Introduction

The Third Workshop on Syntax and Structure in Statistical Translation (SSST-3) was held on 5 June 2009 following the NAACL-HLT 2009 conference hosted by the University of Colorado at Boulder. Like the first two SSST workshops in 2007 and 2008, 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 Alfred V. Aho deliver this year's invited keynote talk. Along with Lewis and Stearns' (1968) seminal "Syntax-directed transduction" introducing syntax-directed transduction grammars or SDTGs—to which synchronous CFGs are also equivalent—the classic pair of Aho and Ullman's (1969) articles "Syntax-directed translations and the pushdown assembler" and "Properties of syntax-directed translations" established the foundations of formal transduction approaches to translation. The motivation behind their formal language theory work was compiler translation of programming languages. But much of the current work at SSST reflects the evolution of those ideas into today's state-of-the-art approaches to modeling of syntax and structure in statistical machine translation of human languages.

Nowhere is it better demonstrated that research is healthily driven by the pollination of ideas across these disciplines. These formal language pioneers foresaw clearly the importance of formalizing compositional models of transduction. But without empirical research across various natural languages, formal language theorists would not perhaps have anticipated the wide applicability in human language translation of an intermediate restricted class of transductions between the syntax-directed transductions that can be described by SDTGs (or synchronous CFGs) at one extreme, and the very restricted finite-state transductions that can be described by FSTs at the other—like the broad equivalence class of inversion transductions that can be described by ITGs (which include synchronous/transduction grammars whose rules are all binary rank, ternary rank, or monotonically straight or inverted in reordering permutation). Nor might they have foreseen the success of the rich variety of statistical machine learning techniques that have been developed to induce such synchronous/transduction grammars, such as the techniques introduced by the hierarchical phrase-based translation approach. As has so often happened over the years of cross-fertilization cycles between formal language and natural language research, the theoretical and empirical lines of research provide mutual inspiration.

We selected ten papers for this year's workshop. Studies on alignment ranged from the theoretical (Søgaard) to data analysis (Nakazawa and Kurohashi; Søgaard and Kuhn; Jiang, Li, Yang and Zhao), to empirical impact on actual translation performance (Saers and Wu; Hashimoto, Yamamoto, Okuma, Sumita and Tokuda). New contributions to translation decoding included purely unsupervised methods leveraging compositional structure constraints (Saers and Wu), methods using explicit syntactic information (Chang, Tseng, Jurafsky and Manning; Khalilov, Fonollosa and Dras), as well as methods attempting to blend the two (Hashimoto, Yamamoto, Okuma, Sumita and Tokuda; Hanneman and Lavie). The program was rounded out by a paper considering the use of explicit syntax in automatic evaluation (Wang, Zhao, Yang and Li).

We would like to thank our authors and our Program Committee for making this year's SSST workshop another success.

Dekai Wu and David Chiang