Inference in Computational Semantics ICoS-5

Buxton, England

April 20–21, 2006

Workshop Proceedings

Johan Bos and Alexander Koller (Eds.)

About ICoS

Natural Language Processing has reached a stage where the exploration and development of inference is one of its most pressing tasks. On the theoretical side, it is clear that inference plays a key role in such areas as semantic construction and the management of discourse and dialogue. On the practical side, the use of sophisticated inference methods is expected to play a crucial role in such application areas as natural language generation, automatic question answering, and spoken dialogue systems. At the same time, inference tools and resources have matured to the point that they can become useful for actual applications. Automated theorem provers and other inference tools are becoming increasingly robust and efficient. And the world knowledge bottleneck is being addressed from different angles, by the manual development of knowledge resources, the merging of existing such resources, and the automated extraction of world knowledge from corpora.

The Inference in Computational Semantics (ICoS) workshops are intended to bring together researchers from areas such as Computational Linguistics, Artificial Intelligence, Computer Science, Formal Semantics, and Logic, to discuss approaches to, and applications of, inference in natural language semantics. ICoS-1 took place in Amsterdam, the Netherlands, on August 15, 1999. ICoS-2 was organised in Dagstuhl Castle, Germany, on July 29–30, 2000. ICoS-3 was co-located with the International Joint Conference on Automated Reasoning (IJCAR 2001), which took place on June 18–23, 2001 at Siena, Italy. ICoS-4 took place in Nancy, France, on September 25–26, 2003.

Welcome to ICoS-5

ICoS-5 is organised as a two-day event at the University of Derby College, Buxton, England, taking place on April 20–21. The programme features three invited presentations (by Christian Ebert, Patrick Pantel, and Stephen Pulman). In addition, we have selected twelve regular papers from 24 submissions, which span topics ranging from the use of inference techniques for formal and computational semantics, over new methods for knowledge extraction from corpora, to NLP applications that use inference tools.

We have accepted six of the remaining submissions as "short papers". These papers will be presented as posters and demos at the workshop, and we believe that this new format will be a worthwhile experience both for authors and for participants.

We would like to thank the members of the programme committee, who produced highly informative reviews and made it possible to set up the broad and strong program for ICoS-5. Finally, we are very grateful to the local organisers, Ian Pratt-Hartmann and Allan Ramsay, who did a flawless job of preparing what promises to be a wonderful workshop location for us.

Rome and Saarbrücken, March 2006
Johan Bos & Alexander Koller (co-chairs)

Workshop Organization

Programme Committee

Carlos Areces	INRIA Lorraine
Peter Baumgartner	National ICT Australia
Christoph Benzmueller	University of the Saarland
Raffaella Bernardi	Free University of Bozen-Bolzano
Patrick Blackburn	INRIA Lorraine
Johan Bos (co-chair)	University of Roma "La Sapienza"
Harry Bunt	Tilburg University
Ann Copestake	University of Cambridge
Dick Crouch	PARC
Ido Dagan	Bar Ilan University
Kees van Deemter	University of Aberdeen
Nissim Francez	Technion
Claire Gardent	CNRS/LORIA
Alexander Koller (co-chair)	University of the Saarland
Shalom Lappin	King's College London
Alex Lascarides	University of Edinburgh
Bernardo Magnini	ITC-Irst
Katja Markert	University of Leeds
Dan Moldovan	University of Texas at Dallas
Jeff Pelletier	Simon Fraser University
Maarten de Rijke	University of Amsterdam
Michael Schiehlen	University of Stuttgart
Matthew Stone	Rutgers
Bonnie Webber	University of Edinburgh

Local Organizers

Ian Pratt-Hartmann	School of Computer Science, University of Manchester
Allan Ramsay	School of Informatics, University of Manchester

