Automated Reasoning for Computational Semantics

Patrick Blackburn, Johan Bos, Michael Kohlhase
Universität des Saarlandes, Saarbrücken, Germany

May 7, 1999

Abstract

This paper discusses inference in computational semantics. We argue that state-of-the-art methods in first-order *theorem proving* and *model building* are of direct relevance to inference for natural language processing. We support our claim by discussing the inferential aspects of several higher discourse phenomena and reporting on an experiment where the induced deduction problems are solved by the MATHWEB society of theorem proving agents.

Keywords: Automated Reasoning, discourse, natural language processing, theorem proving

1 Introduction

Semantic analysis – inference on the basis of semantic information and world knowledge – is one of the central cognitive tasks in natural-language processing (NLP) and Artificial Intelligence. It is needed for situation-dependent disambiguation and for the coherent embedding of utterances into the discourse context. Humans obviously have at their disposal very efficient techniques for semantic analysis, in NLP, similarly powerful techniques have yet to be found.

Early attempts from artificial intelligence [Win71, Cul78, Rie75], have had some limited success, but the inference components have failed to scale up to real-world examples. The field of automated theorem proving (ATP[1]) has seen an enormous increase of preformance of inference engines. However, the application of ATP systems as off-the-shelf components for NLP systems has been deemed impossible, since

- First-order predicate logic is not well-suited as a representation language for the semantic structures of natural language discourse (see section 2),
- ATP systems are optimized towards finding deep combinatorially complex proofs of (mathematical) theorems rather than towards the straightforward proofs needed for semantical analysis,
- Many of the inference problems necessary for semantical analysis are satisfiable and termination has not been a priority goal of current automated theorem proving systems.

In this paper, we demonstrate the feasibility of this proposal using a *translation approach* together with ATP: The translation from dynamic logic (see the next section) to first-order logic allows us to get around the first problem and refute the other objections on several discourse inference problems encountered in semantic analysis.

Perhaps the most important fact about current ATP systems is the variety that are available and the speed many of them offer. Now, it is hard to say anything general about what is likely to constitute a good choice of theorem prover for natural language (beyond the fact that in general natural language applications will require theorem provers that handle equality, a stumbling block for many tableaux based systems). Indeed, we argue that the best idea is not to choose at all but to farm out the inference task to many different ATP simultaneously.

---

[1] For the purposes of this paper, we will subsume model generation under ATP.

In an experiment we have combined the Doris[2] system, (**D**iscourse **O**riented **R**epresentation and **I**nference **S**ystem.) with distributed MathWeb theorem proving environment [FK99] (see section 4), which provides the services of many state-of-the art ATP. In this agent-oriented software environment, Doris acts as a client of the MathWeb theorem proving agents.

The Doris system is an implementation of the computational semantics tools provided by [BB98]; it constructs discourse representations for a considerable fragment of English, dealing with phenomena like scope ambiguities, pronoun resolution and presupposition projection. The emphasis of the system is on the semantic analysis phase, where (spurious) ambiguities that are artifacts of the specific semantics construction process are analyzed and eliminated. For this, the system generates first-order deduction problems that are solved by passing them to the society of MathWeb agents that compete for solving them.

In the rest of the paper, we will give a very brief reminder to Discourse Representation theory [KR93] (section 2) and then explain the inferential aspects of various discourse phenomena (section 3), most notably van der Sandt's dynamic theory of presuppositions.

2 Dynamic Representation Formalisms

One of the main problems with first-order predicate logic for representing natural language is that the accessibility of discourse referents (modeled as bound variables) is given by the logical scope induced by first-order quantification, which is insufficient to model phenomena like anaphoric references.

The so-called dynamic approaches to natural language semantics (Discourse Representation Theory (DRT, see e.g. [KR93]) or dynamic predicate logic (DPL [GS91])) have been developed to cope with this (and related) problems and are now well-established as representation formalisms for natural language semantics. They now constitute one major pillars of logic-based natural language semantics research.

We will concentrate on DRT in this paper. There, sentences and discourses are represented as *discourse representation structures* (DRS); objects that are dynamically introduced in a discourse are not represented by bound variables but by so-called *discourse referents* in the DRSes – which collect discourse referents and information about them. Due to space restrictions, we presuppose that the reader is familiar with DRT and otherwise refer the reader to [KR93].

There are two approaches to inferencing in dynamic logics. The first — which we pursue in this note — is to use the (dynamic) deduction theorem to encode the (dynamic) entailment problem as a (dynamic) satisfiability problem (a DRS) and then translate that DRSs to first-order logic (see [KR93]) and test for satisfiability there. The second paradigm is to develop a calculus for (dynamic) entailment or satisfiability that operates on the dynamic structures themselves (see [Sau93, RG94, MdR98, KK99] for theorem proving and model generation calculi). While the second (more specialized) approach might promise better results in the long run, the first approach allows us to make use of the highly developed automated theorem proving systems that are available today.

