TAG+11

Proceedings of the 11th International Workshop on

Tree Adjoining Grammars and Related Formalisms

26-28 September 2012
Paris, France

Production and Manufacturing by
INRIA, France

Eleventh International Workshop on Tree Adjoining Grammars and Related Formalisms
(TAG+11)
URL: `http://alpage.inria.fr/tagplus11`

Endorsed by

Sponsored by

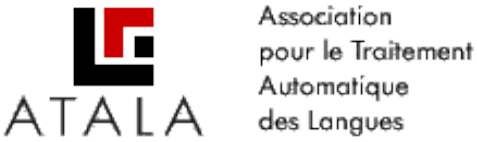

Photo credit: Fotolia
ISBN: 978-1-937284-46-6

Preface

This volume contains papers accepted for presentation at the Eleventh International Workshop on Tree Adjoining Grammar and Related Formalisms, TAG+11 for short, to be held on September 26–28, 2012 in Paris. TAG+ is a biennial workshop series that fosters exchange of ideas among linguists, psycho-linguists and computer scientists interested in modeling natural language using formal grammars. The workshop series, since 1990, has demonstrated productive interactions among researchers and practitioners interested in various aspects of the tree adjoining grammar formalism and its relationship to other grammar formalisms, such as combinatory categorial grammar, dependency grammars, linear context-free rewriting systems, minimalist grammars, head-driven phrase structure grammars, and lexical functional grammars; hence the + in the name of the workshop.

We would like to thank the members of the program committee for their careful and timely work, especially those who participated in discussions on diverging reviews. This meeting would not have been possible without the hard work of all these people. We would also like to thank our invited speakers, Kevin Knight and Bonnie Webber, and the speakers at the tutorial program, David Chiang, Vera Demberg, Laura Kallmeyer and Andreas Maletti. We acknowledge the effort of these speakers in fostering new interest in TAG and more generally in formal research into natural language. Last but not least, we would like to thank the local organizers, Éric de la Clergerie, Djamé Seddah, Laurence Danlos and Chantal Girodon, for their invaluable contribution to the organization of the TAG+11 workshop in Paris, and for securing the necessary funding that made it possible to realize this workshop. We would also like to acknowledge the support staff at INRIA Paris Rocquencourt and at University Paris-Diderot. Funding for TAG+11 was provided by Google, the Association pour le Traitement Automatique des Langues (ATALA), INRIA Paris Rocquencourt, and University Paris-Diderot.

TAG+11 received 36 long abstract submissions from all over the world, and we were able to accept 28 papers out of these 36. This volume contains the 27 research papers to be presented at TAG+11 (one paper had been later withdrawn from the program). 19 papers are to be delivered in oral presentations and eight are to be presented as posters. As at previous TAG+ workshops, the topics addressed by the presentations belong to diverse areas of research, including mathematics of grammar formalisms, parsing algorithms for mildly context-sensitive grammars, language learnability, syntax and semantics of natural languages, and relation between TAG and other grammar formalisms. The oral presentations were thus organized into several different sessions: syntax/semantics, formalisms, derivation trees and applications, grammar extraction and grammar induction, and parsing. By bringing together these different topics under the common theme of Tree Adjoining Grammars, the workshop promises to be a venue for interesting discussion of the latest research in this area.

Chung-hye Han
Giorgio Satta
Program co-Chairs for TAG+11

Organizers

Program Chair:

Chung-Hye Han (Simon Fraser University, Canada)
Giorgio Satta (University of Padova, Italy)

Logistic Arrangements:

Éric Villemonte de la Clergerie (INRIA Paris Rocquencourt, France)
Djamé Seddah (University Paris Sorbonne Paris 4, France)
Laurence Danlos (University Paris Diderot, France)
Chantal Girodon (INRIA Paris Rocquencourt, France)

Programme Committee:

Srinivas Bangalore (AT&T Research, USA)
Rajesh Bhatt (University of Massachusetts at Amherst, USA)
David Chiang (USC Information Sciences Institute, USA)
Benoit Crabbé (University Paris-Diderot, France)
Robert Frank (Yale University, USA)
Claire Gardent (CNRS/LORIA, Nancy, France)
Daniel Gildea (University of Rochester, USA)
Julia Hockenmaier (University of Illinois at Urbana-Champaign, USA)
Liang Huang (USC Information Sciences Institute, USA)
Aravind Joshi (University of Pennsylvania, USA)
Laura Kallmeyer (University of Düsseldorf, Germany)
Makoto Kanazawa (National Institute of Informatics, Japan)
Alexander Koller (University of Potsdam, Germany)
Marco Kuhlmann (Uppsala University, Sweden)
Andreas Maletti (University of Stuttgart, Germany)
Yusuke Miyao (University of Tokyo, Japan)
Mark-Jan Nederhof (University of St Andrews, Scotland)
Steve DeNeefe (SDL Language Weaver, USA)
Owen Rambow (Columbia University, USA)
Maribel Romero (University of Konstanz, Germany)
Tatjana Scheffler (DFKI, Germany)
Anoop Sarkar (Simon Fraser University, Canada)
William Schuler (The Ohio State University, USA)
Stuart Shieber (Harvard University, USA)
Mark Steedman (University of Edinburgh, UK)
Matthew Stone (Rutgers University, USA)
Sylvain Salvati (INRIA Bordeaux Sud-Ouest, France)
Bonnie Webber (University of Edinburgh, Scotland)
Fei Xia (University of Washington, USA)

