Discourse Relations
and
Discourse Markers

Proceedings of the Workshop

Sponsored by
The Association for Computational Linguistics

Edited by
Manfred Stede
Leo Wanner
Eduard Hovy

15th August 1998
Université de Montréal
Montreal, Quebec, Canada

COLING-ACL’98
Preface

There is widespread agreement that the coherence of discourse can be characterized by certain discourse relations that can hold between adjacent spans of text. Different linguistic disciplines approach the definition and interpretation of such discourse relations in different ways, though. Text linguists, for instance, favour a “top down” approach in that they investigate the global role of relations holding between text spans (sometimes called ‘rhetorical relations’), as an instrument of creating coherence. Grammarians, on the other hand, favour a “bottom up” approach: they intend to describe the grammatical role, e.g., of conjunctions, and to that end propose inventories of semantic relations signalled by these conjunctions. As a result of these complementary approaches, the level of description on which discourse relations hold and the particular inventories of discourse relations are somewhat heterogeneous.

One way of approaching the abstract problem of discourse relations on a relatively safe ground is to focus on the surface, investigating the linguistic means by which discourse relations can be recognized. The most important instruments are cue words specifically employed to signal the particular relationship between text spans. These are often called discourse markers—though the term is quite ambiguous. Besides their function to make discourse relations explicit in all forms of discourse, in spoken language they may also provide specific meta-information about the utterance or the interpersonal relationship between dialogue participants. In a wider sense, this function can be interpreted as providing coherence within a dialogue, and therefore the different functions of these markers are quite related.

With the increasing capabilities of natural language processing beyond the sentence level, the problem of discourse structure interpretation, generation, and translation is becoming more acute. It is therefore not surprising that the number of researchers working in this field has increased rapidly over the past few years—a fact that is also reflected in the number of submissions to this workshop.

The workshop addresses the topic of discourse coherence from different angles. The papers presented can be divided into five broad categories: (i) automatic identification of discourse structure relations in discourse (discourse structure parsing); (ii) recognition and generation of cue words that function as discourse markers; (iii) interrelationship between discourse structure relations and the other linguistic strata, grammar and semantics; (iv) analysis and formal representation of both specific discourse structure relations and specific cue words; (v) cue words functioning as discourse markers in speech and dialogue.

We hope that the workshop will provide a forum for stimulating discussions, and that it will generate progress on the issues of discourse coherence and its linguistic signals. Thanks are due to the authors of the papers, who supplied a very good basis for shaping a balanced program. Also, we wish to thank the members of the Program Committee for their valuable help.

Manfred Stede
Leo Wanner
Eduard Hovy

PROGRAM COMMITTEE:
Sandra Carberry (University of Delaware)  Barbara Di Eugenio (University of Pittsburgh)
Eduard Hovy (ISI, USC)  Alistair Knott (University of Edinburgh)
Owen Rambow (CoGenTex Inc.)  Ted Sanders (University of Utrecht)
Donia Scott (University of Brighton)  Wilbert Spooren (University of Tilburg)
Manfred Stede (Technical University of Berlin)  Keith Vander Linden (Calvin College)
Marilyn Walker (AT&T Labs)  Leo Wanner (University of Stuttgart)
Program

9.00 Opening

**Session 1: Discourse Structure Parsing Uses**
Introduction

9:10-9:30 Daniel Marcu (USC/ISI) *A Surface-based Approach to Identifying Discourse Markers and Elementary Textual Units in Unrestricted Texts*

9:30-9:50 Simon H. Corston-Oliver (Microsoft Research) *Identifying the Linguistic Correlates of Rhetorical Relations*

9:50-10:10 Jill Burstein, Karen Kukich, Susanne Wolff, Chi Lu and Martin Chodorow (Educational Testing Service and Hunter College) *Enriching Automated Essay Scoring using Discourse Marking*

10:10-10:25 Discussion

Coffee Break

**Session 2: Cue Words**

Introduction

10:45 Brigitte Grote (Otto-von-Guericke Universität, Magdeburg) *Representing Temporal Discourse Markers for Generation Purposes*

10:50-11:10 Liesbeth Degand (University of Louvain) *On Classifying Connectives and Coherence Relations*

11:10-11:30 Claudia Soria and Giacomo Ferrari (University of Pisa and University of East Piemonte) *Lexical Marking of Discourse Relations — Some Experimental Findings*

11:30-11:50 Simone Teufel (University of Edinburgh) *Meta-Discourse Markers and Problem-Structuring in Scientific Texts*

11:50-12:10 Discussion

Lunch

13:00-14:00 **Poster session**

Laurence Danlos (Université Paris 7) *Linguistic Ways for Expressing a Discourse Relation in a Lexicalized Text Generation System*

Alistair Knott (University of Edinburgh) *Similarity and Contrast Relations and Inductive Rules*

Frank Schilder (Universität Hamburg) *Temporal Discourse Markers and the Flow of Events*

Nigel Ward (University of Tokyo) *Some Exotic Discourse Markers of Spoken Dialogue*

**Session 3: Grammar, Semantics, and Formalisms**

14:00 Introduction

14:05-14:25 Kathleen Dahlgren (Inquizit Technologies, Inc.) *Lexical Marking and the Recovery of Discourse Structure*

14:25-14:45 Jacques Jayez and Corinne Rossari (EHESS and Université de Genève) *Discourse Relations versus Discourse Marker Relations*

14:45-15:05 Marie-Paule Péry-Woodley (Université de Toulouse) *Textual Signalling in Written Text: a Corpus-based Approach*

15:05-15:25 Bonnie Lynn Webber and Aravind Joshi (University of Pennsylvania) *Anchoring a Lexicalized Tree-Adjoining Grammar for Discourse*

