

Possible Attributes For Identifying Discourse Relations For A Reflective Learning Tool

By

Rohana Mahmud

Department of Computation

University of Manchester Institute of Science & Technology (UMIST)

(r.mahmud@postgrad.umist.ac.uk)

Keywords: Discourse Relations, Writing Process, A Reflective Learning Tool

Abstract

A coherent essay is a discourse that has meaning and is made up of sequences of related sentences. It has an internal structure and usually it is characterized by discourse relations (rhetorical relations), the relations that reflect semantic and functional judgments about the text spans they connect. Writing a discourse, an extended sequence of sentences, is a very complex cognitive process especially to beginners in writing like school children. We propose a reflective learning tool that can automatically reveal discourse relations structure to help students structure their essays. Various possible linguistics attributes will be investigated and integrated to automatically identify the discourse relations structure.

1.0 Introduction

As identified by many researchers in discourse research coherent texts have an internal structure and usually they are characterized by discourse relations (rhetorical relations), the relations that reflect semantic and functional judgments about the text spans they connect (Mann and Thompson 1988; Moser and Moore 1993; Grosz, Joshi et al. 1995; Marcu 2000; Knott, Sanders et al. 2001; Ramsay 2003) . They also agree that the relations are not just marked within sentences (inter-sentential) itself but also between sentences (extra-sentential). Many applications such as machine learning, text generation systems, automatic essay evaluation systems, learning how to write and information retrieval require the automatic process of finding discourse relations.

2.0 Discourse structure and student's essays

We want to develop a Reflective Learning Tool to help school children learn how to compose a coherent essay. A coherent essay is a discourse that has meaning and is made up of sequences of related sentences. Writing a discourse, an extended sequence of sentences, is a very complex cognitive process especially to beginners in writing like school children. It is not a simple matter of transcribing language into written symbols. It requires the writer to be creative and responsible for the text that they produce (Sharples, Goodlet et al. 1989; White and Arndt 1991). Indeed, certain cognitive psychologists have described writing as the most complex and demanding of all cognitive activities undertaken by human beings, because there are no rules and no unique solutions (Bracewell, 1981). Usually when we write an extended text, typically we will start with the generation of unstructured content. We will then organise the content in some way and finally translate it into linear text. These processes may occur iteratively and follow a definite generation-organisation-translation progression. (Hayes and Flower 1980; (Sharples and Pemberton 1992)). However, students often have difficulty in structuring their sentences and end up with incoherent essays. Instead of consulting their teachers to explain how to reorganize their sentences, we hope students could use the tool which will automatically reveal the implicit discourse relations structure. The proposed tool should not just focus on improving the quality of the end-product as found in many existing computer aided writing tools (CAW), instead it must focus on the process of writing (Williams 1992; Marjorie 1994; Crafton 1996; Takayoshi 1996).

However, as agreed by (Marcu 2000), automatically deriving the discourse structure of text is a difficult problem. Marcu used obvious surface clues such as connective markers and topic keyword, in producing his discourse parser. We think that richer linguistic clues or attributes may do even better. We therefore plan to carry out more substantial linguistic analysis than Marcu in order to find appropriate clues. This linguistic analysis will have to be carried out on the training set and on the essays that students want

information about. This will pose a number of problems, since deep linguistic analysis of this kind usually depends on the input texts being grammatically well-formed (Torrance and Bouayad-Agha 2001; Burstein, Marcu et al. 2003), which is unlikely to be the case for the kinds of essay we will be looking at. Most approaches in discourse relations also assume discourse has a *tree* structure (Mann and Thompson 1988; Grosz, Joshi et al. 1995; Marcu 2000; Knott, Sanders et al. 2001). However, we know this is **not** always the case, since our students may write very incoherent essays, and may not even be desirable. Thus, this makes the discourse parsing strategy more complex when we know that discourse may be discontinuous. We need rules to assign confidence levels to potential relations between sentences which may be non-adjacent, and a parsing strategy that chooses the optimal set of relations. This is likely to make both the acquisition of rules and their application more difficult than they would be if we assumed that discourses are tree-like.

3.0 A Reflective learning tool

The proposed system is called Structured Tool for Writing Process (STfWP) and has three major modules, namely the PARASITE system module, a Discourse Relation engine (DRe) module and a Graphical User Interface Editor module (GUI-Editor). The following figure 1.0 shows the conceptual model of the system:

3.1 The engine and attributes

The PARASITE system (an acronym for PrAgmatics=ReAsoning about Speaker's InTention), is an existing linguistic engine (Ramsay 2001). It will parse each sentence and carry out an initial linguistic analysis which detects theme-rheme relationships, notes and classifies centering transitions, and records the forms of referring expressions. We expect to use WordNet (Miller, Beckwith et al. 1993; Pallotta 2001), a free lexical online dictionary, to give us some information about the relations between the head verbs of

potentially related sentences. The resulting information should provide us with a richer set of attributes for making guesses about discourse relations. These range from Rhetorical Structure Theory (RST) (Mann and Thompson 1987), Centering Theory (CT) (Grosz, Joshi et al. 1995), Theme and Rheme (Halliday and Hasan 1976), Lexical Relations clues, Topic and Focus and Referring Expression devices such as Anaphors and Pronouns (Halliday and Hasan 1976; Ramsay 2003).

