Introduction

The Fourth Workshop on Syntax and Structure in Statistical Translation (SSST-4) was held on 28 August 2010 following the Coling 2010 conference in Beijing. Like the first three SSST workshops in 2007, 2008, and 2009, it aimed to bring together researchers from different communities working in the rapidly growing field of statistical, tree-structured models of natural language translation.

We were honored to have Martin Kay deliver this year's invited keynote talk. This field is indebted to Martin Kay for not one but *two* of the classic cornerstone ideas that inspired bilingual tree-structured models for statistical machine translation: first, chart parsing, and second, parallel text alignment.

Tabular approaches to parsing, using dynamic programming and/or memoization, were heavily influenced by Kay's (1980) charts (or forests, packed forests, well-formed substring tables, or WFSTs). Today's biparsing models—the bilingual generalizations of this influential work—lie at the heart of numerous alignment and training algorithms for inversion transduction grammars or ITGs—including all syntax-directed transduction grammars or SDTGs (or synchronous CFGs) of binary rank or ternary rank, such as those learned by hierarchical phrase-based translation approaches.

At the same time, Kay and Röscheisen's (1988) seminal work on alignment of parallel texts led the way in statistical machine translation's basic paradigm of integrating the simultaneous learning of translation lexicons with aligning parallel texts. Today's biparsing models generalize this by simultaneously learning tree structures as well. Once again, cross-pollination of ideas across different areas and disciplines, empirical and theoretical, has provided mutual inspiration.

We selected 15 papers for this year's workshop. Studies emphasizing formal and algorithmic aspects include a method for intersecting synchronous/transduction grammars (S/TGs) with finite-state transducers (Dymetman and Cancedda), and a comparison of linear transduction grammars (LTGs) with ITGs (Saers, Nivre and Wu). Experiments on using syntactic features and constraints within flat phrase-based translation models include studies by Jiang, Du and Way; by Cao, Finch and Sumita; by Kolachina, Venkatapathy, Bangalore, Kolachina and PVS; and by Zhechev and van Genabith. Dependency constraints are also used to improve HMM word alignments for both flat phrase-based as well as S/TG based translation models (Ma and Way). Extensions to the features and parameterizations in two S/TG based translation models, as well as methods for merging models, are studied by Zollman and Vogel. The potential of incorporating LFG-style deep syntax within S/TGs is explored by Graham and van Genabith. A tree transduction based approach is presented by Khalilov and Sima'an. Meanwhile, Lo and Wu empirically compare n-gram, syntactic, and semantic structure based MT evaluation approaches. An encouraging trend is an uptick in work on low-resource language pairs and from underrepresented regions, including English-Persian (Mohaghegh, Sarrafzadeh and Moir), Manipuri-English (Singh and Bandyopadhyay), Tunisia (Khemakhem, Jamoussi and Ben hamadou), and English-Hindi (Venkatapathy, Sangal, Joshi, and Gali).

Once again this year, thanks are due to our authors and our Program Committee for making the SSST workshop another success.