Proceedings of FG 2006: The 11th conference on Formal Grammar

Malaga, Spain
July 29-30, 2006

Editor: Shuly Wintner

CENTER FOR THE STUDY
OF LANGUAGE AND INFORMATION

Preface

FG-2006, the 11th conference on Formal Grammar, was held in Malaga, Spain in July 2006. This year's conference included 12 contributed papers covering, as usual, a wide range of topics in formal grammar. In addition to those papers, this volume includes also the abstracts of two invited talks by Josef van Genabith (Dublin City University) and Laura Kallmeyer (Universität Tübingen).

The twenty four submissions to the conference were reviewed by members of the Program Committee; we are grateful to all of them for their help in making the conference a success: Anne Abeille (Paris 7, FR), Pierre Boullier (INRIA, FR), Gosse Bouma (Groningen, NL), Chris Brew (Ohio State, US), Wojciech Buszkowski (Poznan, PL), Miriam Butt (Universitaet Konstanz, DE), Alexander Clark (Royal Holloway University, UK), Berthold Crysmann (DFKI, DE), Philippe de Groote (LORIA, FR), Denys Duchier (LORIA, FR), Tim Fernando (Trinity College, IE), Annie Foret (IRISA - IFSIC, FR), Nissim Francez (Technion, IL), Gerhard Jaeger (University of Bielefeld, DE), Aravind Joshi (UPenn, US), Makoto Kanazawa (National Institute of Informatics), Stephan Kepser (Tuebingen, DE), Alexandra Kinyon (University of Pennsylvania, US), Geert-Jan Kruijff (DFKI, DE), Shalom Lappin (King's College, UK), Larry Moss (Indiana, US), Stefan Mueller (Universitaet Bremen, DE), Mark-Jan Nederhof (Max Planck Institute for Psycholinguistics, NL), James Rogers (Earlham College, US), Ed Stabler (UCLA, US), Hans Joerg Tiede (Illinois Wesleyan, US), Jesse Tseng (LORIA, FR), Willemijn Vermaat (Utrecht, NL), Anssi Yli-Jyrae (Helsinki, FI).

We are indebted to all the authors who submitted papers to the meeting, and to all participants in the Conference. On behalf of the Organizing Committee, which consisted of Paola Monachesi, Gerald Penn, Giorgio Satta and Shuly Wintner, I am happy to present this volume.

Shuly Wintner, February 2007

Contents

1

Constraint-based compositional semantics in lexicalized tree adjoining grammars

LAURA KALLMEYER

Abstract

This talk presents a framework for LTAG semantics that computes semantics based on the LTAG derivation trees such that semantic computation consists of feature unifications parallel to those performed in Feature-Based TAG (FTAG). We show that this framework has sufficient expressive power to deal with a large range of seemingly problematic phenomena, namely quantifier scope, raising verbs, bridge verbs and nested quantificational NPs. Finally, a compositionality proof is sketched for this framework that relies on the fact that the derivation tree locally determines both, syntactic and semantic composition.[1]

Keywords LEXICALIZED TREE ADJOINING GRAMMARS, COMPUTATIONAL SEMANTICS, COMPOSITIONALITY, SCOPE SEMANTICS, UNDER-SPECIFICATION

1.1 Lexicalized Tree Adjoining Grammar (LTAG)

LTAG is a tree-rewriting formalism. An LTAG consists of a finite set of *elementary* trees associated with lexical items. From these trees, larger trees are derived by substitution (replacing a leaf with a new tree) and adjunction (replacing an internal node with a new tree). LTAG derivations are represented by derivation trees that record the way the elementary trees are put together. A derived tree is the result of carrying out the substitutions and adjunctions. Each edge in the derivation tree stands for an adjunction or a substitution.

The elementary trees encapsulate all syntactic/semantic arguments of the

[1] The work presented here can be found in Kallmeyer and Romero (2007) (for the framework and the scope analyses) and Richter and Kallmeyer (2007) (for the compositionality proof).

Proceedings of FG-2006.
Editor: Shuly Wintner.

lexical anchor. They are minimal in the sense that only the arguments of the anchor are encapsulated, all recursion is factored out. Because of this, substitutions and adjunctions roughly correspond to combinations of a predicate with one of its arguments. Consequently, they determine semantic composition and therefore we compute LTAG semantics on the derivation tree.

