UNIVERSITI SAINS MALAYSIA

Peperiksaan Semester Kedua
Sidang Akademik 2003/2004

Februari/Mac 2004

HET 324 - Computational Linguistics

Time : 3 hours

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

Answer **FOUR** questions only.

You have a choice between questions 1 and 2 and between questions 3 and 4. Questions 5 and 6 must be answered.

EITHER

1. Language generation is the reverse of language analysis. Do you agree with this statement? Why or why not?

OR

2. Why does synthetic speech still sound synthetic and why would a successful model of speech articulation not be a model of speaking? Discuss these issues with reference to the problems that still remain in speech synthesis and speech recognition applications.

[20 marks]

EITHER

3. What is Machine Translation? Give a clear account of what it refers to, the different types of machine translation that are available, the difficult problems that surround the application and solutions to the problems. Wherever necessary, give examples to clarify and support your answer.

OR

4. The use of corpora in language studies is important in various areas of language. Yet, the areas of research in which corpus-based approach has made the most impact are semantics, critical discourse analysis, lexicography and language teaching. In what way(s) has corpus-based research impacted these areas? Discuss this with reference to two of the areas listed.

[25 marks]

5. What specific knowledge would a computer program need to have for it to know that sentences *a, d, e* and *g* are acceptable and that sentences *b, c, g* 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] She got on the horse and rode into the sunset

[f] *She rode into the sunset and got on the horse

[g] The dog barked at the cat.

[h] * The cat barked at the dog.

[25 marks]

6. In her discussion on computational lexicon, Klavans (in O'Grady et al, 1996) suggests that a computational lexicon should contain as much information as possible to correctly analyse and generate language. She suggests a list of nine different categories of information that a lexical entry should have. What are the categories? Then using the concordance for the word *declined* that is appended, analyse the meanings of the word, and provide three entries for *declined* using six of the categories suggested by Klavans.

...Appended/-
...4/-

Appended

- ooo 0 ooo -