15:25-15:40 Discussion

Coffee Break
Session 4: Speech and Dialogue

16:05 Introduction


16:50–17:10 Kerstin Fischer and Hans Brandt-Pook (Universität Bielefeld) Automatic Disambiguation of Discourse Particles

17:10–17:30 Daniel Jurafsky, Elizabeth Shriberg, Barbara Fox and Traci Curl (University of Colorado and SRI) Lexical, Prosodic, and Syntactic Cues for Dialog Acts

17:30–18:00 Discussion
18:00 Closing
# Table of Contents

## Preface .......................................................... i

## Program ........................................................... ii

## Table of Contents ..................................................... iv

## Author Index .......................................................... vi

## Workshop Papers

### Daniel Marcu

**A Surface-based Approach to Identifying Discourse Markers and Elementary Textual Units in Unrestricted Texts** .......................................................... 1

### Simon H. Corston-Oliver

**Identifying the Linguistic Correlates of Rhetorical Relations** ..................... 8

### Jill Burstein, Karen Kukich, Susanne Wolff, Chi Lu and Martin Chodorow

**Enriching Automated Essay Scoring using Discourse Marking** ................... 15

### Brigitte Grote

**Representing Temporal Discourse Markers for Generation Purposes** .......... 22

### Liesbeth Degand

**On Classifying Connectives and Coherence Relations** .......................... 29

### Claudia Soria and Giacomo Ferrari

**Lexical Marking of Discourse Relations — Some Experimental Findings** .... 36

### Simone Teufel

**Meta-Discourse Markers and Problem-Structuring in Scientific Texts** ......... 43

### Laurence Danlos

**Linguistic Ways for Expressing a Discourse Relation in a Lexicalized Text Generation System** .......................................................... 50

### Alistair Knott

**Similarity and Contrast Relations and Inductive Rules** ........................ 54

### Frank Schilder

**Temporal Discourse Markers and the Flow of Events** ........................... 58

### Nigel Ward

**Some Exotic Discourse Markers of Spoken Dialog** ................................ 62

### Kathleen Dahlgren

**Lexical Marking and the Recovery of Discourse Structure** ..................... 65

### Jacques Jayez and Corinne Rossari

**Discourse Relations versus Discourse Marker Relations** ........................ 72

### Marie-Paule Péry-Woodley

**Signalling in Written Text: a Corpus-based Approach** .......................... 79

### Bonnie Lynn Webber and Aravind Joshi

**Anchoring a Lexicalized Tree-Adjoining Grammar for Discourse** ......... 86

### Masahito Kawamori, Takeshi Kawabata and Akira Shimazu

**Discourse Markers in Spontaneous Dialogue: a Corpus Based Study of Japanese and English** .......................................................... 93

### Yukiko I. Nakano and Tsuneaki Kato

** Cue Phrase Selection in Instruction Dialogue using Machine Learning** .... 100

### Kerstin Fischer and Hans Brandt-Pook

**Automatic Disambiguation of Discourse Particles** ............................. 107
## Author Index

<table>
<thead>
<tr>
<th>Author</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brandt-Pook, H.</td>
<td>107</td>
</tr>
<tr>
<td>Burstein, J.</td>
<td>15</td>
</tr>
<tr>
<td>Chodorow, M.</td>
<td>15</td>
</tr>
<tr>
<td>Corston-Oliver, S.H.</td>
<td>8</td>
</tr>
<tr>
<td>Curl, T.</td>
<td>114</td>
</tr>
<tr>
<td>Dahlgren, K.</td>
<td>65</td>
</tr>
<tr>
<td>Danlos, L.</td>
<td>50</td>
</tr>
<tr>
<td>Degand, L.</td>
<td>29</td>
</tr>
<tr>
<td>Ferrari, G.</td>
<td>36</td>
</tr>
<tr>
<td>Fischer, K.</td>
<td>107</td>
</tr>
<tr>
<td>Fox, B.</td>
<td>114</td>
</tr>
<tr>
<td>Grote, B.</td>
<td>22</td>
</tr>
<tr>
<td>Jayez, J.</td>
<td>72</td>
</tr>
<tr>
<td>Joshi, A.</td>
<td>86</td>
</tr>
<tr>
<td>Jurafsky, D.</td>
<td>114</td>
</tr>
<tr>
<td>Kato, T.</td>
<td>100</td>
</tr>
<tr>
<td>Kawabata, T.</td>
<td>93</td>
</tr>
<tr>
<td>Kawamori, M.</td>
<td>93</td>
</tr>
<tr>
<td>Knott, A.</td>
<td>54</td>
</tr>
<tr>
<td>Kukich, K.</td>
<td>15</td>
</tr>
<tr>
<td>Lu, C.</td>
<td>15</td>
</tr>
<tr>
<td>Marcu, D.</td>
<td>1</td>
</tr>
<tr>
<td>Nakano, Y.I.</td>
<td>100</td>
</tr>
<tr>
<td>Péry-Woodley, M.-P.</td>
<td>79</td>
</tr>
<tr>
<td>Rossari, C.</td>
<td>72</td>
</tr>
<tr>
<td>Schilder, F.</td>
<td>58</td>
</tr>
<tr>
<td>Shimazu, A.</td>
<td>93</td>
</tr>
<tr>
<td>Shriberg, E.</td>
<td>114</td>
</tr>
<tr>
<td>Soria, C.</td>
<td>36</td>
</tr>
<tr>
<td>Teufel, S.</td>
<td>43</td>
</tr>
<tr>
<td>Ward, N.</td>
<td>62</td>
</tr>
<tr>
<td>Webber, B.L.</td>
<td>86</td>
</tr>
<tr>
<td>Wolff, S.</td>
<td>15</td>
</tr>
</tbody>
</table>