Towards Standards and Tools for Discourse Tagging

Proceedings of the Workshop

Edited by
Marilyn Walker

21 June, 1999
University of Maryland
College Park, Maryland, USA

Published by the Association for Computational Linguistics

Towards Standards and Tools for Discourse Tagging

Proceedings of the Workshop

Edited by
Marilyn Walker

21 June, 1999
University of Maryland
College Park, Maryland, USA

Published by the Association for Computational Linguistics

©1999, Association for Computational Linguistics

Order additional copies from:

Association for Computational Linguistics
75 Paterson Street, Suite 9
New Brunswick, NJ 08901 USA
+1-732-342-9100 phone
+1-732-342-9339 fax
acl@aclweb.org

ACL'99 Workshop
Towards Standards and Tools for Discourse Tagging

Discourse tagging assigns labels from a tag set to discourse units in texts or dialogues. The discourse units range from words or referring expressions to multi-utterance units identified by criteria such as speaker intention or initiative. The motivation for corpora of tagged discourse is the hope that such corpora will lead to major advances in the area of discourse processing similar to the advances in sentence-level language processing that followed the emergence of syntactically annotated corpora. This qill require widely available, large corpora, tagged for multiple phenomena. The goal of this workshop is to contribute to the development of and awareness of useful tools and standard tag sets for tagging discourse phenomena.

While researchers have always labelled and categorized discourse phenomena, there has recently been a large increase in work on discourse tagging. The Discourse Resource Initiative (DRI) was started at a workshop on Discourse Tagging held at the Institute for Research in Cognitive Science at the University of Pennsylvania in March of 1996. See http://www.cis.upenn.edu:80/ircs/discourse-tagging/multiparty.html. Since the first workshop, DRI participants have organized yearly international workshops on the standardization of discourse tagging schemes for dialogue acts, coreference, and higher level discourse structures (http://www.georgetown.edu/luperfoy/Discourse-Treebank/workshops.html). A related effort is the MATE project (http://mate.mip.ou.dk/), co-funded by the European Union, whose aim is to develop tools and standards for tagging spoken dialogue corpora at different levels, including the discourse level. There is also a related effort in Japan (http://www.slp.cs.ritsumei.ac.jp/dtag/).

Even with these three initiatives in place, there is still much work to be done before there are widely accepted (standardized) tagging schemes for various discourse phenomena that could be shared across sites. We hope that this workshop will contribute directly towards the goal of having public, shared corpora, tagged for discourse phenomena, that discourse researchers can use to advance the field of discourse processing. We hope to discuss the following issues at the workshop:

- How can standardization for discourse tagging be achieved? Is it possible to develop a set of standard coding schemes, one for each phenomenon of interest? Or will there necessarily be many schemes with relationships defined between them?
- Cross-level coding: all the initiatives mentioned above promote an approach in which coding schemes are developed at different levels, rather than an approach in which a monolithic scheme addresses all phenomena. Given this methodology, the issue of cross-level coding arises, namely, how can coding schemes for different levels take advantage of each other and

allow coding of cross-level relationships? Is it possible to use corpus annotations at different annotation levels to examine the interdependence of linguistic phenomena?

- Coding schemes and theories of discourse: How can corpora coded for discourse issues help advance our theoretical understanding of discourse phenomena? Is it possible to develop coding schemes that faithfully reflect a discourse theory? If yes, is it desirable?
- Coding schemes and applications: Is it possible to design discourse coding schemes independently from the applications that the tagged corpora are supposed to be used for (eg, to train a speech act recognizer)?
- Coding schemes and reliability: discourse categories are difficult to code for reliably. Whatever the reason (e.g., lack of an overarching theory for discourse, genuine ambiguity and misunderstandings in real dialogue reflected in the coding, etc), how can we devise reliable coding schemes? What reliability measures should be used: are widely used measures (Kappa, Alpha, precision and recall) and the corresponding standards appropriate for discourse tagging? If not, what other measures can we use? When is it appropriate to use naive vs. expert coders? How is reliability affected by whether naive or expert coders are used?
- Tools for discourse tagging: What specific features do discourse tagging tools require? How can we develop tools that decrease the cost and increase the reliability of tagging? Can we simply extend tools developed for other uses, e.g. for syntactic tagging?
- Some paradigms for evaluating dialogue systems take advantage of the use of tagged corpora: how are discourse tagging and tagging for evaluation purposes related? Are there some discourse tags that may be used as evaluation tags or is it advisable to introduce another dimension of tagging?

Thanks to the program committee for reviewing the submitted papers. We hope you enjoy the workshop !

Marilyn Walker (Organizer)	AT&T Labs Research, U.S.A
Morena Danieli (Organizer)	CSELT, Italy
Barbara Di Eugenio (Organizer)	University of Illinois at Chicago, U.S.A
Johanna Moore (Organizer)	University of Edinburgh, U.K.
Jean Carletta	HCRC, University of Edinburgh, U.K.
Laila Dybkjaer	MIP, Odense University, Denmark
Julia Hirschberg	AT&T Labs Research, U.S.A
Diane Litman	AT&T Labs Research, U.S.A
Masato Ishizaki	JAIST, Japan
David Novick	EURISCO, France
Daniel Jurafsky	University of Colorado, U.S.A

TABLE OF CONTENTS

- Steven Bird and Mark Liberman, *Annotation Graphs as a Framework for Multidimensional Linguistic Data Analysis*, p. 1
- Jean Carletta and Amy Isard, *The MATE Annotation Workbench: User Requirements*, p. 11
- Jean Francois Delannoy, *Argumentation Mark-Up : A Proposal*, p. 18
- A. Ichikawa, M. Araki, Y. Horiuchi, M. Ishizaki, S. Itabashi, T. Itoh, H. Kashioka, K. Kato, H. Kikuchi, H. Koiso, T. Kumagai, A. Kurematsu, K. Maekawa, S. Nakazato, M. Tamoto, S. Tutiya, Y. Yamashita, T. Yoshimura. *Evaluation of Annotation Schemes for Japanese Discourse* , p. 26
- Marion Klein, *Standardisation Efforts on the Level of Dialogue Act in the MATE Project*, p. 35
- Lori Levin, Klaus Ries, Ann Thume-Gobbel, and Alon Lavie. *Tagging of Speech Acts and Dialogue Games in Spanish Call Home*, p. 42
- Daniel Marcu, Estibaliz Amorrortu, and Magdalena Romera. *Experiments in Constructing a Corpus of Discourse Trees*, p. 48
- Jon David Patrick. *Tagging Psychotherapeutic Interviews for Linguistic Analysis*, p. 58
- M. Poesio, F. Bruneseaux, and L. Romary *The MATE meta-scheme for coreference in dialogues in multiple languages*, p. 65
- Claudia Soria and Vito Pirrelli, *A Recognition-Based Meta-Scheme for Dialogue Acts Annotation*, p. 75
- Simone Teufel and Mark Moens. *Discourse-Level Argumentation in Scientific Articles: Human and Automatic Annotation* , p. 84
- Graziella Tonfoni, *A markup language for tagging discourse and annotating documents in context sensitive interpretation environments*, p. 94
- David R. Traum and Christine H. Nakatani, *A Two-Level Approach to Coding Dialogue for Discourse Structure: Activities of the 1998 DRI Working Group on Higher-Level Structures*, p. 101
- Teresa Zollo and Mark Core. *Automatically Extracting Grounding Tags from BF Tags*, p. 109