. Wherever necessary, give examples to further clarify and support your answer.

- [a] Corpus –based approach
- [b] Speech synthesis
- [c] Computational morphology
- [d] CALL

[25 marks]

3. What is Natural Language Processing (NLP)? According to Barnbrook, the extent to which applications of computational linguistics such as machine translation, lexicography, word processing, information retrieval, speech recognition and others make use of NLP varies greatly. Choose **TWO** applications and discuss how the provision of NLP capability could improve their existing ability and benefit users.

[25 marks]

4. 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.

[30 marks]