1.2 LTAG Semantics with Semantic Unification

In our approach, each elementary tree is linked to a pair consisting of a semantic representation and a semantic feature structure description. These feature structure descriptions are used to compute assignments for variables in the representations using conjunction and additional equations introduced depending on the derivation tree.

The semantic representations consist of a set of labeled Ty2 formulas and a set of scope constraints of the form $x \geq y$ where x and y are propositional labels or propositional meta-variables. $x \geq y$ signifies that y is a component of the term x. Meta-variables indicate that terms have not been specified yet. The assignment computed based on the feature structure descriptions specifies values for some of the meta-variables in the semantic representations while leaving some of them open. This allows for under-specified representations for scope ambiguities.

1.3 Scope Phenomena

In the talk we present analyses for the scope ambiguities exemplified in (1)–(5):

(1) Exactly one student admires every professor
$\exists > \forall, \forall > \exists$

(2) John seems to have visited everybody
$\textit{seem} > \forall, \forall > \textit{seem}$

(3) Three girls are likely to come
$\textit{three} > \textit{likely}, \textit{likely} > \textit{three}$

(4) Mary thinks John likes everybody
thinks > everybody, *everybody > thinks

(5) Two policemen spy on someone from every city
$\forall > \exists > 2$ (among others), $*\forall > 2 > \exists$

Our analysis models the differences in scope behavior as follows:

1. Quantifiers scope within a kind of scope window delimited by an upper boundary MAXS and a lower boundary MINS, no matter where they attach inside a finite clause.

2. Operators on the verbal spine such as adverbs, raising verbs and bridge verbs take scope where they attach, i.e., among such operators, the attachment order specifies the scope order.
3. Adverbs and raising verbs are not concerned with the MAXS–MINS scope window. Therefore, quantifiers can scopally interleave with them.
4. Bridge verbs embed a finite clause and in particular, they embed the MAXS limit of this clause. Therefore they block quantifier scope.
5. The maximal scope of a quantifier embedded in a quantificational NP is the proposition of the embedding quantifier. Therefore, if it scopes over the embedding quantifier, then this has to be immediate scope (no other quantifier can intervene).

1.4 Compositionality

At first sight, feature logic-based computational semantics systems such as LTAG do not seem compatible with a notion of compositionality. The derived trees clearly do not determine the meaning of a phrase in a compositional way. However, a crucial property of LTAG is that the derivation process (i.e., the process of syntactic combination) can be described by a context-free structure, namely the derivation tree. (This is why LTAG is mildly context-sensitive.) The way our LTAG semantics framework is defined, this context-free structure also specifies the process of semantic combination. In other words, we can define semantic denotations for the nodes in the derivation tree in such a way that the semantic denotation of a node depends only the denotations of the daughters, the semantic representation from the lexicon chosen for this node and the way the daughters combine with the mother. In this sense, LTAG semantics is compositional.

References

Kallmeyer, Laura and Maribel Romero. 2007. Scope and Situation Binding in LTAG using Semantic Unification. To appear in *Research on Language and Computation*.

Richter, Frank and Laura Kallmeyer. 2007. Feature Logic-based Semantic Composition: A Comparison between LRS and LTAG. To appear in *Postproceedings of the Workshop on Typed Feature Structure Grammars, the 22nd Scandinavian Conference of Linguistics*.

2

Parsing and generation with treebank-based probabilistic LFG resources

JOSEF VAN GENABITH

Treebank-based acquisition of “deep” grammar resources is motivated by the “knowledge acquisition bottleneck” familiar from other traditional, knowledge intensive, rule-based approaches in AI and NLP, following the “rationalist” research paradigm. Deep grammatical resources have usually been hand-crafted Butt et al. (2002), Baldwin et al. (2004). This is time consuming, expensive and difficult to scale to unrestricted text. Treebanks (parse-annotated corpora) have underpinned an alternative “empiricist” approach: wide-coverage, robust probabilistic grammatical resources are now routinely extracted (learned) from treebank resources Charniak (1996), Collins (1997), Charniak (2000). Initially, however, these resources have been “shallow”. More recently, a considerable amount of research has emerged on treebank-based acquisition of deep grammatical resources in the TAG, HPSG, CCG and LFG grammar formalisms. This talk provides an overview of research on rapid treebank-based acquisition of wide-coverage, robust, probabilistic, multilingual LFG resources. Grammar and lexicon acquisition O’Donovan et al. (2005) is based on an automatic LFG f-structure annotation algorithm Burke et al. (2004a), Burke (2006). I show how the acquired LFG resources can be used in wide-coverage, robust parsing Cahill et al. (2004) and generation Cahill and van Genabith (2006). I provide an overview of ongoing research on the induction of Chinese Burke et al. (2004b), Japanese, Arabic, Spanish, French and German Cahill et al. (2005) treebank-based LFG resources.