The translation approach can also be varied in the translation that is employed. Jan van Eijck has developed an alternative (linear complexity) translation (see e.g. [vEK96]) using the weakest-precondition-calculus. It remains to be seen how the FOL fragment generated by this translation compares to that of our naive translation.

3 Inference in Semantic Analysis

In this section, we will take a closer look a three classes of inference problems occurring during the semantic analysis phase of natural language processing. At this stage, the discourse has already undergone syntactic processing, semantic construction, and anaphora resolution in Doris which together have generated a set of discourse representation structures. This set can be quite large, due to ambiguities that arise from well known phenomena as quantifier scope and anaphora, and

---

[2] Cf. `http://www.coli.uni-sb.de/~bos/atp/doris.html` for a web-based interface.

one way to deal with this is by imposing pragmatically motivated conditions on the DRSs that decrease the number of readings[3].

These *conversational principles* [Sta79] require that a new utterance in a discourse should be informative and consistent, i.e. it should contribute information that is still unknown and it should not lead to obvious contradictions. Clearly, these principles are only non-trivial if they are applied with respect to a given set of world knowledge and the context of the discourse so far, so that checking them leads to general inference problems.

In the rest of this section, we will give examples that violate these three conditions and show the use of ATP systems. For each example we give the translation into discourse representation structures, and, using the translation to first-order logic, show how we DORIS uses ATP to select readings or to rule out the whole discourse. We will only talk about informativity and consistency, and refer the reader to [BBKdN98b] for the cases of presupposition projection and quantifier scope.

In [VdS92], Van der Sandt models the informativity principle as follows:

> *A DRS $\mathcal{B}'$ is informative with respect to a DRS $\mathcal{B}$, iff $\mathcal{B}$ does not entail $\mathcal{B}'$.*

With the background knowledge that someone who has a husband is married, the discourse

(1) Mia has a husband. She is married.

violates this principle. The DRSs after processing the first and second sentence respectively, are:

(2)
$$a.\ \begin{array}{|c|}\hline U,V\\ \hline U=mia\\ husband(V)\\ of(U,V)\\ \hline\end{array} \qquad\qquad b.\ \begin{array}{|c|}\hline U,V,U'\\ \hline U=mia\\ husband(V)\\ of(U,V)\\ U'=U\\ married(U')\\ \hline\end{array}$$

The background knowledge about marriage is coded into first-order logic:

(3) A woman is married, iff she has a husband.

(4)
$$\begin{array}{l}\forall X.(\exists Y.husband(Y)\wedge of(X,Y))\\ \qquad\equiv married(X)\wedge woman(X)\end{array}$$

Note that in the approach advocated in this paper it is easy to integrate static background knowledge (given in FOL) with DRT, since the latter is translated to FOL anyway.[4]

In this situation, we can test informativity by checking whether

(5) $(3)\wedge(2.a)^{fo}\Rightarrow(2.b)$

is a theorem of first-order logic. In our example we should find a proof, as there is no new information conveyed by the second sentence.

Next we discuss a variation of informativity, the *local informativity* constraint. The local informativity constraint is that if one utters a phrase of the form: *If A then B*, then *A* should be not trivially satisfied. (So, here the if-then from natural language clearly differs from its logical counterpart). In the following example this condition is violated:

(6) Mia has a husband. If she is married, then Vincent dances.

The DRSs belonging to these sentences are:

---

[3] This is essential for practical NLP, since in discourse or dialogue processing applications, the numbers readings of the sentences *multiply* to the number of readings of the whole discourse or dialogue.

[4] In a dynamic deduction approach, the background knowledge would have to be formulated in DRT, or the approach would need to be extended to accommodate for first-order reasoning.

(7) $$\begin{array}{|c|}\hline U,V\\\hline U = mia\\ husband(V)\\ of(U,V)\\\hline\end{array}$$

(8) $$\begin{array}{|c|}\hline \\ \hline \begin{array}{|c|}\hline U'\\\hline U'=U\\ married(U')\\\hline\end{array} \Rightarrow \begin{array}{|c|}\hline V'\\\hline V'=vincent\\ dance(V')\\\hline\end{array}\\\hline\end{array}$$

When these DRSs are combined the result equals:

