

# Finding Discourse Relations in Student Essays

Rohana Mahmud and Allan Ramsay

School of Informatics, University of Manchester

R.Mahmud@postgrad.umist.ac.uk, Allan.Ramsay@manchester.ac.uk

**Abstract.** The aim of the work reported here is to provide a tool to help secondary school (high school) age students to reflect on the structure of their essays. Numerous tools are available to help students check their spelling and grammar. Very little, however, has been done to help them with higher level problems in their texts. In order to do this, we need to be able to analyse the discourse relations within their texts. This is particularly problematic for texts of this kind, since they contain few instances of explicit discourse markers such as *‘however’*, *‘moreover’*, *‘therefore’*. The situation is made worse by the fact that many texts produced by such students contain large numbers of spelling and grammatical errors, thus making linguistic analysis extremely challenging. The current paper reports on a number of experiments in classification of the discourse relations in such essays. The work explores the use of machine learning techniques to identify such relations in unseen essays, using a corpus of manually annotated essays as a training set.

## 1 Introduction

Secondary school students have numerous problems when trying to compose extended texts. They have low-level errors with spelling and grammar, for which a range of support tools exists. But they also have problems with organising their texts into coherent well-structured discourses, and they have particular problems using the devices that the language places at their disposal for indicating the structure to the reader (e.g. lexical cohesion relations and careful construction of referential chains). The underlying aim of the work reported here, then, is to provide a tool which will reveal the discourse cues that are present in such essays, and hence to allow students to reflect on what they have written.

There is, clearly, no such thing as the ‘right’ way to structure an essay. There are correct ways to spell things (though no extant spell-checker gets them all right), and there are correct and incorrect grammatical forms (though no extant parser can be relied on to pass all grammatically correct constructions and flag all grammatically incorrect ones). So it is, at least in principle, possible to produce a tool which will tell you whether you have spelt all the words in some document correctly, and whether all your sentences are grammatically acceptable. But there is no right or wrong structure for an extended text, so it simply makes no sense to talk of showing students where they have made ‘errors’ in the organisation of their texts. The best we can hope for is to show them where they have used discourse structuring markers, and to show them the structure that the clues they have used impose on their texts. The hope is that by