Table of Contents

Abstracts of Invited Talks

Regular Papers

Short Papers

Expressive Power and Complexity
of Underspecified Representations

Christian Ebert

Fakultät für Linguistik und Literaturwissenschaft

Universität Bielefeld

My talk will be about two requirements on *Underspecified Representation Formalisms* in the context of underspecification of scope. The requirement on *partial disambiguation*, stating that partially disambiguated ambiguities need to be represented, does not carry much content unless it has become clear, exactly what those ambiguities are. In line with König and Reyle [1999], I will argue that *all* theoretically possible patterns of ambiguity, i.e. subsets of readings, can occur in natural language, in particular when underspecified representations are assumed to represent patterns of ambiguity that arise through disambiguation by discourse or even by world knowledge. Therefore an underspecified representation formalism can only be regarded as *expressively complete*, if it provides representations for all potential subsets of readings. Taking a closer look at recent prominent approaches to scope underspecification, namely Hole Semantics [Bos, 2002], Minimal Recursion Semantics [Copestake et al., 1999], and Normal Dominance Constraints [Koller, 2004], it turns out that none of these formalisms is expressively complete. Furthermore, these incompleteness results allow for a straightforward comparison of the discussed approaches with respect to expressive power.

The second requirement is the *avoidance of combinatorial explosion*. I will argue that the decisive process that determines the efficiency of an underspecification approach is the construction phase of the representations and not so much the check for satisfiability or enumeration of readings. Thus the desired avoidance of combinatorial explosion comes down to a requirement on a feasible construction procedure and hence to a requirement on the maximal 'size' of the representations, which can only be fulfilled if the involved representations are in some sense more *compact* than the mere listing of the available readings. Unfortunately it turns out that due to the rapid growth of the number of potential patterns of ambiguity, the two requirements of compactness and expressive completeness cannot be fulfilled at the same time. In other words, I will show that any underspecified representation formalism must necessarily miss out on some of the potential patterns of ambiguity or run into the combinatorial explosion problem [Ebert, 2005].

References

Johan Bos. *Underspecification and Resolution in Discourse Semantics*. PhD thesis, Universität des Saarlandes, 2002.

Ann Copestake, Dan Flickinger, and Ivan A. Sag. Minimal Recursion Semantics – An Introduction. Technical report, CSLI, Stanford University, 1999. (Draft).

Christian Ebert. *Formal Investigations of Underspecified Representations*. PhD thesis, King's College London, 2005.

Alexander Koller. *Constrained-Based And Graph-Based Resolution of Ambiguities in Natural Language*. PhD thesis, Universität des Saarlandes, July 2004.

Esther König and Uwe Reyle. A General Reasoning Scheme for Underspecified Representations. In *Logic, Language and Reasoning. Essays in Honour of Dov Gabbay.*, pages 251–277. Kluwer, 1999.

Knowledge Harvesting and Fusion from Small and Large Corpora

Patrick Pantel
ISI, University of Southern California

Inferencing requires large collections of axioms and knowledge bases of interrelated concepts and instances. In the past decade, researchers have explored many approaches to automatically learn such knowledge from textual corpora. In this two-part talk, we will discuss the challenges of harvesting knowledge from various corpus sizes and we will look at some recent attempts at fusing the knowledge into a single semantic repository. On the harvesting front, we will focus on the challenges posed by three types of corpora: i) the Web (order of 10^{12} words) – pattern learning algorithms, models of reliability, and scalable computing; ii) large corpora such as newswire collections (order of 10^9 words) – amenable to complex natural language processing and clustering; and iii) small corpora such as college textbooks (order of 10^5 words) – in which low redundancy requires leveraging external resources such as the Web and ontologies.

This multitude of harvested knowledge inevitably overlaps and is partially inconsistent and incompatible since information can be expressed across data sources in so many ways. However, little effort has been spent on fusing such harvested knowledge into an overarching consistent semantic repository. We will present a recent algorithm for linking semantic resources using grammatical templates. We will also present an algorithm to automatically induce conceptual relations between mid-level ontological concepts and then disambiguate instances with regard to the most appropriate senses in the ontology, using the induced conceptual relations.

Bridging the gap between formal and computational semantics

Stephen Pulman
Computational Linguistics Group, Oxford University.

The literature in formal linguistic semantics contains a wealth of fine grained and detailed analyses of many linguistic phenomena. But very little of this work has found its way into implementations, despite a widespread feeling (among linguists at least) that this can't be very difficult in principle: you just fix a grammar to produce the right logical forms and hook them up to a theorem prover, don't you? In this talk I take a representative analysis of adjectival comparatives and ask what steps one might actually have to go through so as to use this analysis in a realistic question-answering setting. I then try to identify some general conclusions that can be drawn from this exercise.