(9) $$\begin{array}{|c|}\hline U,V\\\hline U = mia\\ husband(V)\\ of(U,V)\\ \begin{array}{|c|}\hline U'\\\hline U'=U\\ married(U')\\\hline\end{array} \Rightarrow \begin{array}{|c|}\hline V'\\\hline V'=vincent\\ dance(V')\\\hline\end{array}\\\hline\end{array}$$

This last DRS violates the local informativity constraint, since

(10) $$(3) \models \begin{array}{l}\forall X,Y.(X = mia \wedge of(X,Y) \wedge husband(Y))\\ \qquad \Rightarrow married(Y)\end{array}$$

The last condition that we check is *consistency*. Say we had continued (2.a) with the utterance She is not married, paraphrased by the following DRS:

(11) $$\begin{array}{|c|}\hline U'\\\hline U'=U\\ \neg\begin{array}{|c|}\hline \\ \hline married(U')\\\hline\end{array}\\\hline\end{array}$$

Clearly, the new information is inconsistent with the information that is already present (implicitly) In this situation, we can check for informativity by checking whether

(12) $(3) \wedge [(2.a) \otimes (11)]^{fo}$

is unsatisfiable.

4 The MathWeb System

The MathWeb system is an object-oriented toolbox that provides the functionality for building a society of software agents that render mathematical services by either encapsulating legacy deduction software or their own functionality. In the current implementation the software bus functionality is realized by a model quite similar to the *Common Object Request Broker Architecture* (CORBA [Sie96]) in which a central *broker* agent provides routing and authentication information to the mathematical services (see [SHS98] for details). The agents are realized in a distributed programming system MOZart[5], which provides the full infrastructure to write distributed applications.

The MathWeb services relevant for Doris include the first-order ATP bliksem, EQP, Otter, ProTeIn, Spass, WaldMeister, the model generator Satchmo (see [SS97] for references) and a service `competitive-atp` that calls sets of ATP concurrently as competing services (this strategy is known to yield even super-linear speedups in practice).

The Doris client generates between 1 and ca. 500 deduction problems for each sentence it processes, distributes them to competing mathematical services (over a network of workstations)

---

[5] See `http://mozart.ps.uni-sb.de`

and collects the results to obtain the desired result. Using the MathWeb approach, the integration of the theorem provers was very simple: the only new parts was a socket connection from Prolog on the Doris side and a new service module for the Doris service[6] on the MathWeb side. Experience with this application shows that distribution using MathWeb does not come for free: for a typical Doris deduction query we have[7]

80–250 ms pure theorem proving time

150-350 ms spent in the service module (opening an inferior shell, creating files,...). This depends strongly on the efficiency of the server file system.

5–500 ms Internet latency (we have measured inter-department (in Saarbrücken) and international (Saarbrücken/Amsterdam) connections)

However, the large number of deduction problems and the possibility of coarse-grained parallelization by distribution lead to a significant increase in overall system performance, compared to an earlier centralized, sequential architecture [BBKdN98a, BBKdN98b].

The current CORBA-like distribution model in MathWeb is sufficient in an agent society, where services and their abilities are relatively fixed and well-known, which is reasonable for the relatively closed projects like Doris. As the number of available services will grow (MathWeb has for instance been adopted by other projects building on Doris), this design will become too inflexible. Therefore the logical next step will be to adopt a more general truly agent-based approach. We have started to extend MathWeb so that it uses the Kqml interlingua (*Knowledge Query and Manipulation Language* [FF94]) as the agent interaction language and the OpenMath [Cap98] standard as a content language.

This move will result in a "plug-and-play" architecture for theorem proving and (in the future) for doing mathematics and computational semantics on the web.

5 Conclusion

In this paper we have reported on an application of current ATP technology in natural language processing. We have shown that first-order ATP systems can successfully be employed as oracles for NLP systems to disambiguate multiple readings.

While the experiment has shown that the naive translation approach to dynamic reasoning is indeed feasible in this application, it is clear that in the presence of larger discourses (the ones tried out so far only consist of tens of sentences), the techniques have to be refined both from the linguistic side as well as from the theorem proving side. For instance the set of formulae supplied to the automated theorem prover can be restricted by taking into account the discourse structure (see for instance [Gar97]).

The general consequences for research in computational semantics are profound: With the use of the highly optimized and efficient theorem proving systems as logical engines and the MathWeb technology to make the integration of them into NLP applications an easy task it will be simple to test inferential theories of meaning in natural language semantics, as we have done in Doris with van der Sandt's anaphoric theory of presuppositions. In fact possibility to work more and larger examples than would be possible by hand have uncovered shortcomings in this theory and have led to a revised account in [BBKdN98b].

References

[BB98] Patrick Blackburn and Johan Bos. Representation and Inference for Natural Language. A First Course in Computational Semantics. Draft available at URL: `http://www.coli.uni-sb.de/~bos/comsem/`, July 1998.

