Proceedings of the

Second Workshop on Text Meaning and Interpretation

Held in cooperation with ACL-2004

25–26 July 2004

Barcelona, Spain

Copyright © 2004 The Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
73 Landmark Center
East Stroudsburg, PA 18301
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

Second Workshop on Text Meaning and Interpretation 25–26 July 2004

Preface

This $1\frac{1}{2}$-day workshop will continue the success of the 2003 Workshop on Text Meaning, which was held at the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics in Edmonton. It aims to:

- re-establish the research community of knowledge-based interpretation of text meaning;
- explicate the implicit treatments of meaning in current knowledge-lean approaches and how they and knowledge-rich methods can work together; and
- emphasize the construction of systems that extract, represent, manipulate, and interpret the meaning of text (rather than theoretical and formal methods in semantics).

Most, if not all, high-end NLP applications—such as machine translation, question answering and text summarization—stand to benefit from being able to use text meaning in their processing. But the bulk of work in the field in recent years has not pertained to treatment of meaning. The main reason given is the complexity of the task of comprehensive meaning analysis and interpretation.

Computational linguistics has always been interested in meaning, of course. The tradition of formal semantics, logics, and common-sense reasoning system has been continuously maintained for many years. But also, much work has been devoted to building practical, increasingly broad-coverage meaning-oriented analysis and synthesis systems. Lexical semantics has made significant progress in theories, description, and processing. Formal aspects of ontology work have also been studied. The Semantic Web has further popularized the need for automatic extraction, representation, and manipulation of text meaning: for the Semantic Web to really succeed, capability of automatically marking text for content is essential, and this cannot be attained reliably using only knowledge-lean, semantics-poor methods.

While there has recently been a flurry of specialized meetings devoted to formal semantics, lexical semantics, semantic web, formal ontology and others, the number of meetings devoted to knowledge-based text meaning processing—content rather than formalism—has been much smaller. The first Workshop on Text Meaning began to remedy this, and ten papers were presented on implemented systems and on related topics.[1]

The call for papers of the present workshop suggested, without limitation, the following topics to potential contributors to the workshop:

- Implemented systems that extract, represent, or manipulate text meaning.
- Broad-coverage semantic analysis and interpretation.
- Knowledge-based text synthesis.
- The nature of text meaning required for various practical broad-coverage applications.
- Manual annotation of text meaning, including interlingual annotations.
- Pragmatics and discourse issues as parts of meaning extraction and manipulation.
- Ontologies supporting automatic processing of text meaning.
- Semantic lexicons.
- Microtheories to support text meaning extraction and manipulation: aspect, modality, reference, etc.
- Text meaning representations in semantic analysis.

[1] The proceedings of this workshop are available at *http://acl.ldc.upenn.edu/W/W03/#W03-0900*.

- Reasoning to support semantic analysis and synthesis.
- Multilingual aspects of meaning representation and manipulation.
- Integrating semantic analysis and non-semantic language processing.
- Semantic analysis and synthesis systems based on knowledge-lean stochastic corpus-oriented methods.

The call for papers encouraged discussion of theoretical issues that are relevant to computational applications, including descriptions of processors and static knowledge resources. It specifically preferred discussions of content and meaning over discussions of formalisms for encoding meaning, and discussions of decision heuristics in processing over discussions of generic processing architectures and theorem-proving mechanisms.

Twenty-seven papers were submitted to the workshop, of which fifteen were selected for presentation and are included in these proceedings. In addition, two panel sessions were organized—see descriptions below in this volume.

Sergei Nirenburg and Graeme Hirst
July 2004

Organizers

GRAEME HIRST, Department of Computer Science, University of Toronto. *gh@cs.toronto.edu*
SERGEI NIRENBURG, Department of Computer Science and Electrical Engineering, University of Maryland, Baltimore County. *sergei@umbc.edu*

Program Committee

JAN ALEXANDERSSON, Deutsche Forschungszentrum für Künstliche Intelligenz, Saarbrücken
COLLIN BAKER, International Computer Science Institute, Berkeley
PETER CLARK, Boeing
DICK CROUCH, Palo Alto Research Center
RICHARD KITTREDGE, University of Montreal
PAUL KINGSBURY, Penn
TANYA KORELSKY, CoGenTex, Inc.
CLAUDIA LEACOCK, Educational Testing Service
DAN MOLDOVAN, University of Texas at Dallas
ANTONIO MORENO ORTIZ, University of Málaga
MARTHA PALMER, University of Pennsylvania
GERALD PENN, University of Toronto
VICTOR RASKIN, Purdue University
ELLEN RILOFF, University of Utah
GRAEME RITCHIE, University of Edinburgh
MANFRED STEDE, University of Potsdam
KARIN VERSPOOR, Los Alamos National Labs
YORICK WILKS, University of Sheffield

