UNIVERSITY OF LONDON

IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE

EXAMINATIONS 1996

MEng Honours Degrees in Computing Part IV
MSc Degree in Foundations of Advanced Information Technology
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

NATURAL LANGUAGE PROCESSING

Thursday, May 2nd 1996, 10.00 - 12.00

Answer THREE questions

For admin. only: paper contains
4 questions
1 pages (excluding cover page)

1a What are the components of a definite clause grammar (DCG) and in which ways do they differ from a context-free grammar? Give examples.

b Show, with examples how each constituent of a DCG maps into Prolog.

c How can DCGs be used to express the features of the following sentence and thus eliminate incorrect variants:

He catches them in big nets.

2a For each of the following classes of words, show in detail the types of lexical entries that are necessary to capture both the surface and logical correspondance:

Nouns, Verbs, Determiners, Prepositions

Show how these entries are used by a parsing system in converting the surface to logical form.

b Explain how the logical form is used in evaluating queries against a database and give the top level of a meta-level interpeter. Show how problems of quantification are resolved.

3a Define the terms "well-formed substring" and "dotted production" as used within a chart parser. Show how they are used to form the chart and how this can avoid reevaluating the parse tree.

b Draw a representation of the chart for the following sentence showing the passive part and some indication of the active part:

he eats butter and margarine is safer

You should supply suitable rules and show a sufficient part of any partial parses that might reasonably be formed to demonstrate possible ambiguity.

c Explain informally how a chart parser operates in *either* a top-down *or* a bottom-up fashion to produce the above chart.

4a Define the different levels at which language can be analysed. What power of formalism is appropriate to each of these levels? Suggest the type of programming languages appropriate for each of these and why. Which is the most difficult to handle?

b Describe one language processing technique which does *not* follow this structure. List the advantages and disadvantages of this approach compared with the fuller linguistic approach. Suggest two applications of language processing where it might be more suitable.

c How can statistics be used in natural language processing? Suggest two uses for a large corpus of processed text.

End of paper