Invited Speakers:

Kevin Knight (University of Southern California, USA)
Bonnie Webber (University of Edinburgh, Scotland)

Tutorials:

David Chiang (USC Information Sciences Institute, USA)
Vera Demberg (Saarland University, Germany)
Laura Kallmeyer (University of Düsseldorf, Germany)
Andreas Maletti (University of Stuttgart, Germany)

Table of Contents

Conference Program

Wednesday, September 26th

9:15–9:30 Opening remarks

9:30–11:00 **Tutorial by David Chiang**
Synchronous Grammars

11:00–11:30 Coffee

11:30–13:00 **Tutorial by Vera Demberg**
Tree-Adjoining Grammars from a psycholinguistic perspective

13:00–14:30 Lunch

14:30–16:00 **Tutorial by Laura Kallmeyer**
LCFRS+: Linear Context-Free Rewriting Systems and Related Formalisms

16:00–16:30 Coffee

16:30–18:00 **Tutorial by Andreas Maletti**
Trees abound: A primer on tree automata and tree transducers

Thursday, September 27th

9:00–10:00 **Invited Talk by Bonnie Webber**
Alternatives, Discourse Semantics and Discourse Structure

10:00–10:20 Coffee

10:20–12:20 **Talk Session on Syntax/Semantics**

Delayed Tree-Locality, Set-locality, and Clitic Climbing
Joan Chen-Main, Tonia Bleam and Aravind Joshi

Deriving syntax-semantics mappings: node linking, type shifting and scope ambiguity
Dennis Ryan Storoshenko and Robert Frank

Tree Adjunction as Minimalist Lowering
Thomas Graf

A Frame-Based Semantics of Locative Alternation in LTAG
Yulia Zinova and Laura Kallmeyer

12:20–13:50 Lunch

Thursday, September 27th (continued)

13:50–15:20 **Talk Session on Formalisms**

A Logical Characterization of Extended TAGs
Uwe Mönnich

Synchronous Tree Unification Grammar
Timm Lichte

Synchronous Context-Free Tree Grammars
Mark-Jan Nederhof and Heiko Vogler

15:20–15:40 **Poster Quickfire Session**

Incremental Neo-Davidsonian semantic construction for TAG
Asad Sayeed and Vera Demberg

Representing Focus in LTAG
Kata Balogh

Describing São Tomense Using a Tree-Adjoining Meta-Grammar
Emmanuel Schang, Denys Duchier, Brunelle Magnana Ekoukou, Yannick Parmentier and Simon Petitjean

An Attempt Towards Learning Semantics: Distributional Learning of IO Context-Free Tree Grammars
Ryo Yoshinaka

15:40–16:30 Poster session and Coffee

16:30–18:00 **Talk Session on Derivation Trees and Applications**

Delayed Tree Locality and the Status of Derivation Structure
Joan Chen-Main

A Formal Model for Plausible Dependencies in Lexicalized Tree Adjoining Grammar
Laura Kallmeyer and Marco Kuhlmann

Using FB-LTAG Derivation Trees to Generate Transformation-Based Grammar Exercises
Claire Gardent and Laura Perez-Beltrachini

Friday, September 28th

9:00–10:00 **Invited Talk by Kevin Knight**
Transformation Frameworks for Machine Translation: Strings, Trees, and Graphs

10:00–10:30 Coffee

Friday, September 28th (continued)

10:30–12:30 **Talk Session on Parsing**

PLCFRS Parsing Revisited: Restricting the Fan-Out to Two
Wolfgang Maier, Miriam Kaeshammer and Laura Kallmeyer

Decomposing TAG Algorithms Using Simple Algebraizations
Alexander Koller and Marco Kuhlmann

Practical Parsing of Parallel Multiple Context-Free Grammars
Peter Ljunglöf

Idioms and extended transducers
Gregory M. Kobele

12:30–14:00 Lunch

14:00–15:30 **Talk Session on Grammar Extraction, Grammar Induction**

Creating a Tree Adjoining Grammar from a Multilayer Treebank
Rajesh Bhatt, Owen Rambow and Fei Xia