Table of Contents

Interpretation in a cognitive architecture
Harold Paredes-Frigolett 1

Solving logic puzzles: From robust processing to precise semantics
Iddo Lev, Bill MacCartney, Christopher Manning and Roger Levy 9

Constructing text sense representations
Ronald Winnemöller 17

OntoSem and SIMPLE: Two multi-lingual world views
Marjorie McShane, Margalit Zabludowski, Sergei Nirenburg and Stephen Beale 25

Evaluating the performance of the OntoSem semantic analyzer
Sergei Nirenburg, Stephen Beale and Marjorie McShane 33

Question answering using ontological semantics
Stephen Beale, Benoit Lavoie, Marjorie McShane, Sergei Nirenburg and Tanya Korelsky 41

Making sense of Japanese relative clause constructions
Timothy Baldwin 49

Carsim: A system to visualize written road accident reports as animated 3D scenes
Richard Johansson, David Williams, Anders Berglund and Pierre Nugues 57

Inducing a semantic frame lexicon from WordNet data
Rebecca Green and Bonnie Dorr 65

Paraphrastic grammars
Claire Gardent, Marilisa Amoia and Evelyne Jacquey 73

Lexical-semantic interpretation of language input in mathematical dialogs
Magdalena Wolska, Ivana Kruijff-Korbayová and Helmut Horacek 81

Underspecification of 'meaning': The case of Russian imperfective aspect
Barbara Sonnenhauser 89

Text Understanding with GETARUNS for Q/A and Summarization
Rodolfo Delmonte 97

Semantic forensics: An application of ontological semantics to information assurance
Victor Raskin, Christian F. Hempelmann and Katrina E. Triezenberg 105

Interpreting communicative goals in constrained domains using generation and interactive negotiation
Aurélian Max 113

Technical Program Schedule

Sunday, 25 July 2004

08:30–08:40 Welcome

08:40–09:10 *Interpretation in a cognitive architecture*
Harold Paredes-Frigolett

09:10–09:40 *Solving logic puzzles: From robust processing to precise semantics*
Iddo Lev, Bill MacCartney, Christopher Manning and Roger Levy

09:40–10:00 *Constructing text sense representations*
Ronald Winnemöller

10:00–10:30 Break

10:30–12:00 Panel: *Toward a theory of semantic annotation*
David Farwell and Eduard Hovy (conveners)

12:00–13:30 Lunch

13:30–14:00 *OntoSem and SIMPLE: Two multi-lingual world views*
Marjorie McShane, Margalit Zabludowski, Sergei Nirenburg and Stephen Beale

14:00–14:30 *Evaluating the performance of the OntoSem semantic analyzer*
Sergei Nirenburg, Stephen Beale and Marjorie McShane

14:30–15:00 *Question answering using ontological semantics*
Stephen Beale, Benoit Lavoie, Marjorie McShane, Sergei Nirenburg and Tanya Korelsky

15:00–15:30 Break

15:30–15:50 *Making sense of Japanese relative clause constructions*
Timothy Baldwin

15:50–16:10 *Carsim: A system to visualize written road accident reports as animated 3D scenes*
Richard Johansson, David Williams, Anders Berglund and Pierre Nugues

16:10–17:40 Panel: *Can we move from sentence meaning to text meaning?*
Sergei Nirenburg (convener)

Monday, 26 July 2004

08:30–09:00 *Inducing a semantic frame lexicon from WordNet data*
Rebecca Green and Bonnie Dorr

09:00–09:30 *Paraphrastic grammars*
Claire Gardent, Marilisa Amoia and Evelyne Jacquey

09:30–10:00 *Lexical-semantic interpretation of language input in mathematical dialogs*
Magdalena Wolska, Ivana Kruijff-Korbayová and Helmut Horacek

10:00–10:30 Break

10:30–11:00 *Underspecification of 'meaning': The case of Russian imperfective aspect*
Barbara Sonnenhauser

11:00–11:20 *Text Understanding with GETARUNS for Q/A and Summarization*
Rodolfo Delmonte

11:20–11:40 *Semantic forensics: An application of ontological semantics to information assurance*
Victor Raskin, Christian F. Hempelmann and Katrina E. Triezenberg

11:40–12:00 *Interpreting communicative goals in constrained domains using generation and interactive negotiation*
Aurélian Max

Workshop Panels

Toward a Theory of Semantic Annotation

Conveners: **David Farwell**, New Mexico State University and Universidad Politecnica de Catalunya, and **Eduard Hovy**, Information Sciences Institute, University of Southern California

Panelists: **Manfred Pinkal**, Universität des Saarlandes, and **Martha Palmer**, University of Pennsylvania

Given the increasing number of annotated corpora being created, it is opportune to consider what one needs to do to ensure that the annotation effort succeeds. What, indeed, is "success" for an annotation effort? What desiderata should annotation efforts conform to in order to maximize chances of success? When compromises on the desiderata are required for practical reasons, which desiderata are first to go? What is the resulting impact on the effort?

