UNIVERSITY OF LONDON
IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE

EXAMINATIONS 1997

MEng Honours Degrees in Computing Part IV
MEng Honours Degree in Information Systems Engineering Part IV
MSc Degree in Advanced Computing
for Internal Students of the Imperial College of Science, Technology and Medicine

This paper is also taken for the relevant examinations for the
Diploma of Membership of Imperial College
Associateship of the City and Guilds of London Institute

PAPER 4.85 / I4.24

NATURAL LANGUAGE PROCESSING

Friday, May 9th 1997, 10.00 - 12.00

Answer THREE questions

For admin. only: paper contains 4 questions

1a Explain the phenomenon of unbounded movement using the example:

the sheep that John dreamed of in the field was asleep

What is the underlying structure of this sentence? How can this phenomenon be represented in a grammar? Explain carefully how this may be dealt with using DCGs.

b How is quantification handled in a general manner when parsing a sentence into an internal logical form?

Show how the logical form may be generated for the sentence:

Every goalkeeper saved a penalty

showing the parse tree annotated with all necessary variable bindings.

There is an ambiguity in this sentence. State what it is and explain briefly what method could be used to treat it in the same framework.

2a Why is most information in a modern grammar placed in the lexicon rather than in the grammatical rules?

Indicate what type of information in required in lexical entries for:

nouns
pronouns
verbs

and show some sample entries for these.

b Many words in the lexicon have a number of meanings and these frequently have different parts of speech. Explain how the ambiguities can be reduced or eliminated by statistical means. What are the advantages and disadvantages of this approach compared with simply applying the constraints implicit in the parsing process?

Turn over ...

3a Explain the difference between the terms syntax, semantics and pragmatics when applied to natural language. Demonstrate how each of them would be used in processing a query:

Can you tell me how to get to Edinburgh in the morning?

b Explain the terms anaphora and ellipsis as used in linguistics.

Annotate the dialogue which consists of example sentence in part a and the following sentences to show the anaphors and ellipses it contains:

The train will get you there by 12.15. It leaves from Euston at 8.10.
Earlier?
Then you need to take a plane. There is one at 8.40 which gets to Edinburgh at 9.45 and you will get to the centre 45 minutes later.

Explain in general terms how one might detect and handle these.

The two parts carry, respectively, 30% and 70% of the marks.

4a What is the difference between between the power of the grammatical formalism needed to represent the morphological and syntactic structures of a language?

How is this reflected in the formalisms that are normally used for these aspects? Give examples of constructs which distinguish the two.

b Explain how the two-level morphology model works in deriving the following spellings:

boxes bigger holier

You may make reasonable assumptions and generalisations to generate appropriate rules.

c Why have named features become universal in modern grammar rules? What are the main uses of such features and how are they treated in the grammar. Give examples of any one notation for representing these.

The three parts carry, respectively, 30%, 40%, 30% of the marks.

End of paper