

Term paper proposal —
Machine learning for modelling inter-sentential
linguistic phenomenon in text

Fredrik Olsson

November 14, 2003

Proposed task

This is a proposal for a term paper for the course in machine learning taught at the Graduate School of Language Technology during the winter semester of 2003.

Machine learning techniques and methods have been successfully applied to a number of language technology tasks. Previous work has to a large extent been focused on intra-sentential phenomenon, such as part of speech tagging, phrase chunking and named entity recognition to name but a few.

The purpose of this term paper is to investigate the efforts made in applying machine learning methods for handling *inter*-sentential linguistic phenomenon, as well as to see if the things modelled have counterparts in prevailing discourse theories.

Goal

An optimal outcome of this term paper is an exhaustive survey of the literature in the field as well as an implementation and/or test of some of the methods covered by the survey. A more realistic outcome of the term paper, given the time available, is a literature survey covering prominent approaches to machine learning for discourse processing.

Constraints

To constrain the task to a presumably manageable size, the subject matter under investigation will be narrative text from single document.