

MPhil in Advanced Computer Science

Deep Language Modelling

Leader:	Ann Copestake
Timing:	Lent
Prerequisites:	Basic understanding of syntax and semantics for language processing. e.g., the Introduction to Language Processing module or Chapters 12, 13, 17 and 18 in Jurafsky and Martin.
Structure:	16 Lectures/seminars (2 sessions per week)

AIMS

This module aims to introduce students to research which relates to ‘deep’ language processing (roughly speaking, techniques utilising explicit semantic representation). The emphasis will be on broad coverage approaches using real data, but we will primarily consider models which are linguistically and psycholinguistically motivated rather than application requirements. A range of current research topics will be discussed, including some relatively under-explored areas. Interconnections between subareas will be emphasised.

SYLLABUS

Most of the sessions in the course will be seminars rather than lectures: students will be expected to do the assigned reading before the session and come prepared to discuss the material. In some sessions, students will present selected research papers.

1. Syntax-semantics interface
 - (a) Generalised quantifiers. Scope underspecification. (2 L)
 - (b) Constructions. Idioms and other multiword expressions. (3 L)
2. Language generation.
 - (a) Introduction to generation. (1 L)
 - (b) Realisation with deep grammars. (2 L)
 - (c) Content selection. (1 L)
 - (d) Lexical selection. (1 L)
3. Psycholinguistics and computational linguistics.
 - (a) Concepts and lexical semantics. (3 L)
 - (b) Quantifier interpretation. (1 L)
 - (c) Incremental parsing and generation. (2 L)

All lectures/seminars will be taught by Ann Copestake.

OBJECTIVES

On completion of this module students should:

- Have a general understanding of deep language processing techniques and an in-depth understanding of selected topics.
- Understand the factors underlying the choice of semantic representations in a variety of contexts.
- Be able to explore the research literature on their own (including relevant work in other disciplines) and to summarise research papers on a topic.

COURSEWORK

The students will be expected to complete the required reading before each seminar (15 hours). They will prepare one 20-30 minute presentation on a research paper which will involve more in-depth reading on a particular topic (10 hours). They will complete a 3,000 word essay on the same topic as their presentation (15 hours). The choice of topic will be discussed with the student by the module leader. Each student will work on a different topic.

PRACTICAL WORK

None.

ASSESSMENT

It will be mandatory for students to give a presentation, but the presentation will not be marked. The module will be assessed by a 3,000 word essay, on a topic assigned by the module leader. The essay will be marked by the module leader and a percentage score given. The essay will be due two weeks after the end of the module (subject to timetabling).

RECOMMENDED READING

Jurafsky and Martin, *Speech and Language Processing*, second edition (2008) will be used for general background. Other readings will be announced before the start of the module.