LINGUISTIC ISSUES
in
LANGUAGE TECHNOLOGY
VOLUME 9

PERSPECTIVES ON SEMANTIC REPRESENTATIONS FOR TEXTUAL INFERENCE

edited by
Cleo Condoravdi, Valeria de Paiva
and Annie Zaenen

LINGUISTIC ISSUES IN LANGUAGE TE

VOLUME 9

PERSPECTIVES ON SEMANTIC REPRESENTATIONS FOR TEXTUAL INFERENCE

edited by

Cleo Condoravdi, Valeria de Paiva and Annie Zaenen

CSLI
PUBLICATIONS
Center for the Study of
Language and Information
Stanford, California

Center for the Study of Language and Information
Leland Stanford Junior University
Printed in the United States
18 17 16 15 14   1 2 3 4 5

LiLT Volume 9

ISBN: 9781575868448 (pbk)
ISSN: 1945-3604

∞ The acid-free paper used in this book meets the minimum requirements of the American National Standard for Information Sciences—Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

CSLI was founded in 1983 by researchers from Stanford University, SRI International, and Xerox PARC to further the research and development of integrated theories of language, information, and computation. CSLI headquarters and CSLI Publications are located on the campus of Stanford University.

An online version of this journal is available at CSLI Publications
http://cslipublications.stanford.edu/

Acknowledgments

We would like to thank the contributors to the workshops that this volume is based on, the authors of the articles and our dedicated reviewers:

Danny Bobrow	Larry Moss
Robin Cooper	Charles Ortiz
Jan Van Eijck	Stephen Pulman
Nissim Francez	Stephen Read
David Israel	Mark Sammons
Bill MacCartney	Hinrich Schütze
Wei Lu	Mark Steedman
Gerard de Melo	Alessandra Zarcone

We also thank the staff of CSLI and especially Emma Pease for the help in producing this volume.

Introduction

ANNIE ZAENEN, CLEO CONDORAVDI, VALERIA DE PAIVA

The introduction of the RTE (Recognizing Textual Entailment) paradigm in 2004 constituted a turning point in computational work on inference in natural language. That paradigm sees the task of determining the relation between two texts as one where assuming the truth of the first text, T, the thesis, leads to concluding to the (likely) truth or falsity of the other text, called H, the hypothesis. Later variants extended the relation also to contradictions. It has focused work on textual inference on tasks that are on the one hand feasible and on the other hand based on real texts. In this volume we present some of the work that arose from this conceptualization of the task, mostly but not only focussing on methods that involve logical formalizations. The volume is based on three workshops that we organized in 2011 and 2012: the *CSLI Workshop on Natural Logic, Proof Theory, and Computational Semantics* on April 8 and 9, 2011 at Stanford,[1] *Semantics for textual inference* on July 9/10, 2011 at the University of Colorado at Boulder,[2] and the CSLI workshop on *Semantic Representations for Textual Inference* on March 9 and 10, 2012 at Stanford.[3]

The first paper in this volume, "The BIUTEE Research Platform for Transformation-based Textual Entailment Recognition" by Asher Stern and Ido Dagan, is from the lab that introduced the current textual inference paradigm, RTE. The paper is an interim report on BIUTEE, an open source and open architecture platform that allows users to develop components for textual inference systems. The system consists of a preprocessing module that does parsing, named en-

[1] `http://www.stanford.edu/~icard/logic&language/`

[2] `http://www.stanford.edu/~cleoc/Sem-Text-Inf/`

[3] `http://semanticrepresentation.stanford.edu`

LiLT Volume 9
Perspectives on Semantic Representations for Textual Inference.

tity recognition and coreference resolution and several modules that do inference recognition. The architecture allows the user to insert different lexical/knowledge resources (e.g. DIRT, WordNet, Wikipedia) and to add his/her own modules. The inference recognition engine itself is transformation-based, constructing derivations that go from text to hypothesis via rewrite rules. This procedure that can be seen as an instance of a proof, with the difference from traditional formal proofs being that the procedure allows for conclusions to likely true (false) instead of aiming for certainty. The system presented in this paper has meanwhile been integrated in an wider platform, EXCITEMENT, that allows more flexibility.[4] The hope is that ultimately the community interested in textual inference will have a platform comparable to that provided to the (statistical) translation community by MOSES.[5]

There are two other papers describing full systems. They concentrate on building systems that allow true textual entailments and are not geared to allowing conclusions that are only *likely* to be drawn by native speakers.

