Proceedings of

SSST

NAACL-HLT 2007 / AMTA Workshop on

Syntax and Structure in Statistical Translation

Dekai Wu and David Chiang (editors)

26 April 2007
Rochester, New York, USA

Production and Manufacturing by
Omnipress Inc.
Post Office Box 7214
Madison, WI 53707-7214

Association for Machine Translation in the Americas (AMTA)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
`business@amtaweb.org`

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
75 Paterson Street, Suite 9
New Brunswick, NJ 08901
USA
Tel: +1-732-342-9100
Fax: +1-732-342-9339
`acl@aclweb.org`

Introduction

The NAACL-HLT 2007 / AMTA Workshop on Syntax and Structure in Statistical Translation (SSST) took place on 26 April 2007 following the NAACL-HLT conference hosted by the University of Rochester in New York. It was organized in response to growing interest in statistical, tree structured models of relations between natural languages. Our hope was to bring together researchers working on various aspects of this subject, and coming from various traditions. One way that the diversity of these traditions can be seen is in their nomenclature: transduction grammars originated in formal language theory (Lewis and Stearns 1968, Aho and Ullman 1969), and as interest in them was renewed in the computational linguistics literature in the 1990s, they came to be also known as synchronous grammars. Pushdown transducers and tree transducers, also introduced in the late 1960s, embody a less declarative, rather more procedural view, but, in many cases, have transduction-grammar equivalents.

Another dimension of diversity is the variety of applications of synchronous/transduction grammars, which is richly reflected in our workshop program. We selected fourteen papers, which include papers on formal properties of synchronous/transduction grammars from both theoretical (Shieber) and comparative experimental (Zhang and Gildea; Huang; Dreyer, Hall and Khudanpur) perspectives, and papers applying synchronous/transduction grammars to machine translation as well as generation (Hall and Němec) and semantic interpretation (Nesson and Shieber). The invited speaker for the workshop was William C. Rounds of the University of Michigan, a pioneer of tree-transducer theory who was one of the first to explore the usefulness of tree transducers for natural language.

The papers included a wide spectrum of experiments trying different tradeoffs between representational adequacy versus efficiency. Some models adopted binary-rank ITG or inversion transduction grammar constraints (Cherry and Lin; Huang; Dreyer, Hall and Khudanpur), while others permitted up to STAG or synchronous tree-adjoining grammar expressiveness (Nesson and Shieber; Shieber), with others in between at the SDTG or syntax directed transduction grammar a.k.a. SCFG or synchronous context-free grammar level (Zhang, Zens and Ney; Zhang and Gildea). Transduction rules ranged from mildly hierarchical, heavily lexical transduction rules on one hand (Cherry and Lin; Zhang, Zens and Ney; Venkatapathy and Bangalore; Dreyer, Hall and Khudanpur), to abstract transduction rules emphasizing compositional syntax on the other (Nesson and Shieber; Hall and Němec; Shieber).

A number of papers investigated machine learning techniques for inducing synchronous/transduction grammars (Zhang, Zens and Ney; Cherry and Lin). Some of these focused on improving algorithms for binarizing or reducing the rank of synchronous/transduction grammars (Zhang and Gildea; Huang). The workshop also witnessed a number of papers proposing new ways of integrating tree structured models into statistical methods in machine translation (Hopkins and Kuhn; Venkatapathy and Joshi; Bonneau-Maynard, Allauzen, Déchelotte and Schwenk; Font Llitjós and Vogel; Owczarzak, van Genabith and Way; Venkatapathy and Bangalore).

The Association for Machine Translation in the Americas sponsored $1000 in scholarships for several students to attend the workshop. We thank AMTA for their generosity, and we also thank the Program Committee for their extremely quick reviews.

Dekai Wu and David Chiang

Organizers:

Dekai WU, Hong Kong University of Science and Technology (HKUST), Hong Kong
David CHIANG, USC Information Sciences Institute, USA

Program Committee:

Srinivas BANGALORE, AT&T Research, USA
Marine CARPUAT, Hong Kong University of Science and Technology (HKUST), Hong Kong
Daniel GILDEA, University of Rochester, USA
Kevin KNIGHT, USC Information Sciences Institute, USA
Daniel MARCU, USC Information Sciences Institute, USA
Hermann NEY, RWTH Aachen, Germany
Owen RAMBOW, Columbia University, USA
Philip RESNIK, University of Maryland, USA
Giorgio SATTA, University of Padua, Italy
Stuart M. SHIEBER, Harvard University, USA
Christoph TILLMANN, IBM, USA
Enrique VIDAL, Universidad Politécnica de Valencia, Spain
Stephan VOGEL, Carnegie Mellon University, USA
Andy WAY, Dublin City University, Ireland
Taro WATANABE, NTT, Japan
Richard ZENS, RWTH Aachen, Germany

Table of Contents

Conference Program

Thursday, April 26, 2007

9:00–9:05 Opening

9:05–9:30 *Chunk-Level Reordering of Source Language Sentences with Automatically Learned Rules for Statistical Machine Translation*
Yuqi ZHANG, Richard ZENS and Hermann NEY

9:30–9:55 *Extraction Phenomena in Synchronous TAG Syntax and Semantics*
Rebecca NESSON and Stuart M. SHIEBER

9:55–10:45 Invited Talk by William C. ROUNDS

10:45–11:15 Coffee Break

11:15–11:40 *Inversion Transduction Grammar for Joint Phrasal Translation Modeling*
Colin CHERRY and Dekang LIN

11:40–12:05 *Factorization of Synchronous Context-Free Grammars in Linear Time*
Hao ZHANG and Daniel GILDEA

12:05–12:30 *Binarization, Synchronous Binarization, and Target-side Binarization*
Liang HUANG

12:30–14:00 Lunch

14:00–14:25 *Machine Translation as Tree Labeling*
Mark HOPKINS and Jonas KUHN

14:25–14:50 *Discriminative word alignment by learning the alignment structure and syntactic divergence between a language pair*
Sriram VENKATAPATHY and Aravind JOSHI

14:50–15:15 *Generation in Machine Translation from Deep Syntactic Trees*
Keith HALL and Petr NĚMEC

Thursday, April 26, 2007 (continued)

15:15–16:00 Posters with Coffee

Combining Morphosyntactic Enriched Representation with n-best Reranking in Statistical Translation
Hélène BONNEAU-MAYNARD, Alexandre ALLAUZEN, Daniel DÉCHELOTTE and Holger SCHWENK

A Walk on the Other Side: Using SMT Components in a Transfer-Based Translation System
Ariadna FONT LLITJÓS and Stephan VOGEL

Dependency-Based Automatic Evaluation for Machine Translation
Karolina OWCZARZAK, Josef VAN GENABITH and Andy WAY

Probabilistic Synchronous Tree-Adjoining Grammars for Machine Translation: The Argument from Bilingual Dictionaries
Stuart M. SHIEBER

16:00–16:25 *Three models for discriminative machine translation using Global Lexical Selection and Sentence Reconstruction*
Sriram VENKATAPATHY and Srinivas BANGALORE

16:25–16:50 *Comparing Reordering Constraints for SMT Using Efficient BLEU Oracle Computation*
Markus DREYER, Keith HALL and Sanjeev KHUDANPUR

16:50–17:50 Panel Discussion

17:50–18:00 Closing