Search Space Properties for Learning a Class of Constraint-based Grammars
Smaranda Muresan

State-Split for Hypergraphs with an Application to Tree Adjoining Grammars
Johannes Osterholzer and Torsten Stüber

15:30–15:50 **Poster Quickfire Session**

Is Syntactic Binding Rational?
Thomas Graf and Natasha Abner

Incremental Derivations in CCG
Vera Demberg

On the Form-Meaning Relations Definable by CoTAGs
Gregory M. Kobele and Jens Michaelis

A linguistically-motivated 2-stage Tree to Graph Transformation
Corentin Ribeyre, Djamé Seddah and Éric Villemonte de la Clergerie

15:50–16:30 Poster session and Coffee

16:30–17:30 **Talk Session on Syntax/Semantics**

Scope Economy and TAG Locality
Michael Freedman

The Shape of Elementary Trees and Scope Possibilities in STAG
Robert Frank and Dennis Ryan Storoshenko

17:30–17:45 Closing remarks

Invited Talks

Transformation Frameworks for Machine Translation: Strings, Trees, and Graphs

Kevin Knight
University of Southern California, USA

Accurate machine translation (MT) of human languages is a longstanding challenge for computer science. Probabilistic string, tree, and graph automata provide nice formal frameworks for large-scale statistical MT systems. This talk addresses the power of these frameworks and how well they fit observed human translation data.

Alternatives, Discourse Semantics and Discourse Structure

Bonnie Webber
University of Edinburgh, Scotland

Sentence-level modality and negation give rise to "alternative" events and/or situations that contribute to discourse semantics in interesting ways. But they are not the only linguistic elements that do this. In this talk, I will try to characterise the range of elements that give rise to alternatives and the nature of these events and situations. I will then show how these alternatives are distinct from the coherence relations that provide a low-level of discourse structure.

Tutorials

Synchronous Grammars

David Chiang
USC Information Sciences Institute, USA

Synchronous context-free grammars (CFGs), first proposed in the 1960s, have become a popular and powerful tool in machine translation, semantic parsing, and other areas of natural language processing. Synchronous tree-adjoining grammars (TAGs) were first proposed in the 1990s and are starting to see interesting applications. The theory behind synchronous grammars, and the algorithms that power their applications, are sometimes natural extensions of those of conventional grammars, but there are also some surprising twists and turns. I will give an introduction to synchronous CFGs and TAGs, present some of their key formal properties, and describe the main algorithms that use them. I will also describe some synchronous grammar formalisms beyond synchronous TAG, like synchronous hyperedge replacement grammars.

Tree-Adjoining Grammars from a psycholinguistic perspective

Vera Demberg
Saarland University, Germany

We will first review some psycholinguistic experiments that are revealing about certain properties of human language comprehension, such as to what degree sentence processing is incremental and possibly even connected, as well as studies that indicate that humans actively predict upcoming input. We will thereby cover syntactic as well as semantic and discourse-level effects.

In the second part of the tutorial, I will discuss how these effects can be modelled using Tree-Adjoining Grammars.

LCFRS+: Linear Context-Free Rewriting Systems and Related Formalisms

Laura Kallmeyer
University of Düsseldorf, Germany

Recently, there has been an increased interest in Linear Context-Free Rewriting Systems (LCFRSs), due to their mild context-sensitivitiy and their capacity to describe discontinuous constituents and non-projective dependencies. LCFRS research has particularly intensified in the area of parsing.

In this tutorial, LCFRSs will be motivated and introduced. Furthermore, closely related formalisms such as Multiple Context-Free Grammars (MCFG) and simple Range Concatentation Grammars (RCG) will be defined and related to LCFRS. The link between LCFRS and the notion of mild context-sensitivity will be discussed and, in this context, the question whether one might even want to go beyond LCFRS will be raised. The more powerful formalisms of (unrestricted) RCG and Literal Movement Grammars (LMG) will be introduced that both are natural extensions of LCFRS, depending on whether the LCFRS rules are understood as manipulating strings or manipulating concrete occurrences of substrings of some input string. The former leads to LMG while the latter leads to RCG.

The aim of the tutorial is to give an overview of the formal grammar landscape ranging from CFG to LCFRS, RCG and LMG, relating the different types of rewriting rules and the different language classes defined by these formalisms. The expressive power and the limitations of these grammars are illustrated by numerous examples.

Trees abound: A primer on tree automata and tree transducers

Andreas Maletti
University of Stuttgart, Germany

We introduce tree automata and tree transducers formally and on examples. We also link them to the (synchronous) grammar notions that are better known in NLP. We then proceed to review most of the basic tree automata and tree transducer results with tie-ins into current results obtained in the NLP community. Finally, we cover some interesting advanced results that so far received little interest from the NLP community.