The first paper "Is there a place for logic in recognizing textual entailment" by Johan Bos describes Nutcracker, a system that integrates Combinatory Categorial Grammar, Discourse Representation Theory (translated into first-order formulas through the reification of modality) and first order theorem proving. It argues that the main problem for such systems is the acquisition of the relevant *background knowledge* and shows how axioms deriving synonyms and hyponyms relations can be automatically derived from WordNet information. It proposes axioms for verbs of 'saying' that allow one to conclude that a reported event holds in the real world assuming one trusts the source and axioms learned from the RTE sets themselves. Several different theorem provers can be plugged into the system as well as different model builders that search for models up to a specified domain size. The analysis terminate either with (i) a proof, (ii) a finite counter model of size n, or (iii) neither. The system has, as expected, high precision but low recall as the knowledge acquisition problem is only partially solved.

Another take on textual inference is provided in the paper by Lenhart Schubert "NLog-like Inference and Commonsense Reasoning". The paper describes the key properties of the Epilog system, an implementation of Episodic Logic, whose language is meant to provide a target representation for natural language. Epilog performs Natural Logic kind of inferences, but it goes beyond them in two ways: it can

[4] `http://www.excitement-project.eu`

[5] `http://www.statmt.org/moses/`

perform goal-directed and forward inferences, not just inferences with known premises and conclusions, as well as inferences based on lexical knowledge and language-independent world knowledge. Moving beyond narrow textual inference to common sense inference, essential to natural language understanding, requires addressing the problem of the "knowledge acquisition bottleneck". The paper describes various under way and potential efforts for gathering knowledge from texts. These are based on the idea that one can recover some of the background knowledge assumed in one text from what is stated in another. In addition to lexical knowledge, necessary for inferences based on meaning, and world/common sense knowledge, necessary for general reasoning and true understanding, it recognizes semantic pattern knowledge. This knowledge comprises general 'factoids' that guide parsing and interpretation. Such factoids can also be used to acquire world knowledge via factoid strengthening and factoid sharpening as defined in the contribution.

The RTE challenge stressed the importance of working on real text and to take into account all the phenomena that contribute to making an inference valid or not. In this it is contrasted with earlier approaches that would concentrate on getting inferences involving specific semantic phenomena, e.g. quantifiers (Cooper et al. 1996, 1994) but after several years, the need to decompose the TE task into basic phenomena and the way these basic phenomena interact. Two papers address that issue.

Elena Cabrio and Bernardo Magnini in "Decomposing Semantic Inferences" look at the problem from an empirical angle. They analyze a TE data set looking at the nature of the inference (deductive, inductive, adductive) and the linguistic phenomena involved (e.g. synonymy, coreference, negation, active/passive alternation). Their results show that a huge amount of background information is required to approach the TE task. They also decompose the inferences of the thesis-hypothesis TH pairs into smaller atomic inference pairs consisting of a linguistic and an inference pattern. They conclude that "the polarity of most of the phenomena is not predictable for the logical judgments" and point out the consequences for attempts to learn from the annotated RTE data sets.

Assaf Toledo et al., in "Towards a Semantic Model for Textual Entailment Annotation", develop a theoretical model of entailment recognition. The main idea consists in providing an interpreted lexicon, which specifies the semantic types and denotations of content and function words, and which ultimately serves as a target canonical representation for constructions in Text and Hypothesis sentences. After "binding to" an interpreted lexicon, an inferential relation can be proven between T and H using predicate calculus and lambda calculus reduction, or disproven by the construction of a countermodel. Starting from the assumption that the model can incrementally incorporate increasingly complex phenomena, the authors concentrate on three prevalent inferential phenomena in the RTE data bases: intersective, restrictive, and appositive modification. At the same time, they acknowledge that interaction between phenomena might significantly complicate scaling up their model. The contrast between intersective vs. restrictive modification provides a nice illustration of how expressions with the same syntactic structure can have radically different inferential properties and how "binding to" an interpreted lexicon can model the difference.

Four of the remaining papers focus on logic more traditionally construed.

Alex Djalali's "Synthetic Logic" reminds us that derivability plays as central a role in NL semantics as that of entailment and that it makes sense to develop a more proof-based approach to the logic of Natural Language, where one tries to capture the sorts of inferences speakers make in practice. Djalali considers MacCartney and Manning's model of Natural Logic (MacCartney and Manning 2009; MacCartney 2009) as a kind of generalized transitive reasoning system and makes it, in the process, much easier to understand as a system of logic than the original system. His proof rules in Gentzen-style sequent calculus are divided into M-rules (which explain the composition of MacCartney and Manning relations) and D-rules, which correspond to structural properties of the MacCartney relations themselves. Djalali's soundness and completeness proofs are crisp and enlightening, despite, like MacCartney and Manning, dealing only with a fragment of the algorithm developed for the implemented system NatLog.