The various attributes will be investigated and integrated in the DRe module using selected machine learning algorithm. DRe will analyse and reveal the discourse relations structure through the GUI-Editor module. GUI-Editor module is a simple-to-use editor, provided with friendly menus and buttons, which will guide users on how to use the system. The structure will be presented in a graphical format and some examples and further explanations are also provided. Students are hope to be able to do ‘re-vision’ (ie. to see text beyond its surface structure) by themselves. The system will not evaluate whether the essay is good or bad but instead help students improve their writing process by reflecting on the structure of what they have written and hence gaining an insight into the cues that can be used to impose a structure on an extended piece of writing.

4.0 Methodology

The aim of this work is to provide a reflective learning tool. We therefore need to collect appropriate training data from school pupils’ work, and we need to liaise with teachers in order to find out how such a tool might fit into classroom activities. We are currently in discussion with the IT development officer at the local education authority to see how we can collect appropriate data without compromising student confidentiality, and to see how he thinks such a tool might fit into classroom practice. An essay mining tool has been developed in order to annotate the data (students’ essays). A mock-up system is also prepared so that the students and teachers could easily specify their requirements.

The domain of our system will be a monologue discourse and specific genre. It will help students in the process of writing that is by checking the relations between pair of

sentences. The targeted users will be English native speakers (students) with sound grammar and spelling. They are also assumed already familiar with using computer in a classroom.

Conclusion

The paper discusses the overview of our current research. It only describes the conceptual model of our proposed reflective learning tool and mentioned the possible attributes to automatically identify the implicit discourse relations structure. Further work will be carried out in implementing the Discourse Relation engine (DRe) and manipulating those attributes after the training data essays is ready. The tool will then be tested and evaluated by the targeted users.

References:

- Miller A, G., R. Beckwith, et al. (1993). Introduction to WordNet: An On-line Lexical Database, University of Princeton: 85.
- Burstein, J., D. Marcu, et al. (2003). "Finding the WRITE Stuff : Automatic Identification of Discourse Structure in Student Essays." IEEE Intelligent Systems **18**(January-February): 32-39.
- Crafton, R. E. (1996). "Promises, Promises: Computer-Assisted Revision and Basic Writers." Computers and Composition **13**: 317-326.
- Grosz, B. J., A. K. Joshi, et al. (1995). "Centering: A Framework for Modelling the Local Coherence of Discourse." Computational Linguistics **21**(2): 203-225.
- Halliday, M. A. K. and R. Hasan (1976). Cohesion in English. London, Longman.
- Knott, A., T. Sanders, et al. (2001). "Levels of representation in discourse relations." Cognitive Linguistics **12**(3): 197-209.
- Mann, W. C. and S. A. Thompson (1987). "Rhetorical Structure Theory: A Framework for the Analysis of Texts."
- Mann, W. C. and s. A. Thompson (1988). Rhetorical Structure Theory: Toward a functional theory of text organization. Text. **8**: 243-281.

Marcu, D. (2000). The Theory and Practice of Discourse Parsing and Summarization, The MIT Press.

Marcu, D. (2000). The theory and practice of discourse parsing and summarization. Cambridge, MIT Press.

discourse parser

Marjorie, S. (1994). Computer Assisted Writing instruction. Eric Digest.

Moser, M. and J. D. Moore (1993). Investigating Discourse Relations. ACL Workshop on Intentionality and Structure in Discourse Relations.

Pallotta, V. (2001). Introduction to WordNet. **2003**.

Ramsay, A. (2001). PrAgmatics = ReAsoning about the Speaker's InTensions - The Parasite Manual. Manchester, Computation Department, UMIST: 48.

Ramsay, A. (2003). Discourse. The Oxford Handbook of Computational Linguistics. R. Mitkov. New York, Oxford University Press: 112-135.

Sharples, M., J. Goodlet, et al. (1989). Developing writer's Assistant. Computers and Writing: Models and Tools. N. Williams and P. Holt. Oxford, England, BlackwellScientific Publications Ltd.

writer assistant

Sharples, M. and L. Pemberton (1992). Representing Writing: External Representations and the Writing Process. Computers and Writing: State of the Art. P. O. B. Holt and N. Williams. Oxford, England, Kluwer Academic Publishers: 319-336.

Takayoshi, P. (1996). "The shape of Electronic Writing: Evaluating and Assessing Computer-Assisted Writing Processes and Products." Computers and Composition **13**: 245-257.

Torrance, M. and N. Bouayad-Agha (2001). Rhetorical structure analysis as a method for understanding writing processes. International workshop on Multidisciplinary Approaches to Discourse, Ittre, Belgium.

White, R. and V. Arndt (1991). Process Writing. Essex, Addison Wesley Longman Limited.

Williams, N. (1992). New Technology, New Writing, New Problems? Computers and Writing: State of the Art. P. O. B. Holt and N. Williams. Oxford, England, Kluwer Academic: 1-19.