---

[6] I.e. a small (60 line) MOZART program that relays problems, results and statistics between the Doris program and the `competitive-atp` service

[7] These times have been measured on a collection of SUN Ultra machines running Solaris 5 in Saarbrücken and Amsterdam (all timings given in total elapsed time).

[BBKdN98a] Patrick Blackburn, Johan Bos, Michael Kohlhase, and Hans de Nivelle. Automated theorem proving for natural language understanding. In Peter Baumgartner, Ulrich Furbach, Michael Kohlhase, William McCune, Wolfgang Reif, Mark Stickel, and Tomàs Uribe, editors, *CADE-15 Workshop "Problem-solving Methodologies with Automated Deduction"*, 1998. Paper available at URL: `http://www.uni-koblenz.de/~peter/cade-15-ws/`.

[BBKdN98b] Patrick Blackburn, Johan Bos, Michael Kohlhase, and Hans de Nivelle. Inference and computational semantics. CLAUS Report 99, University of the Saarland, Saarbrücken, 1998. To appear at IWCS-3, Tilburg, 1999.

[Cap98] Draft of the open math standard. Open Math Consortium, `http://www.nag.co.uk/projects/OpenMath/omstd/`, 1998.

[Cul78] R.E. Cullingford. *Script Application: Computer Understanding of Newspaper Stories.* PhD thesis, Yale University, New Haven, CT, 1978.

[FF94] T. Finin and R. Fritzson. KQML — a language and protocol for knowledge and information exchange. In *Proceedings of the 13th Intl. Distributed Artificial Intelligence Workshop*, pages 127–136, Seattle, WA, USA, 1994.

[FK99] Andreas Franke and Michael Kohlhase. System description: MATHWEB, an agent-based communication layer for distributed automated theorem proving. submitted to CADE'99, 1999.

[Gar97] Claire Gardent. Discourse tree adjoining grammars. CLAUS Report 89, University of the Saarland, Saarbrücken, 1997.

[GS91] Jeroen Groenendijk and Martin Stokhof. Dynamic predicate logic. *Linguistics & Philosophy*, 14:39–100, 1991.

[KK99] Michael Kohlhase and Karsten Konrad. Model generation for discourse representation theory. SEKI-Report SR-99-01, Dept. of Computer Science, Universität des Saarlandes, Germany, 1999.

[KR93] Hans Kamp and Uwe Reyle. *From Discourse to Logic.* Kluwer, Dordrecht, 1993.

[MdR98] C. Monz and M. de Rijke. Labeled resolution for discourse semantics. In *Proceedings of the First International Workshop on Labelled Deduction (LD'98).* Dept. of Computer Science, University of Freiburg, 1998.

[RG94] Uwe Reyle and Dov M. Gabbay. Direct deductive computation on discourse representation structures. *Linguistics & Philosophy*, 17:343–390, 1994.

[Rie75] Chuck Rieger. The commonsense algorithm as a basis for computer models of human memory, inference, belief and contextual language comprehension. In *Theoretical Issues in Natural Language Processing.* MIT Press, 1975.

[Sau93] Werner Saurer. A natural deduction system for discourse representation theory. *Journal of Philosophical Logic*, 22, 1993.

[SHS98] M. Kohlhase S. Hess, Ch. Jung and V. Sorge. An implementation of distributed mathematical services. In *6th CALCULEMUS and TYPES Workshop*, Eindhoven, The Netherlands, July 13–15 1998. Electronic Procedings: http://www.win.tue.nl/math/dw/pp/calc/proceedings.html.

[Sie96] Jon Siegel. *CORBA: Fundamentals and Programming.* John Wiley & Sons, Inc., 1996.

[SS97] Geoff Sutcliffe and Christian Suttner. The results of the cade-13 atp system competition. *Journal of Automated Reasoning*, 18(2):259–264, 1997. Special Issue on the CADE-13 Automated Theorem Proving System Competition.

[Sta79] R. Stalnaker. Assertion. In P. Cole, editor, *Syntax and Semantics 9: Pragmatics*, pages 315–32. Academic Press, 1979.

[VdS92] Rob A. Van der Sandt. Presupposition projection as anaphora resolution. *Journal of Semantics*, 9(4):333–377, 1992.

[vEK96] Jan van Eijck and Hans Kamp. Representing discourse in context. In Johan van Benthem and Alice ter Meulen, editors, *Handbook of Logic and Language*, pages 179–238. Elsevier Science B.V., 1996.

[Win71] Terry Winograd. *Procedures as a representation for data in a computational program for understanding natural language.* PhD thesis, MIT, Cambridge, 1971.