The paper by Icard and Moss summarizes classic as well as more recent work on monotonicity reasoning in natural language. They first offer an informal overview of work on the Monotonicity Calculus, beginning with van Benthem and Sánchez-Valencia in the late 1980s, and continuing on to the present day, including extensions, variations, and applications. Alongside examples from natural language, they also present analogous examples from elementary algebra, illustrating the fact that the Monotonicity Calculus makes sense as a more general system for reasoning about monotone and antitone functions over (pre)ordered sets. Following a discussion of current logical, computational, and psychological work on monotonicity in natural language, they develop a fully explicit Monotonicity Calculus using markings on simple types, with a well-defined language, semantics, and proof system, and culminating in an overview of soundness and completeness results, pointing to recent and forthcoming work by the authors.

The paper by Ian Pratt-Hartman, "The Relational Syllogistic Revisited" is part of the tradition of extending the original syllogistic calculus. In previous work Moss and Pratt-Hartman introduced the relational syllogistic, an extension of the language of classical syllogisms in which predicates are allowed to feature transitive verbs with quantified objects. They showed that this relational syllogistic does not admit a finite set of rules whose associated direct derivation relation is sound and complete. Thus for the relational syllogistic, indirect reasoning, in the form of *reduction ad absurdum* is essential. Pratt-Hartmann's paper in this volume presents a modest extension of the relational syllogistic language which is sound and complete, as desired for direct proofs. This shows that the impossibility of providing a finite rule-set for the relational syllogistic can be overcome by a modest increase in expressive power. The proof is quite complicated. Still one important conclusion from the existence of a sound and complete proof system defined by a finite set of syllogism-like rules such as the ones here is that adding relations (as transitive verbs) to a basic syllogistic logic does not represent a logical 'boundary' with respect to the expressiveness of fragments of natural language. From the previous result of Moss and Pratt-Hartmann one could get the (wrong) impression that syllogistic extensions could not be provided for transitive verbs while keeping the system sound and complete. The system RE introduced shows that this is not the case, soundness and completeness are within reach.

In "Intensions as Computable Functions", Shallom Lappin deals with an long standing problem of intensional logic, proposing a type theoretical solution. Classical intensional semantic representation languages, like Montague's Intensional Logic do not accommodate fine-grained intensionality well. In the traditional work intensional identity is reduced to equivalence of denotation across possible worlds and logically equivalent expressions are semantically indistinguishable. Thus not only all mathematical truths are the same, but also the denotations of belief statements are all logically equivalent. Lappin's paper shows that terms in the type theory PTCT (Property Theory with Curry Typing) proposed by Fox and Lappin (to appear) constitute an alternative intensional semantic representation framework. PTCT uses two notions of equality: intensional identity and extensional equivalence, and while intensional identity implies extensional equivalence, the converse is not true. Their fine-grained notions allow PTCT to prove the equivalence of mathematical truths, while allowing the non-equivalence of all belief statements. Here, Lappin proposes to characterize the distinction between intensional identity and provable equivalence *computationally* by invoking the contrast between operational and denotational semantics in programming language. Since the terms of PTCT are lambda-expressions that encode computable functions and since Lappin has identified these with the intensions of words and phrases in natural language, given the distinction between denotational and operational meaning, he can interpret the non-identity of terms in the representation language as an operational difference in the functions that these terms express. In other words if the terms compute the same result set through different sets of procedures, they are different. This approach factors modality and possible worlds out of the specification of intensions.

While the series of conferences this volume is based on had several talks on machine learning approaches to semantics, none of the authors could find the time to write up their work in a way to that would have fitted it. But "Frege in Space", the contribution of Marco Baroni, Raffaella Bernardi and Roberto Zamparelli, fills the gap with an extensive discussion of distributional semantics and its relation to traditional compositional semantics. One of the problems of classical approaches to semantics is that most lexical items are unanalyzed ('prime semantics'). Distributional semantics proposes a way to handle this through a 'the meaning of a word, is the company it keeps' approach. The approach has shown to deliver interesting results for noun and noun adjective combinations. It is less clear how to extend it to argument taking predicates and how to handle compositionality. "Frege in Space" proposes an ambitious program to do so and shows the way for a synthesis between both approaches.

