

MPhil in Advanced Computer Science

Lexical Semantics and Discourse Processing

Leader:	Simone Teufel
Timing:	either term
Prerequisites:	Core Module Introduction to NLP
Structure:	16 Lectures

AIMS

This module aims to provide an introduction to NLP research centered around lexical semantics (i.e., aspects of the meaning of single words, or the relation between the meaning of words) and discourse processing (linguistic phenomena which operate beyond a single sentence). The phenomena are described, algorithms are presented to recognise and/or resolve ambiguities with respect to these phenomena. The module will also present applications of these two aspects of NLP, including summarisation.

SYLLABUS

Some sessions will be lectures, others will be mostly lectures, but have a short task-based break-out session. There will be assigned reading before each lecture. All lectures will be taught by Simone Teufel.

1. Introduction to Lexical Semantics; Psycholinguistic Models of Similarity and Relatedness (1L)
2. Word Senses and WSD (3L)
3. Sentiment, Register and Genre (1L)
4. Semantic Spaces (2L)
5. Verb Classes (selectional restrictions; thematic roles) (1L)
6. Figurative Language (1L)
7. Models of Coherence (1L)
8. Anaphora Resolution (2L)
9. Rhetorical Relations (3L)
10. Applications (1L)

OBJECTIVES

On completion of this module students should:

- understand aspects of word-based meaning such as synonymy, similarity, word senses, and aspects of discourse phenomena such as anaphora, co-reference, and rhetorical relations;
- have an understanding of corpus-based and symbolic approaches for representing various aspects of the meaning of words and for solving lexical semantics-based ambiguities;
- understand what a model of discourse structure is, which linguistic aspects it could be based on, and which algorithms could be used to recognise aspects of discourse structure;
- have a deep enough understanding of the methods to be able to re-implement these from descriptions in the research literature;
- be able to perform a literature review in the relevant topics, to identify and understand published solutions to subproblems
- appreciate why these phenomena play an important role in practical NLP applications, and be able to identify in which respect the phenomena interact with the task.

COURSEWORK

- Students are expected to do several non-assessed exercises, each covering 4-5 lectures. Exercises set by Simone Teufel (will not change substantially between years); these are materials for self-study and they may include some corpus-based work/programming. Solution notes will be made available when exercises are due; students mark their own answers. (15 hours)

PRACTICAL WORK

None.

ASSESSMENT

- One final take-home exam with two essay-style questions; may include a literature review. Questions set and marked by Simone Teufel. Due 48 hours after being made available, marked 1 week after submission. All marks will be in percentage.

RECOMMENDED READING

- Jurafsky and Martin, Speech and Language Processing, 2nd edition (2008). Chapters 19-21.
- Additional papers, which will be announced before the start of the module.

Last updated: September 2008