Proceedings of FG-2006.
Editor: Shuly Wintner.

I briefly compare our LFG work with similar research on the treebank-based acquisition of HPSG Miyao et al. (2003) and CCG Hockenmaier and Steedman (2002) resources.

References

Baldwin, Timothy, Emily Bender, Dan Flickinger, Ara Kim, and Stephan Oepen. 2004. Road-testing the English Resource Grammar over the British National Corpus. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC 2004)*, pages 2047–2050. Lisbon, Portugal.

Burke, Michael. 2006. *Automatic Treebank Annotation for the Acquisition of LFG Resources*. Ph.D. thesis, School of Computing, Dublin City University, Dublin 9, Ireland.

Burke, Michael, Aoife Cahill, Ruth O'Donovan, Josef van Genabith, and Andy Way. 2004a. Evaluation of an Automatic Annotation Algorithm against the PARC 700 Dependency Bank. In *Proceedings of the Ninth International Conference on LFG*, pages 101–121. Christchurch, New Zealand.

Burke, Michael, Olivia Lam, Rowena Chan, Aoife Cahill, Ruth O'Donovan, Adams Bodomo, Josef van Genabith, and Andy Way. 2004b. Treebank-Based Acquisition of a Chinese Lexical-Functional Grammar. In *Proceedings of the 18th Pacific Asia Conference on Language, Information and Computation*, pages 161–172. Tokyo, Japan.

Butt, Miriam, Helge Dyvik, Tracy Holloway King, Hiroshi Masuichi, and Christian Rohrer. 2002. The Parallel Grammar Project. In *Proceedings of COLING 2002, Workshop on Grammar Engineering and Evaluation*, pages 1–7. Taipei, Taiwan.

Cahill, Aoife, Michael Burke, Martin Forst, Ruth O'Donovan, Christian Rohrer, Josef van Genabith, and Andy Way. 2005. Treebank-Based Acquisition of Multilingual Unification Grammar Resources. *Research on Language and Computation* **3**(2-3):247–279.

Cahill, Aoife, Michael Burke, Ruth O'Donovan, Josef van Genabith, and Andy Way. 2004. Long-Distance Dependency Resolution in Automatically Acquired Wide-Coverage PCFG-Based LFG Approximations. In *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics*, pages 320–327. Barcelona, Spain.

Cahill, Aoife and Josef van Genabith. 2006. Robust PCFG-Based Generation using Automatically Acquired Treebank-Based LFG Approximations. In *ACL/COLING 2006*. Sydney, Australia.

Charniak, Eugene. 1996. Tree-Bank Grammars. In *Proceedings of the Thirteenth National Conference on Artificial Intelligence*, pages 1031–1036. Menlo Park, CA.

Charniak, Eugene. 2000. A maximum entropy inspired parser. In *Proceedings of the First Annual Meeting of the North American Chapter of the Association for Computational Linguistics (NAACL 2000)*, pages 132–139. Seattle, WA.

Collins, Michael. 1997. Three Generative, Lexicalized Models for Statistical Parsing. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics*, pages 16–23. Madrid, Spain.

Hockenmaier, Julia and Mark Steedman. 2002. Acquiring Compact Lexicalized Grammars from a Cleaner Treebank. In *Proceedings of the 3rd International Conference Language Resources and Evaluation*. Las Palmas, Grand Canaria, Spain.

Miyao, Yusuke, Takashi Ninomiya, and Jun'ichi Tsujii. 2003. Probabilistic modeling of argument structures including non-local dependencies. In *Proceedings of the Conference on Recent Advances in Natural Language Processing (RANLP)*, pages 285–291. Borovets, Bulgaria.

O'Donovan, Ruth, Michael Burke, Aoife Cahill, Josef van Genabith, and Andy Way. 2005. Large-Scale Induction and Evaluation of Lexical Resources from the Penn-II and Penn-III Treebanks. *Computational Linguistics* **31**(3):329–365.