We would like to dedicate the volume to our colleagues of the now defunct PARC NLTT group, especially Danny Bobrow, Dick Crouch, Lauri Karttunen, Ron Kaplan, Martin Kay, Tracy King and John Maxwell. They kindled our interest in the problems of the relation between logic, computation and natural language understanding that the volume aims to be a contribution to.

References

Cooper, R., R. Crouch, J. van Eijck, C. Fox, J. van Genabith, J. Jaspars, H. Kamp, D. Milward, M. Pinkal, M. Poesio, and S. Pulman. 1996. Using the framework. Technical Report LRE 62-051 D-16, The FraCaS Consortium.

Cooper, R., R. Crouch, J. van Eijck, C. Fox, J. van Genabith, J. Jaspars, H. Kamp, M. Pinkal, M. Poesio, S. Pulman, and E. Vestre. 1994. Describing the approaches.

Fox, C. and S. Lappin. to appear. *Formal Foundations of Intensional Semantics*. Blackwell.

MacCartney, B. 2009. *Natural Language Inference*. Ph.D. thesis, Stanford University.

MacCartney, Bill and Christopher Manning. 2009. An extended model of natural logic. In *The Eighth International Conference on Computational Semantics (IWCS-8)*.

Contents

The BIUTEE Research Platform for Transformation-based Textual Entailment Recognition

ASHER STERN AND IDO DAGAN[1]

Recent progress in research of the Recognizing Textual Entailment (RTE) task shows a constantly-increasing level of complexity in this research field. A way to avoid having this complexity becoming a barrier for researchers, especially for new-comers in the field, is to provide a freely available RTE system with a high level of flexibility and extensibility. In this paper, we introduce our RTE system, BIUTEE[2], and suggest it as an effective research framework for RTE. In particular, BIUTEE follows the prominent transformation-based paradigm for RTE, and offers an accessible platform for research within this approach. We describe each of BIUTEE's components and point out the mechanisms and properties which directly support adaptations and integration of new components. In addition, we describe BIUTEE's visual tracing tool, which provides notable assistance for researchers in refining and "debugging" their knowledge resources and inference components.

1 Introduction and Background

Textual inference (the ability to automatically find conclusions that can be inferred from a natural language text) is a capability required for many tasks at the semantic level of *Natural Language Processing (NLP)*. For example, a typical *Information Extraction (IE)* task may be to extract, from a natural language text, the employer-employee

[1]Bar-Ilan University, Ramat-Gan, Israel

[2]BIUTEE: **B**ar **I**lan **U**niversity **T**extual **E**ntailment **E**ngine. It is freely available at `http://www.cs.biu.ac.il/~nlp/downloads/biutee/`

LiLT Volume 9
Perspectives on Semantic Representations for Textual Inference.

relationship. Such a task can be formalized as the task of identifying text fragments from which it can be concluded that "X is employed by Y" for some entities X and Y. Similarly, a typical *Question Answering (QA)* task might be the task of finding answers to the question "By whom X is employed", for some entity X (a person, in this case). Addressing both these examples requires a mechanism that recognizes that the text fragment "X is employed by Y" can be inferred from a given text. *Recognizing Textual Entailment (RTE)* unifies this concept, and aims to serve as a common paradigm for textual inference. By adopting the RTE paradigm, the efforts required to solve the problem of textual inference need not to be duplicated for multiple NLP tasks which require this capability. Rather, the aim is to develop a generic solver for the RTE task, which can then be used as an inference component in task-specific systems.

The formal definition of the *Recognizing Textual Entailment* task is as follows. Given two text fragments, one termed *text* and the other *hypothesis*, the task is to recognize whether the hypothesis can be inferred from the text (Dagan et al. 2006a).

Since first introduced, several approaches have been proposed for this task, ranging from shallow lexical similarity methods (e.g., Clark and Harrison 2010; MacKinlay and Baldwin 2009), to complex linguistically-motivated methods, which incorporate extensive linguistic analysis (syntactic parsing, coreference resolution, semantic role labelling, etc.) and a rich inventory of linguistic and world-knowledge resources (e.g., Iftene 2008; de Salvo Braz et al. 2005; Bar-Haim et al. 2007). The latter methods convert the text and the hypothesis into rich representation levels, like syntactic parse-trees, semantic-role graph, or even a logical representation in which the text is converted into a collection of logical formulas. The next step is the entailment recognition itself in which the available knowledge resources are utilized.