We propose the following desiderata:

- perform annotations that are useful for a wide number of tasks (possibly ones not even foreseen today);
- focus on annotations not easily done automatically without having the corpus available as training data;
- only annotate when high inter-annotator agreement is possible;
- focus on annotation that is fast and cheap (relatively), that doesn't require lots of annotator training, and that doesn't require years to carry out;
- ensure that the annotations are theoretically well-founded and acceptable to a large number of people in various projects;
- build on previous efforts, and use automated tools to speed up annotation if possible;
- pay particular attention to annotation interface design, since this can significantly impact performance.

In order to meet these desiderata, many annotation efforts have made decisions that may be seen as compromises. For example, by using the Penn Treebank texts, one can count on a commonly-understood parse tree syntax. However, the Treebank is not a balanced corpus, and hence may negatively influence the results annotations that reflect phenomena not present in that corpus.

On the panel, members of three semantic annotation projects will describe their work and provide insights as to where they had to make compromises in the light of the desiderata and why they did so:

- SALSA (Manfred Pinkal)
- PropBank (Martha Palmer)
- IL-Annotation (David Farwell and Eduard Hovy)

Can we move from sentence meaning to text meaning?

Convener: **Sergei Nirenburg**, University of Maryland, Baltimore County

Panelists: TBA

Text meaning as a whole has not yet attracted widespread attention. Recent studies usually concentrate on text-meaning components — propositional meaning within a single sentence or even clause, relations among clauses, or co-reference issues. Earlier "holistic" work on

text-level "grammars" or plot units did not reach the stage where the main ideas were ripe for judgments of explanatory power or utility. One can indeed view text meaning as a combination of the meaning of its clauses plus causal, temporal, rhetorical, and other relevant relations among the clause meanings, plus speaker attitudes expressed in the input text. At this level, a central issue is cross-fertilization of heuristic material — how one can use findings in one component of the overall text meaning as heuristics for establishing elements of another component? For example, the propositional meaning of a clause can contribute to establishing a coreferential relation between the meaning of a noun phrase within it and a noun phrase in another clause.

Extracting and manipulating the meaning of an entire text holds the promise of improving the quality of results in information extraction, automatic population of knowledge bases, text summarization, modeling question answering and other intelligent agent systems that communicate with people, and other applications. The needs of specific applications effectively define the scope and depth of text meaning in specific projects. The spectrum of choices here is very broad — from approximating text meaning through textual collocation (the "knowledge-lean" end of the spectrum) to including in text meaning the results of reasoning — for example, judgments about speaker goals and beliefs (the "knowledge-rich" end). The choice is made by balancing two conflicting desiderata — real utility and feasibility. In this discussion, we will analyze the available choices and assess the practicality and the promise of integrating work on different components of text meaning.

Author Index

Amoia, Marilisa 73
Baldwin, Timothy 49
Beale, Stephen 25, 33, 41
Berglund, Anders 57
Delmonte, Rodolfo 97
Dorr, Bonnie 65
Farwell, David ix
Gardent, Claire 73
Green, Rebecca 65
Hempelmann, Christian F. 105
Horacek, Helmut 81
Hovy, Eduard ix
Jacquey, Evelyne 73
Johansson, Richard 57
Korelsky, Tanya 41
Kruijff-Korbayová, Ivana 81
Lavoie, Benoit 41
Lev, Iddo 9
Levy, Roger 9
MacCartney, Bill 9
Manning, Christopher 9
Max, Aurélian 113
McShane, Marjorie 25, 33, 41
Nirenburg, Sergei . . . . . ix, 25, 33, 41
Nugues, Pierre 57
Paredes-Frigolett, Harold 1
Raskin, Victor 105
Sonnenhauser, Barbara 89
Triezenberg, Katrina E. 105
Williams, David 57
Winnemöller, Ronald 17
Wolska, Magdalena 81
Zabludowski, Margalit 25