Building such complex systems requires substantial development efforts, which might become a barrier for new-comers to RTE research. Thus, flexible and extensible publicly available RTE systems are expected to significantly facilitate research in this field. More concretely, two major research communities would benefit from a publicly available RTE system:

1. End application developers, who would use an RTE system to solve inference tasks within their application. RTE systems utilized by this type of researchers should be adaptable for the application specific data: they should be configurable, trainable, and extensible with inference knowledge that captures application-

specific phenomena.

2. Researchers in the RTE community, who would not need to build from scratch a complete RTE system for their research, but could integrate their novel research components into an existing open-source system. Such research efforts might include developing knowledge resources, developing inference components for specific phenomena such as temporal inference (see, for example, (Wang and Zhang 2008)), or extending RTE to different languages. A flexible and extensible RTE system is expected to encourage researchers to create and share their textual-inference components. A good example from another research area is the *Moses* system for *Statistical Machine Translation (SMT)* (Koehn et al. 2007), which provides the core SMT components while being extended with new research components by a large scientific community.

Until now rather few and quite limited RTE systems were made publicly available. These systems are quite restricted in the types of knowledge resources which they can utilize, and in the scope of their inference algorithms. For example, *EDITS*[3] (Kouylekov and Negri 2010) is a distance-based RTE system, which can exploit only lexical knowledge resources. *NutCracker*[4] (Bos and Markert 2005) is a system based on logical representation and automatic theorem proving, but utilizes only WordNet (Fellbaum 1998) as a lexical knowledge resource.

To address the above needs, we provide our open-source textual-entailment system, BIUTEE.[5] Our system provides state-of-the-art linguistic analysis tools and exploits various types of manually built and automatically acquired knowledge resources, including lexical, lexical-syntactic and syntactic rewrite rules. Furthermore, the system components, including pre-processing utilities, knowledge resources, and even the steps of the inference algorithm, are modular, and can be replaced or extended easily with new components. Extensibility and flexibility are also supported by a *plug-in mechanism*, by which new inference components can be integrated without changing existing code.

Notable support for researchers is provided by *a visual tracing tool, Tracer*, which visualizes every step of the inference process as shown in Figures 5 and 6, in Section 4.

This paper is organized as follows: A review of the main algorithmic components of BIUTEE is given in Section 2, followed by the system architecture description in Section 3. The visual tracing tool and its

[3] `http://edits.fbk.eu/`

[4] `http://svn.ask.it.usyd.edu.au/trac/candc/wiki/nutcracker`

[5] See footnote 2 above.

TABLE 1: A sequence of transformations that transform the text *"He received the letter from the secretary."* into the hypothesis *"The secretary delivered the message to the employee."*. The knowledge required for such transformations is often obtained from available knowledge resources and NLP tools.

#	Operation	Generated text
0	-	He received the letter from the secretary.
1	Coreference substitution	The employee received the letter from the secretary.
2	X received Y from Z → Y was sent to X by Z	The letter was sent to the employee by the secretary.
3	Y [verb-passive] by X → X [verb-active] Y	The secretary sent the letter to the employee.
4	X send Y → X deliver Y	The secretary delivered the letter to the employee.
5	letter → message	The secretary delivered the message to the employee.

typical use cases are described in Section 4, while in Section 5 we present experimental results. Conclusions, as well as suggestions for future work are given in Section 6.

2 Algorithms and Components

In this section we describe BIUTEE's main algorithms and components. This description is given at a high-level, while more details of individual components are available in the papers cited along this section.

BIUTEE (Stern and Dagan 2011) is a *transformation-based* inference system, in the sense that it transforms the text, T, into the hypothesis, H. Transforming T into H is done by applying a *sequence* of transformations, such that after applying the last transformation, the resulting text is identical to the hypothesis.[6] In this paper we use the term *proof* to refer to such a sequence of transformations. Table 1 demonstrates a proof for a typical (T,H) pair. The transformation-based paradigm requires three main design decisions:

1. How to represent the text and the hypothesis?
2. Which transformations to apply?
3. How to estimate whether a sequence of transformation preserves entailment?

In the following subsections we discuss each of these aspects. Finally, we deal with another crucial issue, namely:

[6] In practice, this goal is heuristically relaxed in the current version of BIUTEE to having the hypothesis embedded in the obtained transformed text.